'IN-THE-MOMENT' PHYSIOLOGICAL RESPONSES TO RECREATIONAL MUSIC ACTIVITIES IN PEOPLE LIVING WITH DEMENTIA

EMILIE VICTORIA BROTHERHOOD

A thesis submitted to University College London (UCL)
for the degree of Doctor of Philosophy

Dementia Research Centre
Queen Square Institute of Neurology
University College London

2023

DECLARATION

I, Emilie Victoria Brotherhood, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Evidence for music being a 'relative island in a sea of impairments' in people living with a dementia (PLWD) mostly relies on data capture before and after the fact, overlooking the opportunity to explore physiological phenomena observable 'in-the-moment'. This PhD and the work herein explores physiological data acquired using passive recording techniques while PLWD engage in recreational music activities.

The Play It Again study acquired pupillometry responses alongside pleasantness and familiarity ratings of music excerpts in 90 healthy older adults (HOAs; **Chapter 3**) and 39 people living with typical [tAD] and atypical Alzheimer's disease [Posterior Cortical Atrophy; PCA] (**Chapter 4**). Within HOA and tAD groups, curvilinear pupil-affect profiles were observed. In all participant groups, pupillometry profiles were distinguishable to an extent between familiar and unfamiliar songs.

A secondary analysis of the Play It Again dataset (**Chapter 5**) harnessed machine-learning methods to attempt classification between the two syndromic groups using a 'fusion model' of both pupil-based and self-report latency metrics. The model yielded 'excellent' discrimination classification performance, which remained unaltered when considering different missing data thresholds within the pupillometry pre-processing pipeline.

The final empirical study, Our Dementia Choir, (**Chapter 6**) acquired physiological data in 16 PLWD before, during and after six choral rehearsals using discreet wearable devices. Accompanying pre- and post-rehearsal wellbeing and stress scales revealed Bonferroni-corrected significant increases in composite wellbeing scores in three of the six weeks, alongside differential overall within-rehearsal autonomic responses in multiple physiological domains relative to baseline (pre-session) levels.

Methodological limitations were noted, particularly in terms of challenges acquiring physiological data from individuals with oculomotor difficulties, or in real-world settings. Nevertheless, findings of a coupling between physiological responses and (a) self-reported musical experiences or (b) low-burden disease classification indicate justification for future work to integrate biosignal responses to music in dementia classification, care and support.

IMPACT STATEMENT

The findings build on existing evidence and advance our understanding of observable physiological responses to music in PLWD. Key findings are discussed specifically within the context of moving the dementia-music-physiology field beyond academic interest and into clinical relevance. The devices used to acquire the physiological responses are either relatively low-cost (e.g. Empatica® E4) or adaptable to lower-cost commercial (pupillometry acquisition) systems to develop further work in clinical or care settings.

Technical skills acquired from undertaking the current work have contributed to the development and fine-tuning of further dementia research. Procedural synchronisation and automated analysis pipelines [***Our Dementia Choir***] have informed study protocols to trial remote physiological testing in frontotemporal dementia/control dyads [*Music And Neuroscience against Dementia: from Designs to Outcomes through Listening INterventions INclusively INformed for INdividuals*] ("MANDDOLIN4 - Music And Neuroscience against Dementia," 2022). These will also be used to analyse datasets acquired in collaboration with conservatoires [e.g., retrospective analysis of video data collected during Royal Academy of Music (RAM)/Wigmore Hall online music programmes for PWLD, using automated facial emotion recognition software (Opstad, n.d.)]. Plans are underway to extend recruitment to the Play It Again paradigm in future work with other syndromic groups (lvPPA) as part of an upcoming PhD project.

The acquisition of technical knowledge required to correct for confounds of eye position in future pupillometry work [***Play It Again***] will reduce participant burden, in enabling future music paradigms to serve a dual purpose in dementia research:
(1) Distinguishing between syndromic groups via analyses of uncorrected data [***Play It Again (secondary analysis)***]; whilst
(2) Providing an opportunity to explore corrected data for possible further pupillometry effects that may have otherwise been masked or diminished in pre-corrected outputs.

Descriptions of the work on a variety of platforms, including podcasts [Music and Dementia (O'Hare, 2020)] has led, to-date, to two invitations consulting on notable dementia-music projects, aiming to deliver digital platforms for music and dementia resources/knowledge transfer: Musical Dementia Care Pathway (Music For Dementia, 2021), plus one other organisation which cannot be named for confidentiality reasons. An imminent goal to submit the work described to the peer-review process would subsequently result in potential dissemination of the work on these platforms, building on existing relationships with charitable and commercial music organisations currently exploring this area.

Longer-term impact of the work [**Play It Again; Our Dementia Choir**] involves providing nuance to debate in health and social care policy regarding social prescribing. The findings in the current work support the inclusion of physiological data capture within 'gold standard' randomised controlled trial (RCT) testing of music listening activities in PLWD. This approach may satisfy policymakers in terms of scientific rigour, whilst promoting the powerful yet transient effects that are observed during a musical experience, not just pre/post differences. This may help to move the narrative forward in terms of how researchers and policymakers frame the power of music, and its use as a person-centred, non-pharmacological, therapeutic complementary approach in dementia care and support.

RESEARCH PAPER DECLARATION

No publications arising directly from the work. Co-authorship on papers and contributions to work cited within the thesis are indicated by in-text references and reference list entries in bold.

ACKNOWLEDGEMENTS

CONTENTS

List of figures

List of tables

1 INTRODUCTION

This opening chapter:

   I.   Outlines the problem the current work seeks to address;

  II.   Defines and discusses key concepts and themes pertinent to the research, including:

- Music and dementia;
- Neuroanatomical correlates of music processing;
- Dementia subtype profiles of music processing;
- Music activity candidates suitable for empirical enquiry;
- The rationale for researching in-the-moment responses to music; and
- An overview of the autonomic nervous system and potential physiological candidates appropriate for dementia-music research.

1.1 THE PROBLEM

Despite the promise of disease-modifying therapies for dementia (van Dyck et al., 2023), any ongoing and future pharmacological therapeutic endeavours are likely to continue focusing on presymptomatic/early-stage clinical populations. This leaves an ever-increasing number of people **(~944,000 currently living with dementia in the United Kingdom (UK); >55 million worldwide, for whom symptom management and supporting wellbeing are likely to remain the principle objectives** (Luengo-Fernandez & Landeiro, in preparation; World Health Organisation, 2023). The evidence base for the use of arts in health (Fancourt & Finn, 2019) has underpinned the recent UK Department of Health and Social Care's commitment to **personalised care - including social prescribing -** benefitting four million people in the UK by March 2024 (Waitzman, 2022). Collectively, this signals a healthcare system moving clinicians towards promoting arts activities as a public health tool to manage certain symptomatic challenges and increase wellbeing in this clinical population (Chatterjee, Camic, Lockyer, & Thomson, 2018; Whitehouse, Vella Burrows, & Stephenson, 2018).

People living with a dementia (PLWD) deserve the opportunity to participate in meaningful activities; music provision is one such candidate, and of particular importance within dementia care settings (Harmer & Orrell, 2008). That said, whilst evidence for the relative sparing of music perception in PLWD makes it tempting to view it as a non-pharmacological 'silver bullet' in social prescribing terms, senior stakeholders within National Health Service (NHS) England have warned that, to correctly tailor its use in person-centred dementia care, music must be "carefully managed" (Burns & Morris, 2018). **Evidence demonstrates that the use of music in dementia care is not necessarily innocuous** (Playlist for Life, 2019). For example, a randomised crossover trial investigating the effects of Baroque music on behavioural disturbance episodes in PLWD reported a significantly higher number of behavioural disturbances exhibited during weeks in which Baroque music was played compared to control observation weeks (Nair et al., 2011). Failing to understand a person with dementia's relationship with music at a given moment, particularly at later disease stages when communication via language may not be viable, may in fact increase the chances of music initiating the exact adverse reactions it seeks to address.

A second problem relates to how the attractive mantra of 'music as medicine' influences scientific and political conscience; this inadvertently creates parameters for experimental designs in this area, most of which reflect a medical model approach to 'testing' the power of music in PLWD (Howland, 2016). Reticence to increase the promotion of creative health endeavours (including music) in UK social prescribing policy cited a lack of evidence of the "why and whether [they] work" and the use of terms such as "placebo effect", reflecting the desire to view music through a biomedical lens ("Amendment 114: Health and Care Bill - Report (3rd Day) – in the House of Lords at 4:23 pm on 7th March," 2022). In addition, **the current emphasis on long-term benefits of music fails to appreciate the impact of dementia's heterogeneity and progression** on music experience and enjoyment moment-to-moment. This can be changeable when compared to pre-morbid preferences (Fletcher, Downey, et al., 2015; van't Hooft et al., 2022). Arguably, the power of music in supporting emotional, cognitive and social functioning lies not in attempts to equate its characteristics to pharmacological offers (e.g. seeking out long-term benefit), but perhaps in better understanding its real-time impact on PLWD, and using this

information to harness music as a flexible therapeutic tool to reflect an ever-changing context. **'In-the-moment' experiences** can be highly salient to the individuals involved, but do not necessarily translate into lasting effects (MacPherson, Bird, Anderson, Davis, & Blair, 2009).

Currently, **relatively few dementia-music studies capitalise on the real-time objectivity and insight that physiological data capture has to offer**. Employing these techniques in different syndromic groups, alongside indices of disease severity and subjective reporting of music experience/wellbeing in early-mid stages of dementia, could have **application for using these techniques in more severely-affected dementia populations.** PLWD who are post-verbal are still able to engage with and enjoy music; consequently, they may in principle retain the ability to communicate their affective and cognitive experience of music in real time via these biosignals (Carrasco et al., 2020; Liu, Wang, & Yu, 2021). Some data collection techniques employed to index physiological responses to music (e.g. **discreet wearables and infrared pupillometry tracking) may lend themselves to testing people in later disease stages more than others** (e.g. wearables versus collecting saliva samples), **or facilitate physiological data collection in real-world settings**.

The work herein aimed to **build on previous research that captures physiological responses to recreational music activities in PLWD**, using passive recording techniques. Specific aims were:

a). To conduct a **pupillometry-music study** in healthy older adults to establish physiological indices of music familiarity and emotional processing in healthy ageing [*Chapter 3*];

b). To **extend this paradigm** to individuals living with early-mid typical and atypical Alzheimer's disease, comparing within- and between- syndromic and healthy control group response profiles, considering factors such as disease severity [*Chapter 4*];

c). To establish this paradigm's ability to **discriminate between typical and atypical Alzheimer's disease using machine-learning methods** [*Chapter 5*], and

d). To elucidate the physiological responses of PLWD participating in a real-world setting of **a dementia choir whilst wearing discreet wearable physiological recording devices** [*Chapter 6*].

1.2   KEY CONCEPTS

1.2.1   Music and dementia

Music is just one of the creative possibilities listed in the *Arts in Health Taxonomy for people with dementia* (Cousins, Tischler, Garabedian, & Dening, 2020). Yet it appears to have gained more attention than other art forms as a potent tool in supporting cognitive, emotional and social functioning in dementia populations (Tang & Vezeau, 2010). This is perhaps because, for people living with neurological disorders, music is preserved as a "*relative island in a sea of impairments*" (H. Gardner, 1985, p. 128). Countless examples of this phenomenon in the context of dementia are peppered throughout clinical folklore and single case studies [e.g. *Musicophilia* (Sacks, 2007); *Patient EN* (Cuddy & Duffin, 2005)] and readily available in the form of anecdotal evidence [*Power of Music report* (Music For Dementia, 2022)]. A review of randomised controlled trials (RCTs) exploring music-based interventions in neurological rehabilitation reported that the majority of music interventions administered in PLWD have been associated with a positive impact on a range of health outcomes, such as: reduction in neuropsychiatric symptoms and depression, and increases in overall cognitive performance (Sihvonen et al., 2017). As dementia is a neurodegenerative syndrome typically characterised by biological, emotional, and social loss, capitalising on this 'relative island' as a recourse has unsurprisingly captured the imagination of personal and professional caregivers, clinicians, music practitioners and academics alike (Benhamou & Warren, 2020; Edwards & Hodges, 2007).

1.2.2   Neuroanatomical correlates of music processing

The capacity to engage with music being to some extent retained - even in severe dementia stages - relies on the principle that music innervates a sufficiently complex co-opting of a number of bilateral cortical and subcortical networks, not all of which are wholly susceptible to the disease (Särkämö & Sihvonen, 2018). Neuroimaging and neuropsychological music listening studies indicate a complex model of music perception, each relating to distinct but extensive neuroanatomical correlates with broadly separable functions relating to core perceptual and associative music processing (Agustus et al., 2018).

As with any sound, music first engages the inner ear; the stimulus is encoded as a neural signal and transmitted via the auditory nerve to the auditory brain stem (Stewart, von Kriegstein, Warren, & Griffiths, 2006). Auditory stimulation has been associated with cortical tissue integrity of Heschl's gyrus, an area including primary auditory cortex, which is subject to neuroplasticity and bilateralisation of processing in individuals scoring highly on musicality tests (Agustus et al., 2018; Turker, Reiterer, Seither-Preisler, & Schneider, 2017). The first computational challenge upon hearing music is to distinguish it from the environmental auditory background (if present) or from environmental sound itself in an 'auditory scene analysis' (Goll et al., 2012). This involves: integration of spectral and temporal information [lateralised respectively to right and left early auditory cortex (Johnson, Marshall, et al., 2021)], to distinguish basic acoustic properties ([e.g. intensity, pitch, timbre, rhythm]; engaging superior temporal gyrus, posterior cingulate and posterior temporal cortices (Agustus et al., 2018); and processing in inferior and dorsolateral prefrontal cortex, cingulate cortex and inferior parietal cortex comprising attentional and working memory networks (Särkämö & Sihvonen, 2018). A secondary acoustic analysis takes place to enable more advanced feature discrimination [e.g. of melody, harmony/chords, and phrasing] (Janata, 2005).

In addition, higher-order associative music properties (e.g. music familiarity) are processed differentially; familiar melodies engage bilateral inferior frontal and anterior temporal cortex; novel melodies specifically invoke left parietal cortices as well as bilateral temporal regions (Agustus et al., 2018; Slattery et al., 2019). This network is observable in younger as well as older adults (Sikka, Cuddy, Johnsrude, & Vanstone, 2015). There is evidence that rhythm and familiarity processing involve similar cortical areas (e.g. supplementary motor areas [SMA], putamen and left planum temporale [PT]) (Griffiths & Warren, 2002; Jacobsen et al., 2015). The left PT acts as a 'hub' for musical temporal characteristic and familiarity integration (Griffiths & Warren, 2002), which perhaps accounts for the mediating effect of music familiarity on temporal judgement performance [i.e. whereby individuals are more likely to rate a familiar song as predictable in terms of temporal structure than unfamiliar songs, regardless of actual temporal predictability (Agustus et al., 2018)].

Emotional responses to music engage the (para)/limbic system (including amygdalae, hippocampi, and medial orbitofrontal cortex) via the thalamus [innervated at initial perceptual staging (Särkämö et al., 2012)]. Hedonic pleasurable responses additionally innervate reward-processing dopaminergic pathways, the release of which in striatal areas is associated with subjective pleasurable music experiences (Salimpoor, Benovoy, Larcher, Dagher, & Zatorre, 2011). Engaging with music beyond listening (e.g. singing) additionally invokes co-activation of both auditory perceptual [ventral (frontotemporal)] and vocal production [dorsal (frontoparietal)] pathways, comprising motor networks (e.g. cerebellum, basal ganglia, and premotor and motor cortices) and superior temporal cortices, inferior frontal regions and pre-/post central gyri; profiles of activation appear to be distinct to those observed in natural speech (Patel, 2013; Särkämö & Sihvonen, 2018). Engaging with music with others present (e.g. in a choir) induces additional socioemotional processing by virtue of being part of a group.

1.2.3 Syndromic group differences in music processing

Whilst the cortical networks as described above are extensive, it is important to note that distinct clinical and neurobiological profiles arising from different dementia aetiologies and syndromes cause changes to these networks that vary in extent and chronological order between clinical groups. This has implications for championing the capacity for music to act as a 'relative island' in the same way across the condition (Benhamou & Warren, 2020). Much of the work in dementia-music processing at the level of syndromic group has focused on comparing auditory and music processing abilities in frontotemporal dementia (an umbrella term describing impairments in socioemotional awareness (behavioural variant FTD [bvFTD]), language (nonfluent variant Primary Progressive Aphasia [nfvPPA]; logopenic variant PPA [lvPPA]) or sematic memory (semantic variant PPA [svPPA]). These conditions are particularly suited for empirical investigation of music by virtue of the prevalence of different profiles of musicophilia (a marked increase in musical interest and engagement) between phenotypes (van't Hooft et al., 2022). This phenomena has been observed particularly within genetic aetiologies (MAPT and C9orf72 mutations) and specific FTD subtypes [bvFTD]) relative to language-led dementias (e.g. PNFA) with mixed findings

in relation to musicophilia in SD (Fletcher, Downey, et al., 2015; van't Hooft et al., 2022).

Music processing capacities are often investigated in these conditions in the presence of a disease control group (individuals with typical - 'memory-led' - Alzheimer's disease [tAD]) alongside healthy age-matched controls. The investigation of individuals with tAD has yielded a number of findings of relative preservation of music perception and appraisal. Voxel-based morphometry comparing musical rule decoding and subjective pleasantness between individuals with tAD and individuals with an FTD phenotype revealed disturbances of both rule decoding and pleasantness ratings of music excerpts (bvFTD); individuals with SD and tAD demonstrated a double dissociation of impairment on a singular task [SD: impaired reward valuation; tAD: impaired rule decoding, preserved reward valuation] (C. N. Clark et al., 2018). Preserved reward valuation in the presence of tAD is encouraging for further work exploring the role of music in dementia care and support (C. N. Clark et al., 2018). Furthermore, cognitivist studies investigating performance on music playing, familiarity and recall tap into differential preservation profiles of musical memory domains. Case studies report preservation of music-making by virtue of relatively spared procedural memory (W. W. Beatty et al., 1988). There is strong group-level evidence to suggest preserved semantic musical memory in tAD, where individuals continue to perform comparably relative to healthy age-matched controls in recognising music as familiar or unfamiliar; this is to an extent differentiated by impairments on tAD performance of episodic musical recall (Slattery et al., 2019). Musical semantic memory is still considered to be retained even in the face of changes in dimension (e.g. pitch or timbre) or when an individual is unable to recall a previous experience of hearing that music (Groussard, Tyler, Copalle, & Platel, 2019).

Both findings of impaired and spared musical abilities have informed our understanding of neurotypical music processing and organisation, and in turn provided insight into the residual neurobiological profiles of each disease group. These elements are particularly informative for incorporating music into ongoing dementia care and support which considers the specific dementia diagnosis. The work herein describes two studies, which aim to build on this understanding. The first of these focuses on music responses in **two presentations of Alzheimer's disease and**

healthy age-matched controls (Chapters 3-5). The second explores responses to music in a group of individuals whereby the **proportion of different diagnoses broadly reflect generally-reported syndromic prevalence**. This approach reflects the rarity of real-world music provision in care settings being targeted to particular diagnoses, considering instead the dementia condition as a whole (Chapter 6).

1.2.4 Candidate music activities for dementia research

Music elicits a range of behavioural responses in individuals with neurological disorders, which have been collated in the Therapeutic Music Capacities Model (TMCM) (Brancatisano, Baird, & Thompson, 2020). The model summarises how researchers and music practitioners may use music as a way of tapping into different processes that can give rise to a PLWD's affect, cognition and behaviour. This framework contextualises the work described herein (see Figure 1.1).

Music-based interventions, outlined in the TMCM as the intermediary between formal *music therapy (MT)* and *everyday use of music*, are an attractive focus for dementia-music research. This type of music activity, often taken up by, with and for PLWD for recreational purposes, sits outside a formal MT context, but typically elicit high levels of engagement (Särkämö et al., 2016b) relative to everyday use of music (e.g. having the radio on in the background) (Brancatisano et al., 2020). Benefits of recreational music-based activities not only relate to their cost-effectiveness relative to MT delivered by a formally trained facilitator, but also demonstrate - in their own right - an ability to support PLWDs' emotional, cognitive, and social functioning (Raglio, Pavlic, & Bellandi, 2018). These types of musical activities are not intended to be therapeutic per se, but are often mislabelled as such, given their capacity to give rise to moments of enjoyment and engagement linked to health-related goals (Stapleton, 2020). As such, these activities have been termed as "music medicine" (Sihvonen et al., 2017, p. 649).

The two research studies within the current work investigate two distinct recreational music-based interventions that fall neatly within two contexts outlined in the TMCM: **music listening** [*individual; passive*] and **recreational choral singing** [*group; active*].
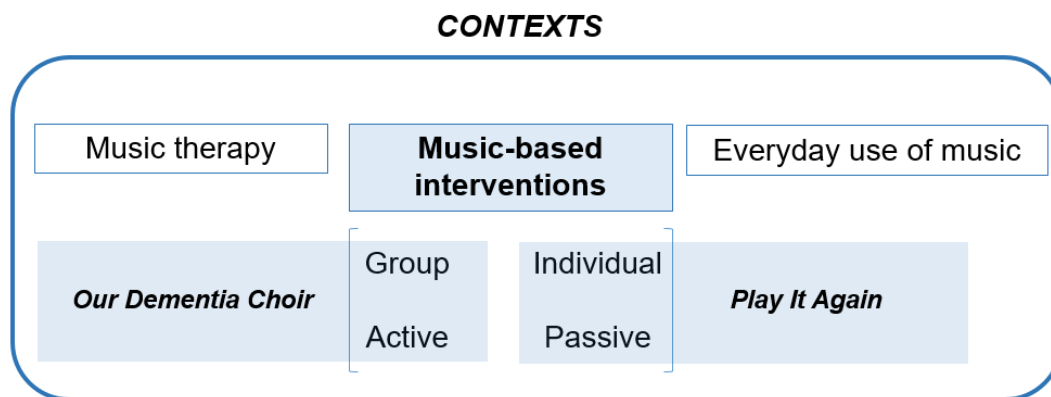
CONTEXTS

Figure 1.1 Contexts outlined within the Therapeutic Music Capacities Model (TMCM); figure adapted from Brancatisano et al. (2020) to reflect the current work described in the thesis (recreational activities highlighted in blue, and labelled accordingly with studies relating to the respective activities: ***Play It Again [individual; passive]*** and ***Our Dementia Choir [group; active]***).

Music listening [individual, passive]

Music listening for PLWD can be either self- and/or caregiver-implemented (E. Johnston, Rasmusson, Foyil, Shopland, & Wang, 2017), delivered via either live performance, or recorded music played through speakers or headphones on a range of devices, in both home [e.g., as investigated in the HOMESIDE study (Baker et al., 2019)] and residential care or clinical locations. While the support of a trained music therapist to deliver music listening interventions is ideal to elicit targeted therapeutic benefits (Raglio, Bai, et al., 2022), funding provision often dictates that music listening for dementia care and support is delivered by healthcare professionals or informal/family caregivers. Due to the ubiquitous implementation of this activity for PLWD, it is an important candidate for study in order to learn more about the processing in PLWD that underlies this type of music experience.

Furthermore, music listening is the one activity which remains accessible throughout the trajectory of the dementia condition, as indicated by the Musical Dementia Care Pathway, inspired by the NHS Dementia Well Pathway (**Music For Dementia, 2021**). The level of music interaction recommendations differ

between early to severe stages. An emphasis on engagement with more active forms of music-making and self-selected music listening is outlined during mild-moderate disease stages. This however changes in the severe disease stages to the promotion of relatively passive music activities (e.g. listening) whereby music selection is more likely to be directed by another person in response to functional change, and ease of delivery in palliative care settings (Bolton, Jiang, & Warren, 2022). The first study described in this thesis explores music listening, focusing on recorded musical excerpts [**Play It Again**, Chapters 3-5].

Studying recorded music listening provides the opportunity to isolate responses to an activity that exclusively comprises musical engagement, relative to the multicomponent nature of a formal MT intervention or live music presentation (Clare & Camic, 2020). MT and performance not only is the skilled deployment of music, but in addition the elicitation of vocal, body and facial expressions of the therapists/performers towards the individual or audience, creating essentially an audiovisual stimulation (Hsu, Flowerdew, Parker, Fachner, & Odell-Miller, 2015). Studying MT interventions therefore invoke the challenge of ascertaining to what extent any impact or benefit are being observed due to the music, the interaction, or the combination of these elements. The multicomponent nature of MT interventions may also explain the mixed results and range of effect sizes that have plagued music therapeutic RCTs and reviews, leading to mixed results regarding the role of music in dementia care and support (Bourne, Camic, & Crutch, 2021; van der Steen et al., 2018).

Choral singing [group; active]

A comprehensive snapshot of choral singing approximated that ~2.14 million people are members of one of the estimated 40,000 choirs in the UK (Voices Now, 2017). Particularly during the early stages of the disease, opportunities for PLWD to participate in a choir are not limited to provisions by dementia-related organisations and charities (Alzheimer's Society, n.d.). However, dementia-aware choirs or choirs whose membership comprise only PLWD may provide the appropriate levels of specialist support to facilitate a person's continued engagement as their disease progresses (O'Neill, 2019). Taking part

in a choir of course engages elements of a music listening activity, however the group and active music-making components give rise to an experience that is perhaps greater than the sum of its parts (Cummings, 2015). Due to the popularity of recreational choral singing, and the take-up of dementia choirs in multiple countries ("Dementia Choir format seals international deals," n.d.) this second music activity is worthy of further investigation. The naturalistic setting of choral investigations addresses in turn the criticisms regarding the generalisability of laboratory-based work. The second study described in this thesis explores participation of PLWD in a dementia choir [**Our Dementia Choir**, Chapter 6].

1.2.5 *The rationale for researching in-the-moment responses*

Investigating responses in PLWD to music activities have alluded to "joyful moments" (couched primarily in the context of MT; see McDermott, Crellin, Ridder, and Orrell (2013), p. 791). The work described herein attempts to address the tendency to overlook what is going on 'in-the-moment' for an individual with dementia as they engage with recreational music activities (Nair, Browne, Marley, & Heim, 2013). There are a number of good reasons that dementia-music research to-date has traditionally focused on pre-morbid or prospective timeframes. Firstly, the phrase "*music medicine*" is evocative of an intervention likely to deliver long-term health benefits; as a result, experimental designs in dementia-music research have favoured the measurement of effects over relatively long periods, with mixed success (van der Steen et al., 2018). Secondly, it is common practice for music practitioners, researchers and clinicians to focus on individuals' past music preferences to select music that could continue to play a role in their current emotional and cognitive functioning. This is a central tenet of music reminiscence therapy, and actively encouraged in protocols and guidelines for incorporating music in dementia care (Garrido, Dunne, Stevens, Chang, & Perz, 2019; Gerdner, 2021). Protocols which promote the use of familiar music alone however have a limitation in that they either rely on self-report from the person with dementia to communicate their preferences for a playlist, or observant relatives and carers to offer suggestions for its contents [e.g. using the Assessment of Personal Music Preference (Family Version) (Gerdner, Hartsock, & Buckwalter, 2000)]. For PLWD in the later stages of the disease, a lack of communication via language can arise from

a combination of biological, social and psychological factors (Norberg, Melin, & Asplund, 1986). Self-reporting new and/or perhaps unexpected responses to music may not therefore always be possible for PLWD, leading to subtle non-verbal signals of engagement with music becoming obscured or more challenging to observe **(Clare et al., 2020).** Complementary to traditional therapeutic efforts is an approach to determining music selection with and for PLWD which is engaging and personally relevant to the individual's desired interaction with music *in that particular moment*; providing environmental stimulation appropriate for their here and now (Dowlen et al., 2021).

Extending the conceptual framework of in-the-moment research to study responses to music in dementia populations has a number of benefits (Keady et al., 2020). Measuring in-the-moment responses advantageously addresses the following challenges in dementia-music research, in that: 1). With or without dementia, one's experience of music can change over the lifecourse (Bonneville-Roussy, Rentfrow, Xu, & Potter, 2013); 2). In some diagnoses, change in musical preference or use can in fact be a canonical facet of the disease itself (van't Hooft et al., 2022); and 3). In restricting current musical experience solely to a person's past, PLWD are denied the opportunity to engage with new musical genres or songs that may not have been pre-morbidly known, or are newly released. This is important when considering that an increased feeling of familiarity can be experienced by people living with tAD in relation to repeated presentations of new songs (Samson, Dellacherie, & Platel, 2009). Using in-the-moment measurement techniques to provide insight into a PLWD's response to music at any given timepoint could inform healthcare professionals' and practitioners' understanding of how best to incorporate music into a person's care and support despite these contextual challenges.

A secondary benefit to exploring in-the-moment measurement is that (to some extent at least) it has the potential to future-proof research in the dementia-music field. A diversification of music listening taste and preference has been aided by streaming services - the most popular means of engaging with music last year (International Federation of the Phonographic Industry, 2022). Ubiquitous access to music inherently means that, while researchers might continue to examine likely pre-morbid musical preferences [e.g. music released in a person's adolescence or 'reminiscence bump'

(Rao, Peatfield, McAdam, Nunn, & Georgieva, 2021)] to aid experimental design and stimulus selection, music experienced within individuals' adolescent timeframes is unlikely to continue to be confined to music released within that time period. As a result, both the pool of music available over the lifecourse - and by virtue the challenge of predicting a person's response to music - is only likely to increase. Methods ascertaining a person's musical preferences must therefore be able to accommodate a much wider range of individualised music taste than we have observed in previous generations. Harnessing data collection techniques now, that enable us to establish profiles of in-the-moment responses in older and clinical populations, creates a sustainable approach in dementia-music research, so that we may continue to explore our understanding of music experiences for dementia populations for years to come **(Brotherhood et al., 2017).**

1.2.6   Psychophysiological responses

The objectivity of the response measures used in the majority of in-the-moment studies in severe dementia to date [within-session field notes, practitioner-rated scales and researcher-rated video analysis (Weise, Jakob, Topfer, & Wilz, 2018)], can be questioned. The use of psychophysiological and endocrinological markers in dementia-music research offers an opportunity to gather data not only in-the-moment, but which are a). objective; b). attainable directly from PLWD (versus relying on proxy estimations); c). ungovernable by conscious control (the obverse to this resulting in potential bias within self-reported ratings; d). lower-cost and burden relative to neuroimaging studies, the inclusion criteria for which can be stringent and obtrusive to participation, and d). selectable to reflect the differential and wide-ranging psychological processing that underlies a musical experience (Bonifacio & Zamboni, 2016; Sakamoto, Ando, & Tsutou, 2013; Särkämö, 2018).

The Autonomic Nervous System (ANS)

Relatively dynamic signals, provided by periphery nervous system responses, reflect the proximity of the 'bodymind' relationship and provide a pragmatic approach to understanding in-the-moment responses to music in PLWD in a relatively objective and low-burden way (Sequeira, Hot, Silvert, & Delplanque, 2009). Before outlining

particular psychophysiological candidates, it is helpful to consider how the biology of the peripheral nervous system gives rise to these observable responses, which in part form the motor output of our experiences.

Central Nervous System components (the brain and spinal cord) receive and send signals to the peripheral nervous system, which in turn comprises the Autonomic Nervous and Somatic Nervous Systems (ANS; SNS) (see Figure 1.2). The ANS is characterised by the nerve systems that regulate tissue and organ function in the body, with the exception of striated muscles, separated into three systems. For the purpose of the current work, focus is directed towards two of these: the *sympathetic* and *parasympathetic* systems. Stimulus-evoked fluctuations in *sympathetic* autonomic responses, and differentiation within these, can indicate internal higher mental processing, such as a range of emotional and cognitive responses (Öhman, Hamm, & Hugdahl, 2000). In contrast, *parasympathetic* nervous system responses relate to processes involved in homeostasis, such as rest and digestion (Sequeira et al., 2009).

Within both the sympathetic and parasympathetic systems, *preganglionic neurons* (the cell bodies of which are located in the spinal cord [sympathetic, parasympathetic] and brain stem [parasympathetic]) send chemical information from the brain/spinal cord via axons to the ganglia, establishing synapses on *(post-)ganglionic neuronal* cell bodies and dendrites (Jänig, 2006). These exist outside the CNS, and in turn transmit information to target organs, which tend to receive information from one or both sympathetic and/or parasympathetic systems (see Table 1.1) (Jänig, 2006). This information is sent and received without conscious control or awareness, thus giving rise to involuntary covert responses, which can be observed even in the absence of subjective differences in response to stimuli. This has been observed in numerous User Experience (UX) experiments studying autonomic responses to a range of images used in advertisements (Romano Bergstrom, Duda, Hawkins, & McGill, 2014).

As denoted in Table 1.1 and Figure 1.2, the sympathetic and parasympathetic nervous systems have interacting effects on the majority of psychophysiological measures (e.g. pupil size, heart rate (HR) and related properties such as heart rate variability (HRV)). This increases the complexity of both selecting physiological candidates for further study in this field, and data interpretation. Jänig (2006) nevertheless cites this

33

Table 1.1 Location of para- and sympathetic preganglionic and postganglionic neurons in regions of interest for current work, adapted from Jänig (2006).

Organ	Sympathetic system		Parasympathetic system	
	Preganglionic neurons	Postganglionic neurons	Preganglionic neurons	Postganglionic neurons
Eye (pupil)	T1-2	Superior cervical ganglion	Edinger-Westphal nucleus	Ciliary ganglion
Heart	T1-4	Stellate ganglion, upper thoracic ganglia (superior and middle cervical ganglion)	Nucleus ambiguus	Cardiac plexus
Skin surface (upper extremity)	T1-5	Stellate upper thoracic ganglia (spinal nerve C5-C8 projection)	-	-

complexity as a "misconception…more the exception than the rule" (p. 28). While the effects of these two systems on each target organ is microscopically discrete, linking sympathetic and parasympathetic activation to neuroanatomical correlates helps to reconcile this. Broadly, the ANS engages both hypothalamic and extra-hypothalamic nuclei, with selective regulation of sympathetic and parasympathetic outflow (Dougherty, 2020). A review of 43 neuroimaging studies demonstrated functional and neuroanatomical specificity reflecting the division of ANS labour which related to autonomic outputs in two categories (sympathetic activity-skin conductance links, and parasympathetic-high-frequency/HRV links) (Beissner, Meissner, Bar, & Napadow, 2013). Using activation likelihood estimation to conduct a voxel-based meta-analysis, the authors demonstrated clustering of executive- and salience-processing networks associated with sympathetic activation, and default mode network (DMN) associations
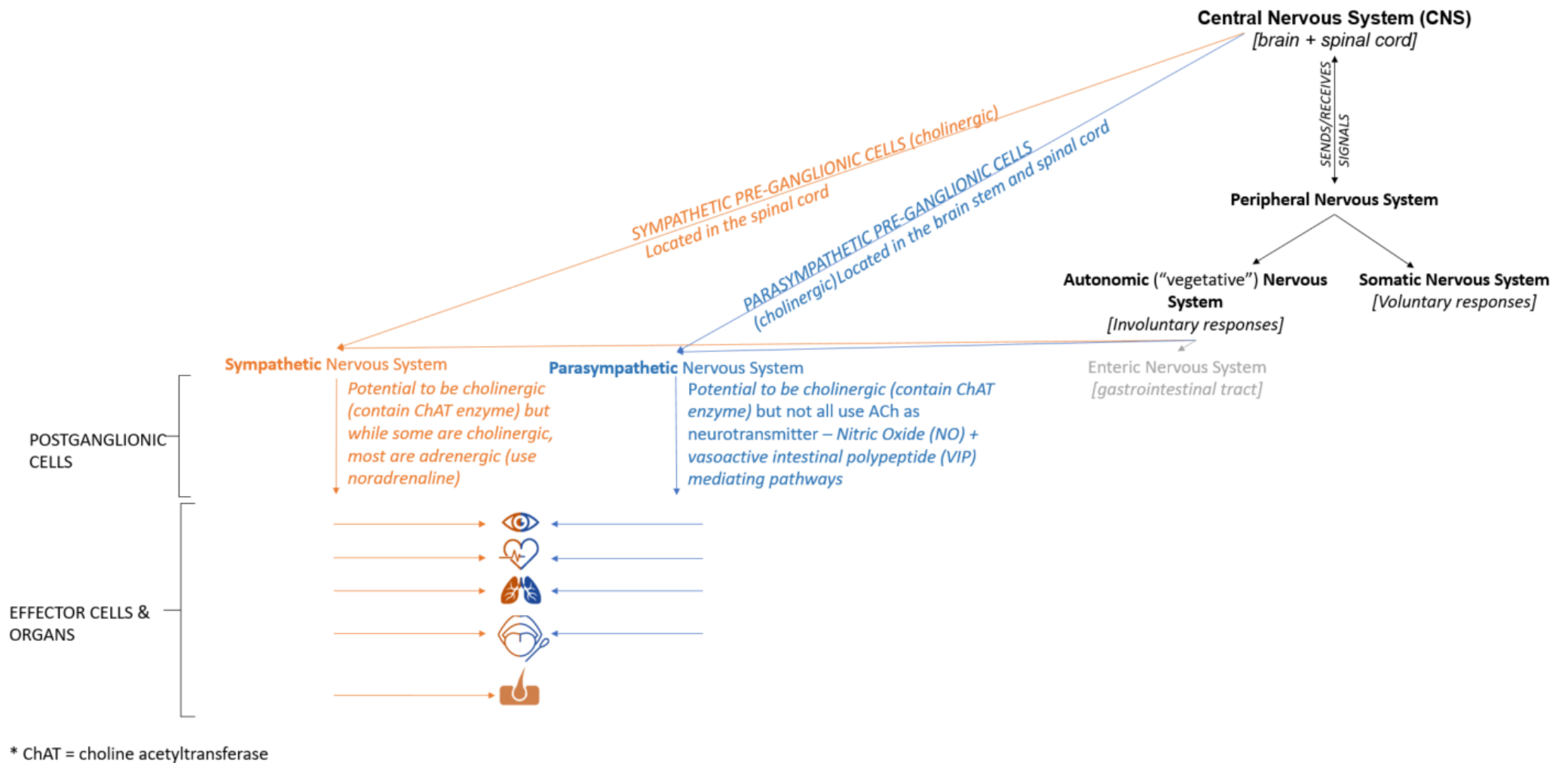
Figure 1.2 Schematic representation of the division of the CNS and pathways to target organs discussed in the current work.

with parasympathetic activity (Beissner et al., 2013). The widespread involvement of multiple networks indicates similar neuroanatomical structures that may relate to both preserved music processing in the dementias, and the capacity to elicit physiological emotion- and cognitive-based changes in recordable indices of sympathetic nervous system activity.

Practical factors for considering physiological candidates

Measures of neuroendocrinological responses in PLWD have been the most widely-used approach to-date in capturing autonomic responses within dementia-related (Sittler, Worschech, Wilz, Fellgiebel, & Wuttke-Linnemann, 2021); and neurotypical (Fancourt, Ockelford, & Belai, 2014) music research. Another approach has been the study of blood biomarkers in response to music-based interventions in PLWD (Kurita et al., 2006; Okada et al., 2009). These studies have given rise to a number of observed differences in salivary properties (namely cortisol, chromogranin and immunoglobulin A) in response to music, the decrease of which during and post-music interventions in PLWD has been proposed to indicate a reduction in stress response.

Findings from these studies indicate the efficacy in saliva sampling in indicating a global increase/decrease in relevant properties in response to a music intervention in its entirety. However, its sampling rate and timeframe of measurement in the experimental design stages increases the potential for missed opportunities in capturing transient physiological responses to music, simply by virtue of lacking the ability to provide a continuous measurement. For instance, significant reductions in cortisol pre-and post-individual interventions were observed at the halfway-point (6th session) and at the final session (12th session) in one study, but no overall significant reductions in cortisol were observed pre- vs. post- overall intervention, despite the positive experience of the participants (Chu et al., 2014). These findings illustrate the importance of investigating each individual session, and adds further support to continuous physiological measurements, thus reducing the risk of overlooking notable - albeit transient - responses. A review considering endocrinological responses to music in PLWD also cited that, due to these factors, together with the relatively slow response rate of change of salivary properties in response to stimuli, these measures

may not be the optimal biosignal for elucidating the impact of music activities on PLWD at the level of detail required for the current work (Thomas, Crutch, & Camic, 2018).

In addition, almost all investigations outlined above commented on the challenge of collecting blood plasma and/or salivary samples from PLWD. In neuroendocrinological studies, this was most commonly cited as being due to participant resistance to the collection procedure (mostly cotton swabs), but also in difficulties in obtaining sufficient levels of saliva secretion for data analysis (Suzuki et al., 2004). A feasibility study investigating HRV and saliva data acquisition in PLWD throughout choral singing and art viewing activities raised doubts regarding the feasibility of collecting saliva samples in this population (Bourne et al., 2019). The authors encountered difficulties in subsequent physiological data analysis owing to insufficient sample acquisition; to an extent by virtue of the increased external room temperature (30°C/86°F) which influenced participant hydration levels (Bourne et al., 2019). Okada et al. (2009) similarly cited that adequate blood plasma cytokine and catecholamine levels were only attainable in a subset of their participants with dementia. These concerns, in addition to inconsistent results when investigating salivary properties such as chromogranin A (CgA) (Suzuki, Kanamori, Nagasawa, Tokiko, & Takayuki, 2007; Suzuki et al., 2004; Valdiglesias et al., 2017) support the need for more discreet and reliable physiological data capture in the current research design. Furthermore, motor demands placed upon individuals to collect saliva samples may not be tolerable for people in later stages of dementia where symptomatic motor challenges arise, and/or in particular dementia syndromes (e.g., in individuals with Posterior Cortical Atrophy).

Passively-recordable biosignals

Noting the limitations above, a re-focus of the methods employed to establish in-the-moment physiological responses to music in PLWD may be timely. Passive recording devices have been shown to be tolerated in PLWD and can provide a continuous time-locked measurement of momentary music responses in PLWD in a relatively unobtrusive way (Au-Yeung et al., 2022; Nesbitt, Gupta, Maly, Okhravi, & Jain, 2019). The first step involves establishing the current knowledge base for physiological responses to music in PLWD. The full range of measures and experimental designs

employed to date to capture passively-recorded physiological responses to music in PLWD are detailed in the literature review (see Chapter 2).

2 PASSIVELY-RECORDED PHYSIOLOGICAL RESPONSES TO MUSIC IN PEOPLE WITH DEMENTIA: A NARRATIVE SYNTHESIS REVIEW

This chapter:

I. Summarises the current understanding of physiological responses to music in PLWD using passive recording techniques;

II. Outlines the main overarching research objective that the current work seeks to achieve;

III. Concludes with a contextualisation of the current work in light of the literature review findings.

2.1 INTRODUCTION

Given the numerous benefits of music for individuals living with a dementia, it is important to explore possible mechanisms by which clinicians and arts practitioners could extend the role of music to supporting those in the later stages of the disease. It may be possible to, for example, use "non-standard' responses, such as non-verbal [facial/body] expressions to determine a person's capacity to access semantic musical memory (Omar, Hailstone, & Warren, 2012, p. 475). Psychophysiological responses to music in PLWD, acquired using passive recording techniques, may serve as one avenue in acquiring such non-standard responses, particularly in individuals who are in later stages of the disease, and/or post-verbal.

One important counter to this proposal, however, is the evidence of autonomic degradation in individuals with dementia, calling this endeavour into question. A review of evidence indicated that autonomic dysfunction may occur only in relation to particular syndromic groups (e.g., individuals with Lewy body dementia (LBD) or Parkinson's disease dementia (PDD) [para- and sympathetic function]) or isolated autonomic processes (e.g. orthostatic hypotension in tAD) (Allan et al., 2007). More widespread autonomic dysfunction in Alzheimer's disease (AD) is however arguably likely owing to its hallmark cholinergic dysfunction (intrinsically linked to both para-

and sympathetic processing) (Femminella et al., 2014). Femminella et al. (2014) also highlight that Braak's (1991) staging of AD was specifically modelled on increased degradation of two brain structures known to be implicated in autonomic processing (the brain stem and insular; also implicated in feature-based and emotional processing of music) (Särkämö & Sihvonen, 2018; Sihvonen et al., 2016). Autonomic dysfunction is apparent within the locus coeruleus (LC) in individuals with tAD, which is a brainstem nucleus specifically implicated in autonomic physiological processing (e.g., pupil psychosensory responses) (David & Malhotra, 2022). Given these findings, it is critical to review our current understanding of what has been reported to-date in the dementia-music-physiology field. This will help to establish whether the approaches outlined in this thesis are likely to yield effects in this clinical population, and indicate potential appropriate biosignal candidates which warrant further investigation.

This review aims to outline the current landscape of passively-collected physiological responses to music in PLWD and subsequent autonomic findings, to inform the current work and to verify, where evidence exists, of preservation of such responses in PLWD relative to healthy control participants. The review seeks to build upon previous dementia-music reviews, many of which sought to include only RCTs. The RCT paradigm, whilst considered the methodological 'gold standard', tends to focus on pre- and post- comparisons of music interventions, thus "the lights are cut precisely when the really interesting stuff takes place, the 'during' period when engagement with music happens" (DeNora & Ansdell, 2014, p. 4). Reviewing the current dementia-music-physiological literature which sits within or on the boundaries of this "power cut" may further our understanding of the underlying mechanisms that may be driving how music plays a role in dementia care and support (DeNora & Ansdell, 2014, p. 4).

Two existing reviews formed the basis of the current search. Thomas et al. (2018) reviewed 13 papers relating to physiological responses to the arts in individuals living with a dementia, reporting that all 'arts' conditions in physiological studies involved music to some extent. Physiological responses collected were broadly categorised into cardiovascular or endocrinological studies. A second review helpfully surveyed the 'what' and 'how' across the varying methodological approaches to dementia-music research (Dowson, McDermott, & Schneider, 2019). Eleven studies were identified in

this review which incorporated a physiological element. In line with the review scope, specific findings were not discussed in great detail beyond a summary depicting each related outcome measure (e.g. gait, nutrition). The outcome measure summary did however highlight a useful factor for consideration: the definition of 'physiological response' in the current context being focused more on autonomic nervous system responses (e.g. as in Thomas et al., 2018) as opposed to encompassing functional changes which perhaps more reflect Activities of Daily Living (ADL) outcomes (Dowson et al., 2019). The author is also aware of more recent papers published subsequent to these reviews which explore a wider range of physiological responses pertinent to the current work. In addition to combining the original search terms from these two reviews for the purposes of the current work, it was necessary to: (a) expand the search to incorporate other dementia syndromes, physiological candidates, and music terms; (b) remove references likely to return studies using more invasive physiological data collection (e.g. salivary and plasma studies) reflecting the current focus; and (c) incorporate more recent publications to reflect the recent increase in interest in this area.

The systematic review adopted a 'narrative synthesis' approach to explore the relevant studies available (Popay, 2006). Previous relevant review discussions highlighting lack of reported effect sizes suggest that a statistical meta-analysis would not be achievable. This view builds on the approach taken in Dowson et al. (2019) in providing a textual identification of similarities and differences between the relevant publications to facilitate further work in this area. Where statistical interrogation via meta-analyses is prevented by virtue of study heterogeneity, the narrative synthesis approach is often perceived as the 'next-best' option in these circumstances, and in addition provides a comprehensive overview of the field for policymaker as well as academic audiences (Popay, 2006). Figure 2.1 sets out the context for the work.

2.1.1 Aims

a) Collate and summarise passive physiological data collection approaches and findings in dementia-music studies;

b) Discuss coupling with emotional and/or cognitive changes as interpreted by study authors;

c) Ascertain the nature and timing of passive physiological data capture in the current literature;

d) Report any challenges or recommendations around approaches to physiological data capture in PLWD, in terms of practicalities or harnessing technological solutions for objective measurement tools; and

e) Establish the level of detail included in studies regarding the choice of music activity selected, and discuss the extent to which critiques of other reviews of music and dementia research are applicable in this particular corpus of research studies.

2.2   METHODS

The search returned scientific articles published between 1960 and 28 November 2022. Searches were performed in PubMed (utilising MeSH terms), APA PsychINFO, EMBASE and Scopus electronic databases. The search strategy was based on a combination of terms included in recent literature reviews investigating the impact of creative arts on PLWD (Dowson et al., 2019; Thomas et al., 2018; Sittler et al., 2021). Final search terms were developed and refined following a number of trial searches.

The search terms are as follows, with variations on the following search strategy adapted as appropriate for each database:

exp Dementia/ or (Dementia or Dementia's or Dementias or Dementias' or Demented or Alzheimer* or Alzheimer or Alzheimer's or Alzheimer`s or Alzheimers or Posterior Cortical Atrophy).ti,ab. AND exp Music/ or Singing/ or (Music or "Arts" or sing or Song or Songs or Choir or Choral or Playlist or Playlists or Listen or Listening or Musical* or Vocal).ti,ab. AND (Psychophysiolog* or bioma* or biol* or Psychobio* or Physiolog* or Heart Rate or Heartrate or Pupil or Pupillometry or Eye-tracking or Eyetracking or Eye tracking or Galvanic Skin Response or Electrodermal or Arousal or pulse or autonomic).ti,ab.
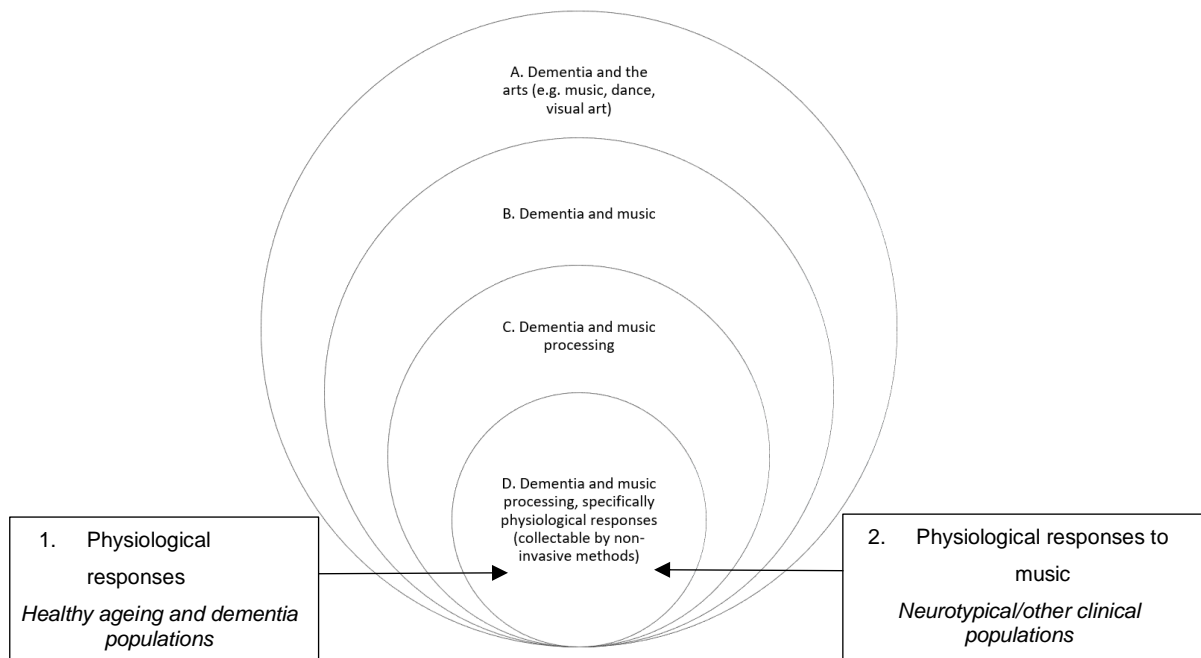
Figure 2.1 Stacked Venn diagram indicating the wider context for dementia-music-physiological studies (A-C), and to highlight the other areas of research that inform it (1-2).

Articles were included in the final review subject to meeting the following criteria: any quantitative or mixed-methods study communicated in English, involving PLWD, citing a physiological response as an outcome measure, and published in a peer-reviewed journal. Studies were required to include music as a standalone activity (e.g. singing) that could be differentiated from other modalities, thus excluding any physiological studies exclusively investigating multimedia presentation (Imtiaz, Anwar, & Khan, 2020). Exclusion criteria for the review was the obverse. A number of studies classified by Dowson et al. (2019) as physiological music investigations in PLWD were, on closer inspection, reporting measures acting as potential indirect markers of emotional/cognitive responses or functional changes (e.g. nutritional intake) (Wong, Burford, Wyles, Mundy, & Sainsbury, 2008). These physiological markers are also arguably observable without invoking the use of sensors and recording devices. In the subsequent review, papers or findings were therefore excluded from subsequent reporting or discussion whereby the physiological outcome was either (a) related to a functional change or (b) could be observed or coded on a global level via observation without recording equipment. Facial electromyographical studies and indices of movement (e.g. accelerometer data) were retained in the current review owing to the ability to record even negligible changes

which are less likely to be observable to the naked eye watching video recordings, and which readily generate a quantitative value to these changes in behaviour.

Combining these terms returned a total number of 480 papers (Scopus (n=60); Embase (n=216); PubMed (n=120) and PsycINFO (n=84). Review papers were excluded from the final selection, but checked for references and any relevant missing literature was added (N = 6). This led to a total initial number of 486 articles. Results were consolidated using EndNote. Duplicates were automatically detected and removed (N = 63), followed by a manual check and removal of further remaining duplicates (N = 89). Eighteen articles were selected for final inclusion in the review (see Figure 2.2).

2.3   FINDINGS

2.3.1 Summary of results

Sample characteristics and methods employed by the 18 selected articles are summarised in Appendix A; main physiological findings are outlined in Appendix B. Eleven of the 18 selected articles exclusively investigated physiological responses to recorded music listening. Five exclusively focused on music therapy sessions or interactive music sessions where there was either an implicit or explicit opportunity for the participants to engage with music- making (e.g. singing, clapping). Two studies incorporated both music formats, in order to compare the two music presentations albeit via different experimental designs [within-subject multiple-case study **(Walker et al., 2021)**; between-groups (Sakamoto et al., 2013)]. One study compared a music experience (recreational choral singing) with a non-musical activity [art-viewing (Bourne et al., 2019)]. Two studies sought to explore differences between syndromic groups (Benhamou et al., 2021; Garrido, Stevens, Chang, Dunne, & Perz, 2018) whereas other studies treated all clinical participants as one, perhaps owing to inadequate numbers in each syndromic group to warrant between-group analyses. Studies were mixed in terms of physiological data capture timings; studies either recorded data continuously, segmented small timeframes pertinent to the research objective (e.g., Benhamou et al., 2021), reported overall mean responses reflecting whole-session data capture, or adopted the more traditional approach of collecting physiological data pre- and post-music session/programme.
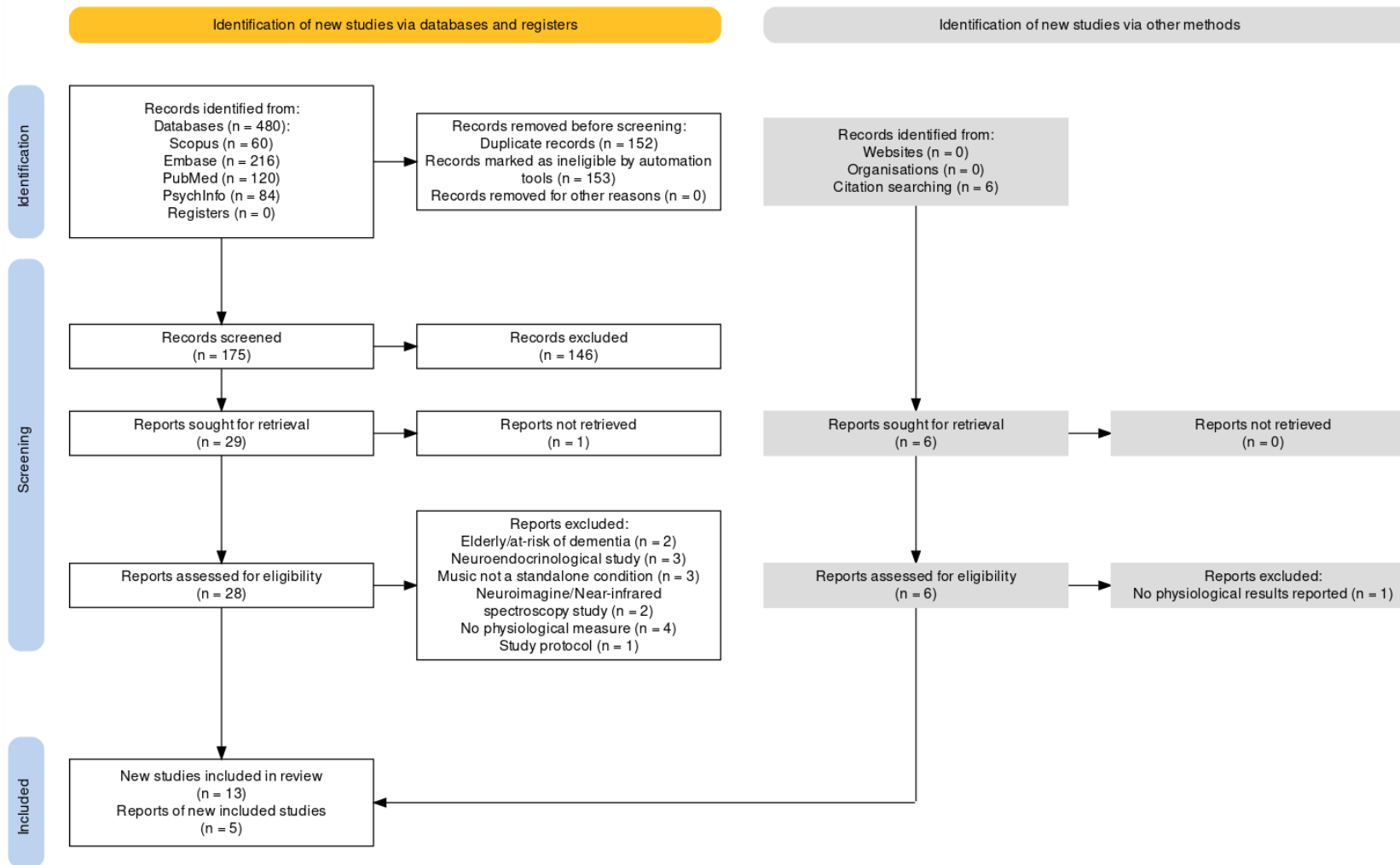
Figure 2.2 Flowchart adapted from the preferred reporting items for systematic reviews and meta-analyses (PRISMA) reporting guidelines for literature review (Haddaway, Page, Pritchard, & McGuinness, 2022).

Candidates for passively-collected physiological responses to music in PLWD could be broadly categorised into: pupillometry, cardiovascular (heart rate and heart rate variability; blood pressure) and pulse oximetry, electrodermal, skin temperature, movement, and facial electromyographical responses. The prevalence of each physiological candidate (where captured) broadly reflects the frequency of published reports on psychophysiological responses to music as outlined in Hodges (2010) with the exception of respiration studies, which was less popular in the current review selection (see Figure 2.3). Measurements such as accelerometer and facial electromyography were included in the current review but were not documented in Hodges (2010).
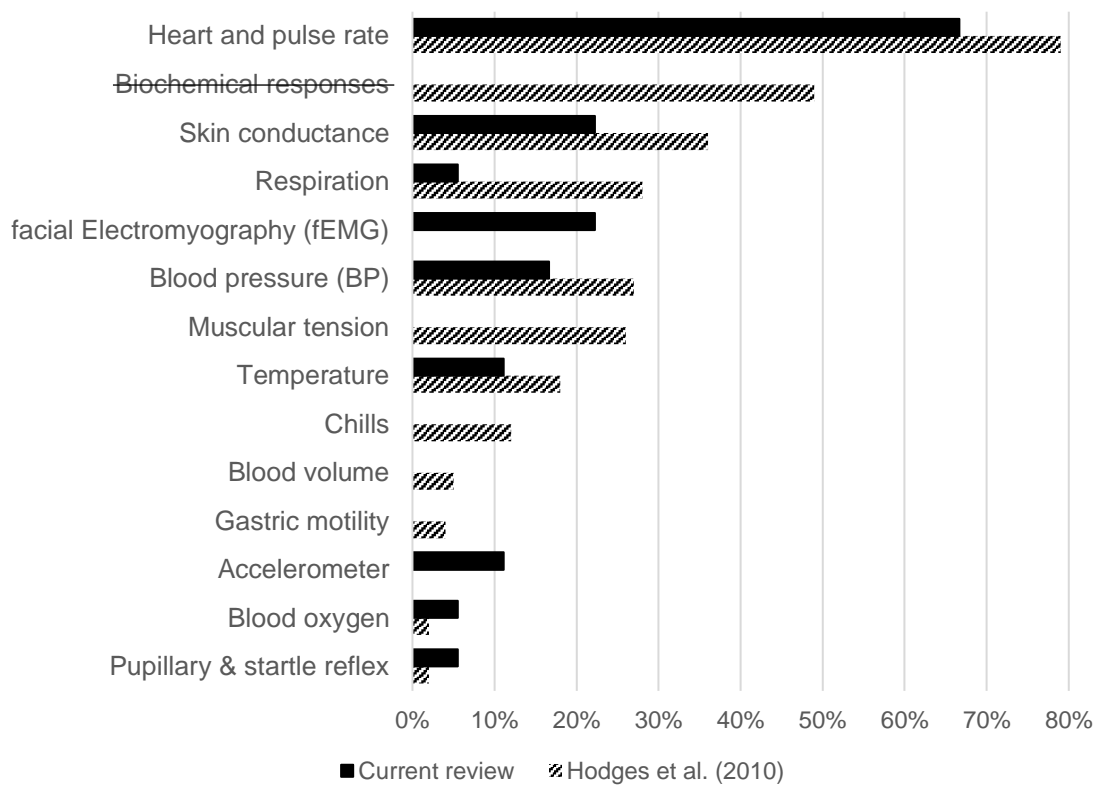
Figure 2.3 Proportion of passively-recordable physiological measures employed in the 18 articles within the current review (shaded black), relative to the prevalence of research using each measurement in the general music-physiology research field according to Hodges (2010, patterned). The strike through 'biochemical responses' reflects that these responses were excluded in the current review. Accelerometer and facial electromyography data were included in the current review, but did not feature in Hodges (2010).

Musical response findings in PLWD within each candidate physiological biosignal are summarised below.

2.3.2 Psychophysiological measurements

2.3.2.1 Pupillometry

Only one study to date has investigated pupillometry responses to music in PLWD (Benhamou et al., 2021). This study aimed to record pupillometry responses to songs manipulated in terms of 'expectation' (by systematically changing acoustic, syntactic or semantic properties). The study compared detection of and pupillometry responses to these manipulations in people living with frontotemporal dementia (including PPA subtypes), typical Alzheimer's disease and healthy older controls. Individuals with FTD and PPA were less sensitive to detecting deviant musical properties than controls, and in some cases than individuals with typical AD. There was no correlation with disease severity in auditory deviant discriminability across the patient groups. Maximum overall pupil responses in each patient group did not differ to healthy controls. Differential pupillary responses were observed in each group (acoustic deviant > syntactic deviant > semantic deviant > no deviant). A surprise value and level was placed on each deviant, and this positively correlated with pupil size in the healthy control, tAD and nfvPPA group; the strength of the correlation did not differ between tAD individuals and healthy controls. Splitting the songs into high/low entropy groups, and correlating the surprise value with pupil size in each, healthy controls' pupil responses correlated positively in both groups; tAD pupil responses correlated positively with low entropy songs only, and nfvPPA pupil responses correlated positively with high entropy songs only. The findings indicate that, at least in typical Alzheimer's disease, that pupillometry responses to and cognitive appraisal of - at least - musical expectancy, is broadly aligned with profiles of psychophysiological behaviour observed in control populations (Benhamou et al., 2021). This is additionally encouraging considering a). the tAD group in this study was significantly older than the control group, and b). there was no correlation with pupillary responses and age overall. This study indicates that investigating pupillometry responses to music, at least in people with typical Alzheimer's disease, may yield responses that couple with self-reported cognitive and emotional appraisals of music.

These findings contrast with other previous reports of observed group differences between individuals with tADs' autonomic processing of other auditory stimuli sets (namely salient sounds, whereby pupil response was shown to be reduced in tAD participants relative to healthy controls, despite no group differences in self-reported response) (Fletcher, Nicholas, et al., 2015a). Moreover, when investigating the relationship between pupil responses and valence ratings for environmental (non-verbal) sounds in individuals with bvFTD, SD, PNFA, AD and controls, overall autonomic reactivity to sound was found to be normal in the tAD group, however coupling between pupillary and affective behavioural responses was altered in this group relative to controls (Fletcher, Nicholas, et al., 2015b). It may be that these autonomic responses in tAD versus controls may be particular to environmental sounds rather than music. It is important for the current work however to note that no studies attempting to link individuals with tADs' pupillometry responses with music pleasantness ratings have yet been undertaken.

2.3.2.2 Cardiovascular responses and pulse oximetry

The largest corpus of passively-recorded psychophysiological responses to music in PLWD relate to cardiovascular responses, reflecting trends in the wider literature (Figure 2.3) (Hodges, 2010). Experimental design has largely comprised pre- and post- overall music listening programme comparisons of heart rate (HR), yielding mixed results. Significantly ($p<0.001$) lowered HR was reported in a case study of a patient with FTD at Session 20 of a music therapy session relative to baseline (Session 1) (Ridder & Aldridge, 2005). A similar finding was observed whereby a relative decrease in HR in 10 PLWD was reported following 24 individualised music sessions (Maseda et al., 2018). Listening to concert music did not however elicit any changes in HR when comparing first and last (8th session) responses (da Rocha, Siqueira, Grella, & Gratão, 2022) nor over the course of a 12-week Turkish music listening randomised controlled trial intervention (Gök Ugur, Orak, Yaman Aktas, Enginyurt, & Saglambilen, 2019). In some instances whereby HR did not yield significant findings, significant post-session decreases in systolic and diastolic blood pressure (SBP; DBP) were observed in PLWD in the music intervention relative to a non-music control group (Gök Ugur et al., 2019; Takahashi & Matsushita, 2006). Within-group analyses of the MT intervention group in Takahashi & Matsushita (2006) however revealed significant

post-session increases in blood pressure (BP) relative to pre-session responses collected at first session, 6 months, 1 year and 2 years after the start of MT. BP has also been shown to be significantly lower in PLWD responses to classical music listening relative to popular music in the absence of HR effects (Corrêa, Caparrol, Martins, Pavarini, & Gratão, 2020), however this biosignal does not always guarantee effects in the absence of HR findings (da Rocha et al., 2022). Studies comparing PLWD responses to type of music intervention (passive and interactive group) reported significant post-session reductions in mean HR relative to pre-session levels, an effect absent in a control (no music) group (Sakamoto et al., 2013). The authors discussed the decreases in HR indexing parasympathetic activation innervated by the passive and interactive music exposure, reflecting "reduction of stress" (p. 782). Interestingly, between-group music activity comparisons (interactive/passive) revealed no significant differences in cardiovascular responses (Sakamoto et al., 2013).

Investigating cardiovascular responses while a music activity takes place indicates support for differential in-the-moment responses in PLWD that may reflect a musical experience. Pulse increase responses were differentially observed (Tune 2 *[Patient A]*; Tune 5 *[Patient B]*) in individual participants (Norberg et al., 1986). The authors discuss the physiological response to Tune 5 in patient B coupling with the observable response "the patient intensively listening inwardly" (Norberg et al., 1986, p. 477). Within-song reductions in cardiovascular responses (HR) were observed relative to baseline in four out of six PLWD (Walker et al., 2021 [Study 2]). Mixed results regarding relative HR changes from baseline were reported when PLWD listened to music of differing tempo (Walker et al., 2021 [Study 2]). This may be attributable to a number of different factors (e.g. participant diagnosis/individual differences in song appraisals) rather than HR changes specifically in relation to an objective musical characteristic. Mixed findings may also be explained by HR responses being in-part mediated by a range of sources (e.g. hemodynamic, humoral and neural [para- and sympathetic] pathways) (Sequeira et al., 2009). While HR responses have been shown to couple with affective image valence and arousal ratings in patients with Alzheimer's disease and healthy controls (Balconi et al., 2015), responses are therefore likely to reflect a multitude of underlying processes rather than pinpoint changes in psychological processing alone.

Heart rate variability (HRV) facilitates the compartmentalisation of parasympathetic and sympathetic innervations in an attempt to more confidently form conclusions as to the impact of music on PLWD. Within-music-therapy sympathetic activity (low frequency: high frequency ratios [LF:HF]) decreases compared to baseline and post-session recordings have been recorded in PLWD (Kurita et al., 2006; Okada et al., 2009). These differences did not reach statistical significance in either study, but yielded significant increases in cardiovascular parasympathetic activity within-music-session relative to baseline in individuals with mild-moderate (Kurita et al., 2006) and severe (Okada et al., 2009) cerebrovascular disease (CVD) and dementia. The control condition (non-MT) did not reveal any cardiovascular change (Okada et al., 2009). Between-acoustic-activity comparisons (hearing a personal memory, a news story, a favourite song and unfamiliar song) did not elicit differential HR responses in five PLWD in the severe disease stages (Sun, Baird, Gelding, de Wit, & Thompson, 2021). However, between-activity (music/visual arts) investigations have revealed increases in PLWD cardiovascular responses during a choral session relative to baseline, but no effects in the same group when participating in an art-viewing session (Bourne et al, 2019). The inclusion of a caregiver control group in Bourne et al. (2019) revealed that the differential within-choir cardiovascular responses were exclusive to the PLWD group.

2.3.2.3 Electrodermal activity

Autonomic responses indexed by electrodermal activity (EDA) are described in one study meeting review criteria. EDA data were acquired during four different songs in a Portuguese sample of 20 people with possible/probable AD, and 20 healthy age-matched controls (Barradas, Juslin, & Bermúdez i Badia, 2021). Each song was selected from previous studies to elicit a different psychological mechanism innervating an emotional response: 'brain stem' (a song inducing a startle reflex), 'emotional contagion' (a song previously validated to induce sadness), 'episodic memory' (a song commonly used in Portuguese weddings) and 'musical expectancy' (a song with low key clarity, and which did not confirm to harmonically conceived tonality) (Juslin et al., 2015; Juslin et al., 2013). A main effect of mechanism was reported for the electrodermal activity, but no interaction with group (Barradas et al., 2021). This encouragingly indicates that the presence

50

of possible or probable AD did not significantly impact the autonomic response to the different songs. This finding corroborates previous AD-control comparisons of electrodermal activity during a sustained attention to response task, and in relation to the emotional appraisal of affective images (Irish et al., 2006; A. Joshi et al., 2014)

Investigation of EDA responses in PLWD indicate, to some extent, differential responses depending on participant characteristics and musical properties. Measuring galvanic skin response (GSR) to music of different tempo in 99 individuals with dementia revealed GSR increases in response to both slow and fast-tempo music, and an interaction of time (music vs. baseline) and enjoyment group (high/low) in the fast-tempo condition (Garrido, Stevens, Chang, Dunne, & Perz, 2019). Walker et al. (2021) reported consistently higher EDA responses during the first song relative to the baseline EDA output in a community music setting [*Study 1*], and during a music listening session [*Study 2*]. In-session vs. baseline EDA responses in two interactive music sessions [*Study 2*] were more mixed, however the majority [N = 4/6] demonstrated an increase in-session compared with baseline. The authors interpreted this as indication of increased pleasure; indeed, increased electrodermal activity has been demonstrated in response to happy music in neurotypical adults (Lundqvist, Carlsson, & Juslin, 2009; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009).

The authors continued to observe mixed EDA responses to fast and slow tempo songs in eight participants, with five eliciting significantly higher EDA responses during slower-paced music, and the obverse demonstrated in the remaining three participants (Walker et al., 2021). Interestingly, schematic representation of the data also indicated that the largest EDA response for all participants in all sessions was in relation to a familiar melody, however this was not statistically interrogated (Walker et al., 2021). These findings support the proposal that perhaps subjective appraisals of music (e.g. a person's familiarity with the song) may be driving differential observed effects in physiological (EDA/HR) responses relative to objective musical properties.

Between-activity analyses in Walker et al. (2021) indicated that the majority of participant EDA responses were significantly higher in the interactive music sessions compared with the control (music listening) session. These EDA findings contrast however with the HRV findings in Sakamoto et al. (2013) and with observations from the EDA findings in Sun et al. (2021), whereby no effect of condition was observed. This may be attributable to the severity of the participants' condition relative to the populations tested in the other studies. The authors also cite methodological limitations as potential explanatory factors; the lack of EDA findings may have resulted from the fact no stable baseline was ascertained (unlike Barradas et al., or Walker et al., both 2021), and stimulus presentation order was not controlled (Sun et al., 2021). Noting the small sample size in Sun et al. (2021), and given EDA effects have been observed in PLWD elsewhere (Alam, Roy, Holmes, Gangopadhyay, & Galik, 2016; Melander, Kikhia, Olsson, Walivaara, & Savenstedt, 2018) it seems plausible to suggest that the findings are more likely to reflect the experimental design rather than warranting no further exploration of EDA responses in a more severely-affected population.

2.3.2.4 Skin temperature responses

Sun et al. (2021) found that, when measuring electrodermal, cardiovascular and skin temperature responses in five non-responsive people living with severe dementia, individuals' skin temperature was the only physiological metric which showed a differential response to music versus two other auditory conditions (personal memory and new story). However, the two music conditions (favourite/unfamiliar song) did not elicit differential responses. The authors interpreted an increase in skin temperature on the wrist as a reduction in arousal, by virtue of increased blood flow to peripheral parts of the body. This proposal is supported by previous music studies (H. Lai, 2004) and to some extent in Walker et al. (2021), which demonstrated higher ST responses in the majority of participants (N = 4/6) during music listening compared to baseline. While this increased physiological response is encouraging to observe in people living with severe dementia, the findings in Sun et al. (2021) in particular must be interpreted with caution in the absence of a stable baseline condition and no controls in place for order effects (i.e. that skin temperature simply increased over time). In addition, comparing baseline ST responses to ST within the first presented song during either

a music community setting [*Study 1*] or two sessions of a music programme [*Study* 2] revealed mixed findings, comprising either an increase, decrease, or no change in skin temperature (Walker et al., 2021).

A more consistent result was reported in Walker et al. (2021) when comparing skin temperature responses to fast and slow tempo songs; six out of nine participants' skin temperature was significantly higher for fast tempo music than slow tempo music in the music community setting, whereas only one participant showed the obverse ([*Study 1*]). This was not replicated in a second study whereby skin temperature responses tended to be lower for fast music compared to slow tempo music, more so in the music listening condition than interactive music session (Walker et al., 2021 [*Study 2*]). The authors concluded that in the interactive sessions, musical tempo may play less of a role in physiological responses owing to other variables that, by virtue of the real-world setting, were challenging to control. These findings, however, may add support to the proposal that individual differences in subjective experiences may be driving physiological within-music responses, relative to their acoustic properties.

2.3.2.5 Movement/accelerometer

Walker et al. (2021) was the only study to incorporate movement metrics into their analysis, by virtue of its addition to the Empatica® E4 suite of multistrand physiological data collection. The authors reported increased movement during the first intervention song relative to baseline, the interpretation of which was proposed to reflect either increased engagement, similar to the EDA profiles observed, or a reduction in apathy and/or depressive symptoms (Walker et al., 2021). The collection of movement mirroring EDA responses as identified in Walker et al. (2021) may indicate a potential confounding variable to account for when analysing EDA data; this and other potential confounding factors have been raised in EDA-related publication guidance by Boucsein et al. (2012). There is however, an argument to say that movement may be a biosignal candidate in its own right. Movement and music have always been intrinsically linked, as evidenced in the neuroanatomical correlates in music production and perception as outlined in Section 1.2.2; indeed, other recent work has successfully predicted dancers' musical preference from their musical-induced movement when

moving freely to 30-second long excerpts (Agrawal, Carlson, Toiviainen, & Alluri, 2022).

2.3.2.6 Facial electromyography (fEMG) and computer-assisted fEMG

Facial electromyography (fEMG) refers to the recording of muscular micromovements in the face, indexing a change in emotional state. Barradas et al., (2021) reported an interaction of group (AD/age-matched controls) and mechanism (song eliciting either a startle, episodic memory, emotional contagion, or musical expectancy) on fEMG zygomaticus response. Dementia participants elicited less activity in response to the 'brain stem' song compared with controls, and a greater response compared to controls when listening to songs evoking 'emotional contagion' and the 'expectancy' condition. The authors concluded that the fEMG responses were valence-sensitive, in contrast to EDA responses that reflected changes in autonomic arousal.

This review has highlighted a movement towards objective coding of task-evoked changes in video-recorded responses. Of the six studies collecting concurrent video recordings alongside physiological data (Barradas et al., 2021; Garrido et al., 2018; Garrido, Stevens, et al., 2019; Liu et al., 2021; Ridder & Aldridge, 2005; Sakamoto et al., 2013) four studies subjected the video data to computer-assisted facial electromyography software (e.g., FaceReader). The computer software package detects subtle changes in facial expressions in video data, and categorises the extent to which a person is exhibiting an expression aligned with one of the Ekman emotions (Happiness, Sadness, Anger, Fear, Disgust, Surprised, Contempt) in comparison with a Neutral baseline on a continuum (Ekman, 1999). This novel approach to analysing facial movements in PLWD is non-intrusive, and therefore preferable to traditional electromyography data capture, which involves attaching small electrodes to the skin. Adopting this approach has revealed subtle changes in evoked response to music activities in PLWD, and may be useful to incorporate into the analysis of video data collected in future research.

2.3.3 Findings relating to experimental design and data collection techniques

2.3.3.1 Interpreting findings as sympathetic/parasympathetic responses

The articles included in the review reference increases and decreases of physiological responses, and propose interpretations as to whether these index (para)/sympathetic activation, and what changes in emotional states this may be reflecting in the participants. Kurita et al. (2006) and Okada et al. (2009) both reported trends of LF:HF decrease during MT compared with before and after (an index of sympathetic activity), but notably these indices did not reach statistical significance in either study. The lack of sympathetic change, compared with other HRV metrics indexing parasympathetic activity, was concluded in both studies to be a result of music innervating the parasympathetic nervous system and attenuating the sympathetic nervous system, by virtue of the musical properties (e.g. tempo and rhythm) eliciting feelings of relaxation and/or 'distraction' reactions within respective limbic and hypothalamic systems. Nevertheless, it was noted in Kurita et al. (2006) the difficulty in establishing sympathetic activation using the LF:HR ratio due to the potential contamination of respiration on these measurements. The EDA studies were much clearer in their interpretation of sympathetic innervation, by virtue of this measurement not being contaminated by parasympathetic activity (see Table 1.1).

2.3.3.2 Coupling of physiological response to self-report

Some of these findings indicate a potential challenge coupling physiological responses with self-report in dementia groups. Of particular interest to the current work are the findings of association between FaceReader valence output and PLWD's subjective pleasure, arousal and acoustic comfort ratings (Liu et al., 2021). While the use of FaceReader is relatively new in this field, these findings indicate the potential for video-recorded data to act as a proxy for phenomenological music experiences in PLWD. Furthermore, while Barradas et al. (2021) reported no significant interaction of *mechanism* and *group* (AD/controls) in electrodermal response, there was a reported interaction of the two factors in four out of five scale responses (*happiness-elation, nostalgia-longing, anxiety-nervousness*, and *sadness-melancholy*). This indicates the potential for a de-coupling of physiological responses to self-report as a function of disease; this was demonstrated to some extent in Benhamou et al. (2021), albeit from

a more cognitivist approach in terms of asking individuals to identify musical properties as opposed to 'felt' appraisals of the musical stimuli.

2.3.3.3 Later-stage dementia studies

Encouragingly, differential physiological responses were observable even in later stage dementia (Norberg et al., 1986; Ridder & Aldridge, 2005; Kurita et al., 2006; Okada et al. 2009; Sakamoto et al., 2013). Kurita et al. (2006) in particular remarked on the notable change in HRV during MT demonstrated in an 89 year-old woman with dementia, whose ADLs were all undertaken by nursing staff. Sakamoto, et al. (2013) demonstrated parasympathetic responses to passive and interactive music activities in people living with severe dementia, with a larger sample size than in previous studies (N = 39 [13 in each experimental and control group]), indicating previous reports were not demonstrating spurious results. It is important to note that most demonstrated results in severe dementia reflected parasympathetic responses. Sympathetic activity measured using a variety of approaches (e.g. LF:HR ratio in HRV; (Kurita et al. (2006) and Okada et al. (2009) and in some cases Walker et al. (2021)) yielded no significant differences between baseline and experimental conditions. Walker et al. (2021) did however demonstrate significantly increased HR and EDA in six out of eight participants with severe dementia in response to the first song of a live music session in comparison to baseline readings. Behavioural observations indicated that pleasure was predicted by dementia severity in Garrido et al (2018), but a potential link between dementia severity and physiological responses was not explored.

2.4 DISCUSSION

The findings from the articles included in the current review demonstrated physiological responses to music in PLWD. The studies reflect a heterogeneity within the field, in terms of: range of physiological markers, timings of physiological data capture, syndromic group (where interrogated), disease severity, as well as musical context and level of engagement (from passive listening to music-making opportunities). A common thread and the focus of this review was to consolidate music response findings acquired in PLWD via use of passive data recording techniques; the

prevalence of different techniques reflected previous psychophysiological work in this area (Hodges, 2010). Some novel approaches were identified as part of this work, not only via the capture of accelerometer data (albeit in only one study described), but an introduction of facial electromyography, the automation of which shows promise for further investigations (Garrido et al., 2018; Garrido, Stevens, et al., 2019). In particular, this review indicates that pupillometry and EDA responses appear to be primary candidates for further investigation of music experience-biosignal coupling in PLWD.

The review has provided reassurance around the general principle of investigating physiological responses to music in PLWD, and identified notable gaps in our knowledge. Differential psychophysiological responses to music reflecting syndromic group differences have been formally explored very little, probably owing to small sample sizes of different diagnoses in the current review. Precise descriptions of types and stages of dementia were specifically recommended by four studies (Corrêa et al., 2020; Garrido et al., 2018; Garrido, Stevens, et al., 2019; Ridder & Aldridge, 2005). It should be mentioned that Garrido et al. (2018) did attempt to delineate the relatively large sample size into different diagnoses, but conceded that this was only achieved by examining mini mental state examination (MMSE) performance and attributing drop in visual performance relative to memory domains as indicative of other non-AD dementia (e.g., LBD). Whilst the principle of exploring syndromic group differences is informative, ideally experimental designs looking to compare syndromic groups should seek to establish diagnoses through more formal methods (Baird & Samson, 2015; Benhamou et al., 2021).

A call for further description ascribed to participant background, clinical and musical stimuli was recommended, owing to potential influences on psychophysiological responses (e.g. mood, medication [Kurita et al., 2006; Okada et al. 2009; Garrido et al., 2018]; music experience [Walker et al., 2021]; intervention content and approach [Ridder & Aldridge, 2005]). Work in this area still appears to be susceptible to similar concerns raised within wider-range dementia-music methodological reviews, particularly regarding lack of participant and stimulus characterisation (Clare & Camic, 2020; Fancourt et al., 2014). Whilst challenging to describe or implement music therapy interventions which can be repeated, the take-up of relatively simple music paradigms in experimental design (e.g. music listening) - and providing a

comprehensive overview of the procedures involved - are advantageous for empirical work.

Spiro (2010) posited that, "*The popularity of music…with its increasing resource investment, invites and requires the development of the understanding of the mechanisms involved in the perception…of music by individuals with dementia and its effect on physiological and neurological processes*" (p. 891). Reviewing the available literature indicates that attempts to move measuring in-the-moment responses to music beyond the realms of clinical folklore and into systematic empirical enquiry in dementia research has proven promising, but not without methodological challenge. The main criticisms in current literature reviews (namely relating to music therapy research) called out a general lack of methodological rigour in or description of investigations, and in some cases this has been reflected in the current review, making it difficult to ascertain the specific processes driving physiological change (Vink & Hanser, 2018). Various issues such as: a) small sample size and/or poorly characterised participant groups; b) quality of data design and collection; and c) gaps in definitions of conditions and paradigms employed, mask the possibility for replicating results (Fancourt et al., 2014; Sihvonen et al., 2017). Optimistically speaking, the findings reported in the review do however indicate that it is justifiable to further explore physiological responses to music in PLWD.

Notably, changes in physiological responses were mostly reported as an isolated modality, or in relation to objective musical properties such as tempo, genre, or musical predictability. It is arguably important to explore, where possible, a potential physiological-phenomenological coupling of a musical experience. Links between the physiological and phenomenological musical experience should be explored within PLWD who remain able to communicate the latter via self-report. Should these physiological responses be demonstrable in individuals with early-mid stage disease, a longer-term application of the work may be to seek evidence for relatively preserved profiles of autonomic responses to music in people living with later stages of the disease, who are no longer able to communicate their responses via language. Clare et al. (2020) highlight, "nonverbal communication in later-stage dementia may be overlooked or underestimated by busy care staff and families" (p. 1115). An ability to capitalise on and demonstrate these subtle non-verbal communicative endeavours via

passive physiological data collection could justify continued musical provision to individuals living with later-stage dementias. Passive physiological recording techniques also overcome some ethical issues of data collection in people with later stages of dementia, by virtue of being lower-risk relative to invasive techniques. Whether physiological responses to music continue to be observable in real-time in syndromic groups may provide moment-to-moment insight into how music is experienced over the progressive disease trajectory.

The overall objective of the current research is therefore to establish profiles of physiological responses in PLWD in relation to (a) their subjective appraisal of music [Play It Again] and (b) wellbeing/stress scores [Our Dementia Choir], using passive data capture techniques.

2.4.1 The current work in the context of the literature review findings

The methodological approaches, study findings and researcher recommendations from the literature review has heavily informed the study design of the research outlined herein. A typical approach taken in research in the wider dementia-music field to date comprises the collection of standardised scale and subjective self-report data, to garner a holistic understanding of responses in PLWD to a range of musical stimuli (Tang & Vezeau, 2010). The current work seeks to maintain this approach, whilst adding in-the-moment physiological data capture to enable statistical modelling which can establish potential music-based physiological-phenomenological associations in PLWD. The different settings for the two studies included in the thesis reflect two distinct pieces of work, which were not designed to be directly compared. As a reflective exercise, it is however perhaps helpful to outline their positions on a continuum, and highlight in the relevant chapters the implications of each for the studies' designs and techniques employed within to acquire physiological data (see Figure 2.4).

EXPERIMENT   *EXPERIENCE*

1. Play It Again *[2. secondary analysis]*   3. Our Dementia Choir

Figure 2.4 A representation of the *experiment - experience* continuum. The first investigation, entitled *Play It Again* is a laboratory-based *experiment* (Chapters 3-5). The second research study outlined in this report, *Our Dementia Choir*, explores the physiological and subjective responses to a real-world participatory *experience* of a 12-week choral programme for PLWD (Chapter 6).

3   PLAY IT AGAIN: A MUSIC-PUPILLOMETRY STUDY IN HEALTHY OLDER ADULTS

3.1   ABSTRACT

This chapter outlines the first empirical study within the thesis. The study aimed to elucidate the extent to which healthy older adults elicited differential pupil responses to musical excerpts rated in terms of varying pleasantness and familiarity. Healthy older adults (N=90) were recruited opportunistically via departmental research databases/Join Dementia Research. Pupillometry responses were recorded while participants listened to 30 musical excerpts, each subjectively rated for pleasantness, and indices of familiarity [(1) *a priori* predicted (known/unknown); (2) self-rated by participants in terms of degree (not familiar-extremely familiar); and (3) depth (whether the excerpt evoked a memory)]. Background, basic health and music experience information was collected. Separate linear regression models obtaining cluster robust standard errors were run to predict log-transformed maximum trialwise pupil response from self-rated pleasantness, *a priori* predicted and subjective ratings of music familiarity, controlling for demographic factors including age and gender. Statistical modelling revealed evidence for: a curvilinear relationship between pleasantness ratings and pupil response, associations of larger pupil response with predicted and self-rated familiar music relative to unknown stimuli, alongside evidence for associations of decreased pupil response with age and increased pupil response in male participants. The results are discussed in relation to the preliminary finding of preserved differential pupillometry responses to pleasantness and familiarity appraisals of music in healthy older adults. While observed associations between pupil size and age raise questions in terms of the utility of pupillometry in older adults, this study serves as a useful proof-of-principle for pupillometry to act as a non-invasive biosignal for establishing in-the-moment response to music in healthy older adults, warranting extension to clinical populations.

3.2   INTRODUCTION

3.2.1   Emotional responses to music

Music listening is a readily-available pastime which many of us engage with on a regular basis. A UK survey revealed that 74% of adults identify music as important to their quality of life [*This Is Music report (UK Music, 2021)*]. Global music listening rose by 10% between 2021 and 2022, reflecting the ever-increasing accessibility of music; in a recent survey, 80% of 55-64 year olds stated they believe there are more ways to access music than ever before [*Engaging with music report* (International Federation of the Phonographic Industry, 2022)]. Music is created to elicit a response from the listener; semi-structured interviews conducted with older adults indicated that music listening was an emotional recourse for many during the Covid-19 pandemic (Groarke, MacCormac, McKenna-Plumley, & Graham-Wisener, 2022). While physical manifestations of emotional responses to music are described even in common parlance (e.g., "chills"), elucidating the psychophysiological relationship to music remains complex (de Fleurian & Pearce, 2021). This is, in part, due to the fact that emotional appraisals of music are difficult to predict based on acoustic properties, relying to a greater extent on an interaction between the music, listener, and situation (Juslin & Laukka, 2004). The complexity increases when we consider that, in addition to individual differences in emotional responses to music, an individual may react differently to a particular song depending on the situation or context in which they find themselves (Juslin, Liljeström, Laukka, Västfjäll, & Lundqvist, 2011). Capitalising on physiological measurements of responses to music may elucidate the underlying musical experience for an individual at any particular moment.

3.2.2   Selecting candidate biomarkers to measure in-the-moment responses in PLWD

Autonomic arousal is mediated by the sympathetic nervous system, but not contaminated by parasympathetic nervous activity, making it easier to disentangle in terms of the types of psychophysiological response elicited (Critchley, 2002). Autonomic arousal responses can be relatively isolated using a number of physiological marker candidates, including changes in *pupil dilation* (Mathôt, 2018). Psychophysiological measures of autonomic arousal, acquiring response data through passive recording devices, may therefore offer a reliable approximation of in-the-moment affective and cognitive responses to music in PLWD.

3.2.3 The pupil

While our pupils respond to light and stimuli near-fixation, they also dilate in cognitive and emotional response (either to stimuli, or spontaneously), known as the *psychosensory pupil response [PPR]* (Mathôt, 2018; Szabadi, 2018). These responses, which occurs in the absence of changes in illumination, are a recognised peripheral marker of autonomic arousal (Mathôt, 2018). PPRs are mediated by norepinephrine (NE) and acetylcholine (ACh), contributing to the innervation of independent musculature within the eye (DiNuzzo et al., 2019):

Sympathetic 'arousal' pupil activation (task-evoked *psychosensory pupil dilation responses [PDR]*) occur via the subcortical pathway (S. Joshi, Li, Kalwani, & Gold, 2016). Starting at the hypothalamus, activation innervates the release of NE fired from the brain's main source of norepinephrine (Locus Coeruleus; LC), the activity of which correlates with pupil size as demonstrated in functional neuroimaging studies (S. Joshi et al., 2016). Impulses from the LC are transmitted to the spinal cord (intermediolateral column; IML), then outside the spinal cord (the superior cervical ganglion) which, via a complex nerve network, connect to the radial dilator pupillae muscle (Mathôt, 2018). LC-innervated pupil responses are relatively slow to occur (500-1000ms in animal studies) (Larsen & Waters, 2018). There is both limited evidence elucidating the functional role of the PDR; as Steinhauer states, "Nobody really knows for sure what these changes do" (Fong, 2012). Furthermore, understanding of the microscopic contributions of its underlying neural pathway, owing to the involvement of the LC and hypothalamus, reflects a wide range of functional processing (Mathôt, 2018). Animal models have associated sympathetic pupil dilation with activation in the dorsomedial prefrontal cortex, which includes the supplementary motor area (SMA), frontal eye fields (FEF) and supplementary eye fields (SEF) (Claron et al., 2022).

Parasympathetic activation (pupil constriction mechanism) is mediated by the circular sphincter pupillae muscle by virtue of ACh (acetylcholine) release (Gingras, Marin, Puig-Waldmuller, & Fitch, 2015; Szabadi, 2018). ACh also plays a role in pupil dilation by its release, inhibiting the brain stem's oculomotor nucleus, which in turn relaxes the sphincter pupillae muscles resulting in increasing pupil response (dilation).

As we can see from the descriptions outlined above, interactions between these two pathways by virtue of involvement of similar neural areas (e.g. LC both activating the sympathetic pathway and inhibiting the EWN which reduces parasympathetic activity) give rise to pupil dilation, albeit by different mechanisms. This highlights the complexity in interpreting the underlying processes that give rise to PPRs. Jänig (2006) however specifically uses the pupillary response as an example of how this claim is often overstated, proposing that when this antagonism between sympathetic and parasympathetic systems occurs, it is synergistic in nature, or under different functional condition. Supporting the latter interpretation, Gingras et al. (2015) cite that the complexity of the experimental paradigm itself is a key factor to interpreting the likely mechanisms underlying a pupil dilation response. For example, a cognitively complex task is likely to lead to a pupil dilation response by virtue of increased cognitive 'load' or processing to complete the task; this response is mediated by the parasympathetic system inhibition, thus leading to pupil dilation (Steinhauer, Siegle, Condray, & Pless, 2004). However, in low-cognitive load tasks (e.g. passive listening paradigms as one such example) PPRs elicited are more likely to reflect emotional or cognitive responses innervated by sympathetic processing (Gingras et al., 2015). In the current work, pupillometry data are captured in response to music excerpts in healthy older adults (current chapter) and individuals living with typical and atypical Alzheimer's disease (Chapters 4-5).

3.2.4 *The curvilinear pupil-affect relationship*

Researchers have manipulated numerous musical properties to investigate different pupil response profiles [e.g. elevated pupil response to vocal vs. instrumental excerpts (Weiss, Trehub, Schellenberg, & Habashi, 2016); elevated pupil size for music featuring deviations in terms of rhythm or melodic predictability (Bianco, Gold, Johnson, & Penhune, 2019; Bianco, Ptasczynski, & Omigie, 2020; Fink, Hurley, Geng, & Janata, 2018)]. The complexity however of detangling what exactly underlies an emotional response to music is a challenge for psychophysiological experimental design. This complexity is indirectly highlighted by a music-physiology review stating only 25% of reviewed literature within the wider music-physiology field had attempted to couple psychophysiological responses to music with emotional appraisals (Hodges, 2010).

Music which is self-selected to elicit joy activates the autonomic nervous system, indicating that pleasantness may be readable by candidate biosignals, such as pupil responses (Salimpoor et al., 2009). Evidence for differential physiological reactions in response to emotional appraisal of auditory stimuli has been particularly well documented in pupillometry research, a least in healthy younger adults. A curvilinear pupil-affect relationship, characterised by a larger pupil response to subsequently-rated positive and negative stimuli (relative to neutral; see Figure 3.1 for schematic representation), was reported in 30 neurotypical adults ($\bar{x}_{AGE(years)}$ = 23.5) (Partala & Surakka, 2003). Pupil responses for selected positive, negative and neutral environmental sounds were compared, exploring the potential of relating the PDRs to subjective ratings of valence and arousal (Partala & Surakka, 2003). A similar significant curvilinear relationship between maximum pupil size and valence scores of positive, neutral and negative environmental sounds (r^2 = 0.44) was also reported in 26 older healthy controls ($\bar{x}_{AGE(years)}$ = 67) (Fletcher, Nicholas, et al., 2015b).
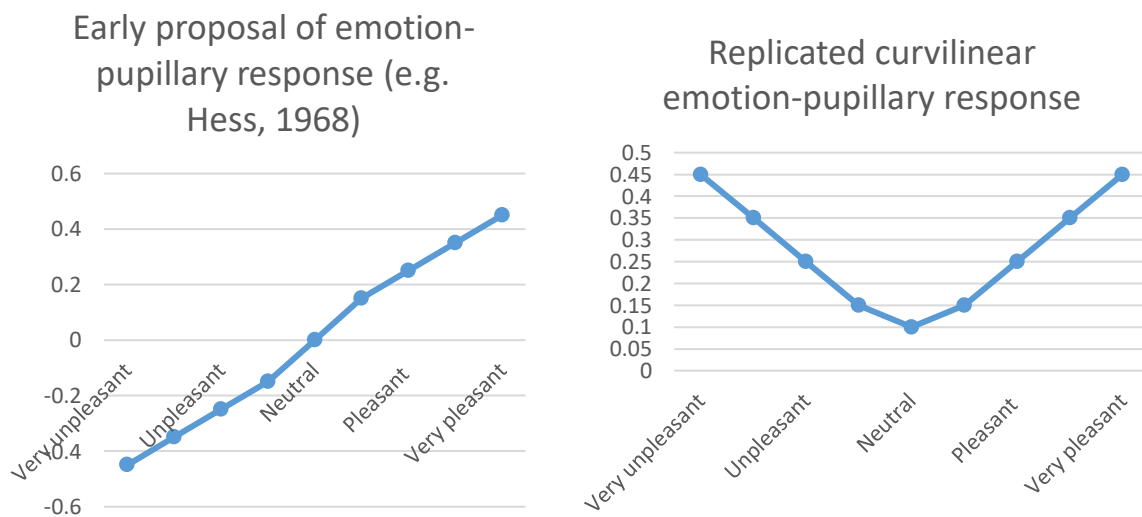
Figure 3.1 Schematic representation of the two proposed emotion-pupil response profiles.

The curvilinear pupillometry dynamic described above underscores the complexity of the pupil-emotion relationship. The emotional appraisal of a stimulus, as proposed by the *circumplex model of affect*, arises from a combination of two intersecting axes: **valence**[*pleasant-unpleasant*] and **arousal**[*activation-deactivation*] (Russell, 1980). In this context, the curvilinear pupil profile is thought to reflect an autonomic arousal response to emotional stimuli versus neutral stimuli, rather than a particular 'valence' response (whereby evidence for whether or not different physiological profiles can be

determined for each emotion is mixed) (Barrett, 2014). The curvilinear pupil-affect response profile is mirrored in other modalities, namely electrodermal activity (Bradley, Miccoli, Escrig, & Lang, 2008).

Further pupillometry studies have sought to explore additional potentially mediatory effects of the emotional response to music in healthy adults. Gingras et al. (2015) added a third dimension to the circumplex model, investigating associations of 30 younger neurotypical adults' pupil responses with participant-rated appraisals of music excerpts' pleasantness, arousal and tension indices, as well as familiarity (included to verify that all musical excerpts were of relatively low familiarity to avoid mediating effects). Bisecting the musical excerpts in terms of arousal (high/low), Gingras et al. (2015) reported a significantly larger pupil response associated with high arousing musical excerpts, the effect of which was weaker in individuals who reported an overall higher liking for the excerpts. Similar findings were demonstrated for the same musical excerpts bisected by tension ratings (high/low), discussed in terms of the two properties of arousal and tension being highly correlated (Gingras et al., 2015). No significant association of pleasantness ratings with pupil response was reported. In a study investigating the mediatory effect of predictability and pleasantness on pupil responses in non-musicians, Bianco et al. (2019) reported an increase in pupil dilation response for melodies which were liked (high>medium>low-liking ratings), however this dilation response was greater when the melody was more predictable. This evidence indicates a variety of factors contributing to pupil responses to music in addition to merely stated preferences.

3.2.5 Familiarity-emotion links in music

A subjective familiarity effect on music preferences has been demonstrated in neuroimaging studies [higher activation of both reward networks and limbic and paralimbic emotion-related regions when participants listened to self-rated familiar compared to unfamiliar music], psychophysiological studies and cognitive psychology experiments (Pereira et al., 2011). Psychophysiological studies investigating the relationship between emotional arousal, pleasure and familiarity derived from music listening revealed that increased EDA was only observed in response to 35 unfamiliar music excerpts during repeated exposure (van den Bosch, Salimpoor, & Zatorre, 2013). The 'BRECVEMA' framework proposes that one emotional response to music

is evoked via episodic memory (the others being brain stem reflex, rhythmic entrainment, evaluative conditioning, contagion, visual imagery, musical expectancy and aesthetic judgment) (Juslin, 2013). Notably, in a subsequent study whereby four of these psychological mechanisms were systematically manipulated through musical stimuli, the music designed to elicit an episodic memory was the only condition whereby significantly higher electrodermal activity compared to baseline was observed in younger adults (Juslin et al., 2013).

As outlined in Chapter 1, musical memory and engagement appears to be relatively spared in individuals with Alzheimer's disease (Cuddy, Sikka, & Vanstone, 2015), and there exists an emphasis on music familiarity in a variety of music protocols aiming to reduce agitation, and/or provide pleasurable moments for people with dementia (Garrido, Dunne, et al., 2019; Gerdner, 2021). However, studies investigating the music-familiarity link in PLWD indicate that music familiarity (manifesting as an episodic memory or otherwise) may not be sufficient information to reliably or accurately determine how likely a person is to have a pleasant experience engaging with that song (Garrido et al., 2018; Garrido, Stevens, et al., 2019). This perspective is corroborated by Barradas et al.'s (2021) replication of Juslin et al.'s (2013) study exploring autonomic responses to music in PLWD. Using the same four conditions within the BRECVEMA framework, the authors found that individuals with dementia rated a song designed to elicit an episodic memory lower in terms of their liking, in comparison to the control group's rating of the same song. This finding highlights the importance of investigating the extent to which we can couple individuals' contiguous self-reported emotional and cognitive appraisals of musical stimuli with in-the-moment continuous physiological recordings. The current study aims to establish whether differential physiological responses indexing distinct profiles of pleasantness and familiarity can be observed in response to a music listening task, by measuring pupillometry responses and acquiring self-reported indications of pleasantness and familiarity for each song. This chapter describes such a paradigm, and its investigation in a healthy older adult population [*Note:* Chapter 4 extends this investigation to apply the same paradigm within- and between different syndromic groups of people living with typical/atypical Alzheimer's disease].

3.2.6 The pupil old-new effect

One PDR which has attractive applications to the field of dementia research is the 'pupil old/new effect'. This profile reflects greater increases in pupil dilation for familiar stimuli relative to novel items (Võ et al., 2008). The effect has been observed in studies investigating responses to musical stimuli. While Gingras et al. (2015) included familiarity ratings for the purpose of controlling for its potential effect, they did report a positive - albeit non-significant - correlation between familiarity and pupil dilation. This non-significant finding is presumably attributable to the fact that all stimuli included were intended to be of low familiarity, thus reducing the potential for the effect to be demonstrated statistically. Published findings profiling the response time for differential patterns of task-evoked pupil dilation events did however indicate larger when presented with snippets of familiar relative to comparable unfamiliar songs in a sample of young healthy participants (Jagiello, Pomper, Yoneya, Zhao, & Chait, 2019). These findings indicate young adults are able to process whether a song is familiar or not within 100-300ms post-stimulus onset. This contrasts to Weiss et al. (2016) who reported a significant interaction of time bin (seconds) and song familiarity (old/new), in that mean pupil dilation was significantly larger whilst young adults listened to familiar songs, with this effect only being observed in the latter five seconds of each 12-second excerpt. These different findings may reflect either the differences in paradigms between the two studies (the stimuli set in Weiss et al. (2016) were researcher-selected, whereas Jagiello et al. (2019) utilised participant-selected personally salient familiar songs) or the wider timeframe (1-second chunks) analysis approach in Weiss et al. (2016) which may have masked any more granular rapid effects. Regardless of methodological differences, the demonstration of this effect in response to music in younger neurotypical populations indicates that this is a candidate PDR worthy of further investigation first in healthy older adults, with potential application to clinical populations. The current study concentrates on these two particular dynamic pupil response profiles which evidence suggests respectively index affective and familiarity appraisals of a variety of stimuli.

3.2.7 The impact of healthy ageing on pupil dilation response

For the current work, the clinical application relies to an extent on the ability to observe both these emotional and cognitive effects on pupillometry responses firstly in healthy older adults. Rationale for investigating responses specifically in this group relates to findings which suggest that - relative to younger individuals - pupillary parameters may reduce as a function of normal physiological aging (Yildirim Biçer, Zor, & Küçük, 2022). This reduction occurs by virtue of age-related pupil changes (e.g. atrophy to the dilator pupil muscle coupled with decreasing sympathetic activity which diminishes dilator muscle tone) (Yildirim Biçer et al., 2022). These pupillary response changes may impede the ability to observe the aforementioned psychophysiological pupil dilation responses even in healthy disease control groups.

Studies investigating age-related pupil changes have mainly focused on response changes to light; however, age-related changes have also been observed in the psychosensory pupil response, particularly in relation to cognitive load (Van Gerven, Paas, Van Merrienboer, & Schmidt, 2004). One study investigated comparative pupil dynamics between younger and older adults recorded during a listening task to harmonic sounds (Zhao, Bury, Milne, & Chait, 2019). The paradigm was designed to induce pupil dilation by virtue of requiring sustained attention to complete the experimental task of detecting gaps in concurrent, spectrally distinct tone streams, in the presence of distractors. Zhao et al. (2019) found that, relative to younger control, the older group (aged 63-79) exhibited lower within-subject variability and higher between-subject variability relative to younger adults. The findings were discussed in the context of pupillometry being justifiable in healthy older populations; "obtainable from older listeners but with certain caveats" (Zhao et al., 2019). Encouragingly, similar pupillary response profiles have been observed in older adults relative to younger adults in response to a short-term memory load auditory task (Piquado, Isaacowitz, & Wingfield, 2010). Most studies reporting age-related pupillary differences have specifically focused on changes in pupil dilation velocity, citing a relative slowness to the pupil response in healthy older adults in comparison to younger individuals (Tekin et al., 2018). For the purposes of the current study which seeks to investigate size differences between conditions, this justifies the selection of a single pupil feature (maximum pupil size) to explore in this population. This also reflects previous pupil

feature selection in studies of older adults (acting as a control group) and PLWD (Benhamou et al., 2021; Fletcher, Nicholas, et al., 2015b).

The current study explores whether the curvilinear affect-pupil relationship and the pupil old-new effect can be observed in healthy older adults in response to music that is respectively rated in terms of pleasantness and familiarity.

3.2.8 Research questions and hypotheses

Pupillometry responses

1. Are healthy older adults' self-reported pleasantness music ratings associated with pupillometry change from baseline? [*Associated hypothesis (RH1)*]: *We predicted a curvilinear association of pleasantness ratings with pupil response change from baseline.*

2. Are independent classifications of a song's familiarity associated with healthy older adults' pupil response change from baseline? [*Associated hypotheses (RH2)*]*: We predicted an association of a priori familiarity predictions with pupil response change from baseline.*

3. Are healthy older adults' subjective familiarity music ratings associated with pupillometry change from baseline? [*Associated hypotheses (RH3)*]: *We predicted an association of self-reported familiarity ratings with pupil response change from baseline.*

4. Is memory evocation of a song in healthy older adults associated with pupillometry change from baseline? [*Associated hypotheses (RH4)*]: *We predicted an association of binary memory ratings with pupil response change from baseline.*

Secondary hypotheses related to associated subjective music ratings

5.   Are independent classifications of a song's familiarity associated with older adults' pleasantness ratings? [*Associated hypothesis (RH5)*]*: We predicted an association of a priori familiarity ratings with pleasantness rating*

6.   Are older adults' subjective judgements of a song's familiarity associated with their pleasantness ratings? [*Associated hypothesis (RH6)*]*: We predicted an association of subjective familiarity rating with pleasantness rating*

7.   Are independent classifications of a song's familiarity associated with older adults' likelihoods of reporting that song as familiar? [*Associated hypothesis (RH7):]* We predicted an association of a priori familiarity predictions with self-reported familiarity scores*

8.   Are older adults' subjective judgements of a song's familiarity associated with their likelihood of reporting that song as evoking a memory? [*Associated hypothesis (RH8)*]: *We predicted an association of subjective familiarity scores with memory evocation ratings*

9.   Are songs released within older adults' putative 'reminiscence bump' (between 13 and 19 years old) associated with subjective music ratings? [*Associated hypothesis (RH9)*]*: We predicted an associated of a song being released between each participants' age of 13 and 19 with participants' subjective music ratings*

3.3   METHODS

3.3.1 Participants and characterization

100 healthy control participants were opportunistically recruited, either by interrogating the Dementia Research Centre research participant database, or via the Join Dementia Research platform (https://www.joindementiaresearch.nihr.ac.uk/). Ten participants were excluded from the subsequent analysis due to technical issues (N =

6), and reporting a neurological condition potentially affecting physiological responses (N = 4). Neurological conditions disclosed by the four control participants included: vestibular neuritis, multiple sclerosis and a comorbid presentation of restless leg syndrome and essential tremor in one individual (medicated with clonazepam). Background participant information on the remaining 90 participants is shown in Table 3.4. Health status (e.g. visual/hearing impairments, neurological comorbidities, family history of a neurological condition) was documented using the UCL Dementia Research Centre demographics questionnaire, also used in clinical assessments. Participants additionally completed a questionnaire detailing their previous and current levels of music engagement. Musical experience (0-4) was ascertained from selected variables (*played instrument, engaged in singing, highest grade attainment, length of time played/singing*) and recoded in a similar approach to Benhamou et al. (2021). Participant year of birth was subtracted from song release year to indicate which songs were released in the reminiscence bump period in accordance with Rao et al. (2021) (13-19 years of age, inclusive). A binary *reminiscence bump* variable was created to reflect this characteristic of each stimulus for each participant. The variable *age* was rescaled to reflect units in decades rather than years for clarity in model interpretation (dividing participant age by ten). All participants attempted the Wechsler Abbreviated Scale of Intelligence (WASI Test) - Matrices (Shortened version).

3.3.2   Ethical approval and funding

The study was approved by the Queen Square NHS Research Ethics Committee (17/LO/0099) both adhering to British Psychological Society ethical standards and in accordance with the principles outlined in the Declaration of Helsinki. The study was part of the Created Out of Mind residency ('Created Out of Mind: Shaping Perceptions of Dementias, Grant Ref: 200783/Z/16/Z), funded by the Wellcome Trust as a part of the Hub Award (Principal Investigator S. Crutch; see **Brotherhood et al., 2017**).

3.3.3   Musical stimuli selection

Stimuli and relevant metadata are outlined in Table 3.1. 32 songs were selected from a British music database (two practice items, 30 experimental trials) to comprise a variety of genres and release dates (5 classical, 10 songs from the 1960s, 10 songs

from the 1970s, and 5 'nonsense' samples, created *de novo* by reversing 5 of the selected songs and increasing their speed. Recorded music was selected by virtue of ascertaining relative chart success of famous (known) melodies in year of release (and notoriety subsequent to release date.). 'Unknown' songs released in the 1960s and 1970s were selected by identifying lesser-known artists who created music of a similar style in the same decade (Eerola, 2011). Five nonsense excerpts were created as novel instrumental stimuli that could act as control stimuli (i.e. known to be not familiar), yet remained interpretable as music. Excerpts were selected from each song, and fixed at a total sample length of 12000ms. The excerpt timeframe was selected arbitrarily in the unknown songs; for famous 'known' songs, the excerpt selection for experimental stimuli was discussed and agreed amongst the research team (the aim being to reflect a notorious riff or part of the song). Twelve-second excerpts were selected to consider longer latencies of PDRs reported in older and clinical populations. In line with previous pupillometry studies, volume was normalised across samples to avoid potential confounds of sound intensity on pupillary response (Gingras et al. 2015).

Stimulus properties of each excerpt (Beats Per Minute [BPM] and key) was established by submitting each stimulus file to an automated BPM and key detection online application, 'Tunebat' (https://tunebat.com/Analyzer). Efforts were made to balance the a priori known groups in terms of these potential confounding acoustic properties, as well as lyrical content, as previous studies have demonstrated these properties may influence physiological responses (Walker et al., 2021; Garrido et al., 2018). Similarity of excerpt BPM between the *a priori familiarity* groups was confirmed using an independent samples *t*-test (t(28) = -0.890, p = 0.381), and between *genre* using a Kruskal-Wallis test (H(4) = 2.413, p = 0.660). Comparison of excerpt key (recoded as major/minor) indicated no significant differences between *a priori familiarity* (p = 1.000, Fisher's Exact Test), or *genre* (p=0.557, Fisher's Exact Test).

Table 3.1 List of stimuli for the Play It Again paradigm, split by a priori prediction and group

A priori prediction (familiarity)	Group	Artist	Song title	Excerpt Beats per Minute (BPM)	Excerpt key
Known	Classical	Mozart	Serenade in G Major k525	68	Major [G]
		Holst	Jupiter	101	Major [G]
		Beethoven	Für Elise	106	Minor [A]
		Elgar	Nimrod	107	Major [B♭]
		Handel	Harpsichord Suite No. 4	119	Minor [D]
	1970s	Queen	Bohemian Rhapsody	70	Major [B♭]
		Elvis	Love Me Tender	82	Major [E♭]
		Grease	You're The One That I Want	106	Major [C]
		Boney M	Rivers Of Babylon	115	Major [G]
		Blondie	Heart of Glass	115	Major [E]
	1960s	Del Shannon	Runaway	76	Major [B♭]
		Nat King Cole	L-O-V-E	78	Major [D]
		Engelbert Humperdinck	Release Me	106	Major [B♭]
		Frank Sinatra	Fly Me To The Moon	119	Minor [A]
		Beatles	I Wanna Hold Your Hand	135	Minor [E]
Unknown	Nonsense	A	-	77	Major [D]
		B	-	104	Major [E♭]
		C	-	123	Major [C#]
		D	-	136	Major [C]
		E	-	139	Major [B♭]
	1970s	Hollins & Starr	Feelin' Good	68	Major [C]
		Richard Wagner	Oceans	73	Major [E♭]
		Marion Black	Who Knows	117	Minor [Dm]
		Fancy	Touch Me	123	Minor [B]
		Speedy Keen	Old Fashioned Girl	132	Major [A]
	1960s	Mellotones	Feel Good	86	Minor [G]
		Eddie Cooley	Hey You	87	Major [B♭]

Little Joe & the Thrillers	Please Don't Go	107	Major [A♭]
Randy Starr	You Growin' Up	115	Major [C]
Billy Storm	Please Don't Mention Her Name	124	Minor [F#]

3.3.4   Self-report data: Active categorisation task via Likert-scale and binary prompts

The majority of musical memory in PWLD comprise three paradigm approaches: familiarity decision-based tasks, recall of composer or melody title, or singing the last note of a musical excerpt. The current work sought to capitalise on contiguous musical ratings after each excerpt, reflecting the potential application of the current work. As Section 1.2.3. points out, an active categorisation task of familiarity following each trial is likely to elicit familiarity ratings comparable to controls in particular syndromic groups.

Despite *a priori* predictions of stimuli familiarity being a sufficient marker in other studies relating to psychophysiological response (Otero, Weekes, & Hutton, 2011), it was also important in the current work to model potential associations of pupil responses to subjective ratings of the music. This enabled us to firstly ascertain whether our familiarity predictions of each excerpts' notoriety were accurate, and also to account for individual differences in terms of musical preference and exposure (Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2007). Ascertaining familiarity ratings in terms of a Likert response in terms of degree (1 - not familiar to 5 - extremely familiar) and a binary question prompting a participant to ascertain the depth of that familiarity (memory evocation question) also enabled different statistical models to consider pupil psychosensory responses in relation to a song's recollective strength. Including a familiarity decision-based task predominantly engages semantic or conceptual memory, which as outlined in Section 1.2.3. has been shown to be relatively spared in these two syndromic groups (Slattery et al., 2019). Using non-standardised Likert scales replicates the methods utilised in numerous music-psychophysiology studies, in neurotypical (Gingras et al., 2015), older adult and clinical populations (Fletcher, Nicholas, et al., 2015b). Taking this approach rather than a 'name that tune' design enabled the paradigm to capture responses from an emotivist or 'felt' point of view, rather than create a cognitivist paradigm with correct hit/error rates. Whilst verbal

75

semantic memories (e.g. name that tune) and semantic musical knowledge (a feeling of familiarity) are strongly associated, the paradigm was designed without a 'name that tune' question; an intentional design decision supported by Groussard et al. (2019) and in consideration of the challenges with verbal semantic memory specifically observed in Alzheimer's disease. In accordance with one proposed model of musical memory (which Groussard et al. (2019) use a face recognition model (Bruce & Young, 1986) analogy), name generation of either song title or artist is the highest processing capacity accessed when experiencing a musical stimulus. Therefore, whilst sporadic exclamations of song/artist name in the current patient population may reflect relative preservation - at least in some participants - of verbal semantic knowledge of the excerpts included, it must be noted that we recruited individuals with mild-moderate disease severity. If we accept this musical memory model by Groussard et al. (2019), verbal semantic memory is likely to be one of the initial musical appraisal elements to be impacted by disease progression in Alzheimer's disease. In addition, participant insight of there being a 'correct/incorrect' answer may unduly influence psychophysiological responses; pupil dilation may instead reflect the cognitive effort required to reach a 'correct' response. This approach also reflects a more real-world scenario, whereby - for the most part - we are all able to enjoy a song in-the-moment, even if the level of familiarity we attribute to it or the metadata we can recall about it is factually inaccurate or unavailable to us.

Self-report measures comprised two five-point Likert-scales and a binary yes/no question. The Likert-scales prompted the participant to respond in accordance with their self-rated levels of *pleasantness (1-5)*, *familiarity (1-5)*, and *memory evocation (binary)* (for all response options, see Table 3.2).

3.3.5 Equipment

Stimuli were presented on a Dell Latitude E6540 laptop elevated 240mm from the desk, from a viewing distance of 600mm, fixed using an SR Research Head Support. The visual stimulus throughout music presentation was a dark grey fixation cross (RBG 75, 75, 75) presented against a lighter grey contrast background (RGB 150,150,150, in line with Gingras et al. 2015). Pupil area was recorded using a desktop-mounted infrared eye tracker (Eyelink 1000 Plus® v5.09; SR Research)

Table 3.2 Response options for self-report questions

Self-report question	Levels(n)	Levels(response)
Pleasantness *['How pleasant did the piece make you feel?"]*	1	Very unpleasant
	2	Unpleasant
	3	Neutral
	4	Pleasant
	5	Very pleasant
Familiarity *["How familiar are you with the piece?"]*	1	Not familiar
	2	Slightly familiar
	3	Somewhat familiar
	4	Moderately familiar
	5	Extremely familiar
Memory evocation *["Did the piece evoke a memory?"]*	-	Yes
	-	No

sampling at 1000Hz. Accommodating a high sampling rate (1000Hz) was integral for the eventual dual application of the paradigm, in terms of allowing various downsampling options for primary analyses [e.g. to lower sampling rates recommended to reduce the risk of multicollinearity (van Rij, Hendriks, van Rijn, Baayen, & Wood, 2019)], and to accommodate the capacity for submitting the same dataset for algorithm testing and development. The Eyelink 1000 also has the capacity to collect a wider range of features relative to other eyetracking systems, enabling multiple features to be explored within secondary analyses. The feasibility of testing different dementia syndromic groups using this system as demonstrated in previous studies (Benhamou et al., 2021; Fletcher, Nicholas, et al., 2015b) increases the chances of successful data acquisition in the current work. A baseline period was inserted into each trial to reconcile any changes in overall pupil size over the course of the experiment. As Partala and Surakka (2003) reported stimulus effects within two-second post-stimulus offset, a 5-second 'washout' period was introduced subsequent to each stimulus offset to avoid between-trial contamination of response.

Data were collected binocularly for the majority of participants; monocular recording took place if the participant declared a visual impairment in one eye deemed to impact

a physiological response (e.g. Horner's syndrome) and/or a strong glare from a participant's glasses obscured recording from one eye that could not be rectified by reposition the camera/participant, and whereby wearing glasses was necessary for the participant to view the fixation cross. Data were collected in pupil and corneal reflection mode (SR Research, 2010). The experiment took place in quiet uniformly lit rooms across two sites (Wellcome Trust N = 94; Dementia Research Centre N = 6). Stimuli were presented at a comfortable listening level (approximately 70dB) through headphones (Behringer HPM1000) where tolerated or via the Dell Latitude E6540 speakers at a comfortable listening level attuned to individual participant preference (average speaker volume = 67.24 dB sd = 2.49). Audiovisual data were captured using the Veho VCC-005-MUVI-HD10 Mini Handsfree Action Cam (at 1280 x 720p at 60fps), fixed on a small tripod next to the Eyelink camera lens (see Figure 3.2). Accelerometer(ACC), blood volume pulse (BVP), heart rate (HR), electrodermal activity (EDA) and skin temperature data (TEMP) were collected via two unobtrusive wristbands (Empatica® E4 device).

3.3.6 Procedure

Participants completed a written consent form. Following this, one Empatica® E4 device was fitted to each wrist. To allow for a baseline recording period to be established following device fitting, the demographic questionnaire was administered to participants before starting the eye-tracking task. Participants received standardised instructions and were advised to avoid blinking during the music presentation. A five-point randomised target calibration (deemed sufficient for pupillometry studies) (see Gingras et al. 2015) was performed (see Figure 3.3). Two practice trials were conducted to familiarise the participants with the paradigm.

The researcher conducted the experimental procedures out of the participants' line of sight. Participants were presented with a fixation cross appearing centrally on-screen 500ms prior to the onset of the excerpt (12000ms). Stimulus presentation order was randomised. Upon stimulus offset, an interval of 5000ms preceded three self-report

Figure 3.2 Photograph of the experimental setup

questions appearing sequentially on the display screen. The participant responded verbally to each of the post-trial questions in turn, either using the number or phrase appropriate to their reported levels of familiarity, pleasantness, and 'yes' or 'no' in response to whether the music evoked a memory. No time limit on self-report responses was imposed. The experimenter input the number on the keyboard relating to the participant's response at the time of response. Pupil dilation responses were collected from the trial (N = 30) and practice (N=2) songs, however practice trials were not further analysed.

Following the presentation of the excerpts, all participants attempted the Wechsler Adult Scale of Intelligence Matrices Test (Short Version). Approximately 5 minutes after this task was completed (or discontinued in accordance with the test manual), the Empatica® devices were removed. Participants were fully debriefed once the experimental procedure was completed.

3.3.7 Data reported within the chapter

Sections 3.3.5 and 3.3.6 refer to additional physiological data strands collected as part of this study (audiovisual recordings for automated emotional response categorisation, and data acquired via the Empatica® device (ACC, BVP, HR, EDA and TEMP). Reporting these data strands falls outside the scope of this thesis, therefore subsequent physiological response reporting refers to pupillometry data only.
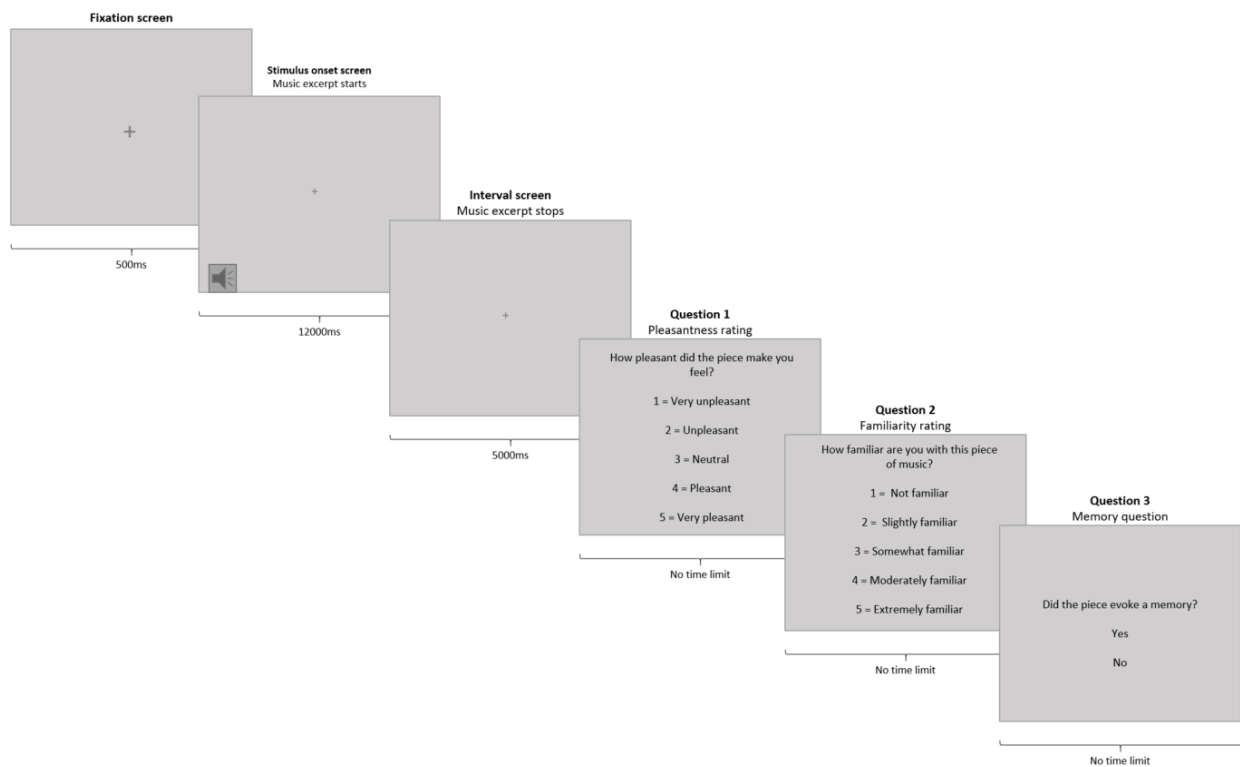
Figure 3.3 Sequence of one trial within the Play It Again experiment.

3.3.8 Missing pupillometry in data collection

Due to a fault on the host computer, one day of data collection was lost (N = 6, as outlined in Section 3.3.1). Other missing data attributable to artefacts such as eye blinks were concluded to be missing at random, and therefore missing values were imputed using established pupillometry pre-processing techniques so that the final pre-processed datasets contained no missing entries (see below).

3.3.9 Rating and pupillometry data pre-processing

Self-report ratings: Ratings were retained as ordinal data for initial analyses. Pleasantness/familiarity ratings were additionally recoded as numerical data (reflecting the numerical values ascribed to each pleasantness/familiarity level) to enable further exploratory analysis.

Pupillometry: The following steps reflect approaches and recommendations from Benhamou et al., 2021; Mathôt, 2018, and involving adaptation of functions within the *GazeR* package in R (Geller, Winn, Mahr, & Mirman, 2020). Fixations directed at the 1cm screen periphery were excluded from analysis within Eyelink Dataviewer to reduce potential influence of the Pupil Foreshortening Effect (PFE). This refers to the reduction in pupil size by virtue of the eye rotating to different screen locations, which is most problematic at the point of deviating from a central fixation cross, towards the furthermost edges of the screen (Hayes & Petrov, 2016). Blinks were detected by the Eyelink 1000 algorithm, and data collected 100ms pre- and post-blink on and offset were excluded using the GazeR *extend_pupil* function (Geller et al., 2020). For computational purposes, data were downsampled to 50ms (20Hz) in line with recommendations from van Rij et al. (2019). All pre-processing steps took place on mean pupil area data collected within the 50ms time bins, whereby pupil area was averaged across each pupil, or in the case of monocular blinks, the available pupil area was taken. Raw data points were initially converted to millimetres and plotted to confirm that minimal data points fell outside the expected normal pupil size range (roughly between 2mm and 8mm) (Mathôt, 2018). Data exceeding more than three standard deviations above or below each participant's mean pupil area across the available dataset were excluded as outliers, and a further exclusion criterion was set as 50% missing data within trials and within subjects. Applying this subsequently excluded 2.7% overall trials plus two participants from the dataset. Data were smoothed using the Hanning filter, and a cubic interpolation subsequently applied to span missing data points. Data were baselined trial-by-trial by dividing the pupil size during music listening by the mean pupil size 500ms preceding stimulus onset. The log of this ratio was calculated to reconcile skewness, and the resulting baselined pupil sizes were summarised at the trialwise level and exported for statistical analyses. Two additional trialwise outliers were removed by virtue of being greater than 6.5 times

81

larger than the group mean (reflecting patient group outlier removal, Section 4.3.7). All statistical pupillometry analyses were performed on the resulting dataset using maximum baselined pupil size per trial (pupil$_{max}$) as the outcome, reflecting that the *a priori* and subjective ratings had been obtained in relation to the entire excerpt (Gingras et al., 2015). The difference between each participant's pupil$_{max}$ response and the group mean was calculated as a metric of variability (pupil$_{var}$). The *average gaze x* and *y* position variable (original units = pixels) was rescaled for clearer model interpretation, by dividing by 100.

3.3.10 Statistical Methods

All statistical analyses were undertaken in R. The threshold $p < 0.05$ was established as the criterion for statistical significance for all analyses. 95% confidence intervals are reported where applicable.

Demographic, music experience and neuropsychological data

Demographic characteristics and WASI scores were compared between male and female healthy older adults using Chi-square (or Fisher's Exact Test where expected frequencies were lower than 5) for categorical data, and *t*-tests (or Wilcoxon rank sum tests in the case of nonparametric data) for numeric data, subject to normality testing.

Subjective music ratings

Ordinal pleasantness/familiarity ratings: Cumulative link mixed models (CLMM; using Laplace approximation to compute maximum likelihood estimates) were undertaken using the *clmm2* function within the *ordinal* package in R (Christensen, 2022) to examine potential effects on self-reported pleasantness and familiarity Likert scale ratings (see Box 3.1). The modelling and approximation approaches were selected for their capacity to allow the inclusion of multiple random effects in a model with an ordinal outcome. Data were visually inspected to ascertain whether the proportional odds assumption was upheld (whereby the logarithms of odds demonstrate an arithmetic series). This refers to the odds ratios from a proportional odds model corresponding to the ratio of the odds of a rating of 1 to the odds of rating between 2 and 5; which is also the ratio of the odds of rating 1 to 5 compared to the odds of rating of 3 to 5 and so on.

Box 3.1 *Cumulative link mixed models (CLMM)*

The CLMM is an established regression approach that models ordinal outcomes. It extends the cumulative link modelling (CLM) approach by virtue of its capacity to include mixed (fixed and random) effects. This approach is therefore able to handle both ordinal (Likert-scale) data, whilst simultaneously accounting for interindividual and/or interstimuli differences (Agresti, 2010).

This confirmatory approach was undertaken for each CLMM model in the current study. Where it was demonstrated that the assumption was not met, additional models were run using the *clmm* package in R whereby the proportional odds assumption was relaxed and a multinomial model was fitted, and then four separate logistic regressions were fitted so the odds ratios were not constrained to be equal. The log odds for each predictor variable were inspected in these additional models to ascertain if this improved model fit. If the model fit was not improved by this approach, we were satisfied that the original model findings could be reported. This was found to be the case in all CLMM models, and therefore all subsequent pleasantness- and familiarity-rating results are reported using the original CLMM findings. In the current proportional odds models, the odds ratios represent the odds of rating familiarity levels 1 ('not familiar') vs 2-5 ('slightly' to 'extremely familiar'). Odds of a higher familiarity rating in Model 3 therefore relate to either 'moderately familiar' or 'extremely familiar' in the original Likert scale, therefore for conciseness, we describe the coefficients in terms of rating a song as a 'greater' familiarity rating. The same approach has been taken for reporting pleasantness ratings. Kendall's Tau was used to assess the correlation between self-reported pleasantness and familiarity ratings.

Categorical memory evocation data: A multiple logistic regression was undertaken to determine the predictive variables for the binary memory evocation responses.

Rating model structures are outlined in Table 3.3, clarifying both fixed and random effects. Gender was not included as a predictive factor in the self-report models, having

not contributed to previous self-report ratings in a pupil-emotion investigation (e.g. Partala & Surakka, 2003). Exponential values of the model coefficients and confidence intervals were calculated to generate odds ratios (OR) and 95% confidence intervals, reported in square brackets.

Table 3.3 Self-report rating model composition, predicting pleasantness rating, familiarity rating, and memory evocation in healthy older adults. Model type indicated adjacent to each outcome marker.

Variables	\multicolumn{6}{c}{Model$_{(rating)}$ number}	Reference variable level	Motivation to include
	1	**2**	**3a**	**3b**	**4**	**5**		
Pleasantness rating	O^{CLMM}	O^{CLMM}					-	Outcome measure of pleasantness rating
Self-reported familiarity		X	O^{CLMM}	O^{CLMM}	X	X	*Not familiar*	Outcome measure of familiarity rating (Models 3a-b) and main predictor variable in Models 2 and 5 (comparing Slightly familiar - Extremely familiar with 'Not familiar' ratings)
Memory evocation					O^{LR}	O^{LR}	-	Outcome measure (of whether a song evoked a memory or not
A priori familiarity rating	X		X	X	X		*Song a priori predicted to be 'unknown'*	Included to ascertain whether the *a priori* familiarity predictions influenced pleasantness, familiarity and memory evocation ratings - comparing

								songs *a priori* predicted to be 'known' with songs predicted to be 'unknown'.
Genre	X	X		X	X		*Nonsense songs*	Included to explore incidences of different music listening preferences
Reminiscence bump (the song released between participant's age 13-19 inclusive)	X	X		X	X	X	-	See Rao et al., 2021 (peak recollection of songs occurring for songs released between 13 and 19 years of age (inclusive)
Trial number	X	X	X	X	X	X	-	To establish any incidence of fatigue, or perseverance of self-reporting response over the experiment course
Participant **[random effect]**	X	X	X	X	X	X	-	Randomising at the level of participant to take account of individual differences
Song ID **[random effect]**	X	X	X	X	X	X	-	Randomising at the level of song to take account of differences in song properties aside from genre/familiar characteristics

O = **outcome measures;** X **= predictor variables (main predictor in bold)**

Pupillometry data

For the pupil outcome, a modelling strategy was required which could handle crossed random effects as well as slightly skewed residuals which were identified upon visual inspection. We selected cluster robust standard errors, with the 'HC1' method to reflect the large sample size. These robust standard errors could handle the crossed random effects, by clustering on both 'song ID' and 'participant ID', and were robust to outliers caused by the skewness. Since the model estimates a different coefficient for each level of the 'pleasantness' and 'familiarity' variables, it can already take into account a non-linear relationship that may exist between the levels of self-report ratings (Fletcher, Nicholas, et al., 2015b). The R package *sandwich* (Zeileis, 2004; Zeileis, Köll, & Graham, 2020) was used to obtain cluster robust standard errors following linear regressions. Model composition, reference variable levels used in each model for comparison (where applicable), and rationale for variable inclusion are outlined here:

1. $Model_{(pupil)}1$ modelled pupilmax responses from pleasantness rating (the reference variable level selected was songs rated as 'neutral';
2. $Model_{(pupil)}2$ modelled pupilmax responses from *a priori* familiarity rating (the reference variable selected was songs which were *a priori* predicted to be unknown);
3. $Model_{(pupil)}3$ modelled pupilmax responses from familiarity level (the reference variable level selected was songs rated as ''not familiar';
4. $Model_{(pupil)}4$ modelled pupilmax response from memory evocation rating (the reference variable selected was songs rated as not evoking a memory).

All pupil models included the following predictor variables:
*   *Setting*: Included as a nuisance covariate to control for potential influence of experimental setting (Wellcome/Dementia Research Centre) on pupil dilation response;
*   *Trial number*: To establish the prevalence of pupillary hippus in the current dataset, whereby the pupil dilation response (PDR) reduces in size over the course of an experimental session (see Zhao et al., 2019);

- *Age*: To establish whether there is an effect of age (unit: decade) on the pupil dilation response within a healthy older adult population in the current dataset (see Zhao et al., 2019);

- *Gender*: To establish whether gender influences the PDR reflecting sex difference findings (e.g. Partala & Surakka, 2003; Gingras et al., 2015; the reference variable level was female);

- *Average position on the X and Y axes*: To establish whether gaze position has an effect on pupil response by virtue of the pupils deviating horizontally and vertically from the point of central fixation, as outlined in Hayes and Petrov (2016).

Additional exploration of associations of pleasantness/familiarity ratings and pupil size: Recoding the pleasantness/familiarity ratings as numerical data allowed for further estimation of metrics such as participant variability, and ease of schematic representation. Relaxing the Likert scale ordinal requirements additionally enabled formal analyses exploring the potential curvilinear valence-pupil relationship as observed in Fletcher, Nicholas et al. (2015b). This was established by modelling mean $pupil_{max}$ from mean pleasantness rating and (mean pleasantness rating)^2, including participant ID as a random effect. Numerical values also facilitated the investigation of a potential relationship strength (r^2) between pleasantness/familiarity ratings and playlist properties (e.g. excerpt BPM) and between pleasantness/familiarity ratings and other musical properties (e.g. key and lyrical content) using simple linear regression.

3.4   RESULTS

Significant findings at the level of p < 0.001 are denoted by double asterisks; single asterisks denote p-values at the level of 0.001 < p < 0.05.

3.4.1 Participant characteristics and music experience

No significant differences (and therefore candidates for model covariates) were found between male and female older healthy adults in any of the demographic factors, scored music experience, or in neuropsychological performance (see Table 3.4). Comparing detailed music engagement responses between participant genders revealed one marginal significant difference between the groups in terms of higher frequency of women playing an instrument relative to male participants (p = 0.044). This however was not included as a covariate owing to overall music score being comparable; detailed music responses are available in Appendix C. Classical music was the most popular listened-to genre for both male and female healthy older adults.

Table 3.4 Demographic, clinical and musical experience of the healthy older adult sample.

	Gender		Total
	Female	**Male**	
No. in group	*52*	*38*	*90*
Age (years): $\bar{x}$ *[IQR]*	67.0 [60.8;73.0]	71.0 [62.8;75.0]	69.0 [61.0;74.0]
Min./Max.	52.0 / 80.0	51.0 / 78.0	51.0 / 80.0
Handedness (left:right)	4:48	3:35	7:83
UK-born	39 (75%)	30 (79%)	69 (77%)
Residing in a different country to birth for > 1 year	25 (48%)	19 (50%)	44 (49%)
Educated to at least undergraduate degree	44 (85%)	29 (76%)	73 (81%)
Native English speaker	47 (90%)	36 (95%)	83 (92%)
English fluency	51[1] (98%)	38 (100%)	89 (99%)
Visual impairment present[2]	8 (15%)	5 (14%)	13 (15%)
Hearing impairment present[3]	3 (6%)	3 (8%)	6 (7%)
In a 'good' mood	49 (94%)	33 (92%)	82 (93%)

Reporting a close (blood) relative with a neurological condition	18 (35%)	10 (27%)	28 (31%)
Dyslexia (self-reported)	1 (2%)	0 (0%)	1 (1%)
Self-reporting a neurological condition	0 (0%)	0 (0%)	0 (0%)
No. receiving acetylcholinesterase medication	0 (0%)	0 (0%)	0 (0%)
Wechsler Adult Scale of Intelligence: Matrices (Short version) %: $\bar{x}$ [IQR]	80 [80;80]	80% [70;90]	80% [70;80]
Scored music experience: 0-4: $\bar{x}$ [IQR]	1 [0;3]	0 [0;2]	1 [0;3]

[1] As outlined in the Created Out of Mind study protocol (17/LO/0099 and 8545/002: Created Out of Mind), participants were required to be proficient in the English language to take part in the experiment. For all non-native English speakers, their English comprehension was deemed proficient by the research team to continue participation, with the exception of one fluent Italian speaker. For this participant, a bilingual Italian-English academic collaborator on the project sought and obtained consent, and conducted the experiment.

[2] Visual impairments listed relating to the eyes to ascertain precautions around pupillometry recording (e.g. monocular recording). Visual impairments reported here any impairment could not be corrected-to-normal e.g. with the use of spectacles or contact lenses.

[3] Participants required to wear hearing aids were asked to bring them to the experiment. Hearing impairments reported here demonstrate any impairment which could not be corrected-to-normal e.g. with the use of hearings aids, or due to the participant forgetting to bring hearing aids to the experiment.

3.4.2 Pleasantness responses to music

3.4.2.1 Pupil-pleasantness responses

The Model$_{(pupil)}$1 intercept revealed a significantly larger pupil$_{max}$ response change from baseline whilst listening to reference music excerpts (songs which were subsequently rated as neutral; $p<0.001$; see Table 3.5). In line with previous studies, a curvilinear valence-pupil relationship was observed; pupil$_{max}$ responses were larger whilst participants listened to songs which subsequently elicited any pleasantness rating relative to 'neutral' music excerpts. These larger pupil$_{max}$ responses reached statistical significance whilst participants listened to music excerpts subsequently rated as either 'pleasant' (38.9 [37.4, 40.3]%, $p = 0.017$) or 'very pleasant' (41.6 [40.0, 43.4]%, $p < 0.001$), relative to pupil$_{max}$ elicited during music excerpts subsequently rated 'neutral' (37.1 [27.8, 47.1]%). Recoding pleasantness ratings as a continuous variable enabled a formal interrogation of the curvilinear relationship including all the terms below as both linear and quadratic terms (var^2), revealing a significant but weak pupil-pleasantness relationship ($r^2 = 0.147$, $p <0.001$) (see Figure 3.4).   Pupil$_{max}$

response was significantly associated with *experimental setting* (university>Wellcome, $p < 0.001$) and g*ender* (male>female, $p<0.001$). Smaller $pupil_{max}$ responses were significantly associated with increasing *trial number* and *participant age.*
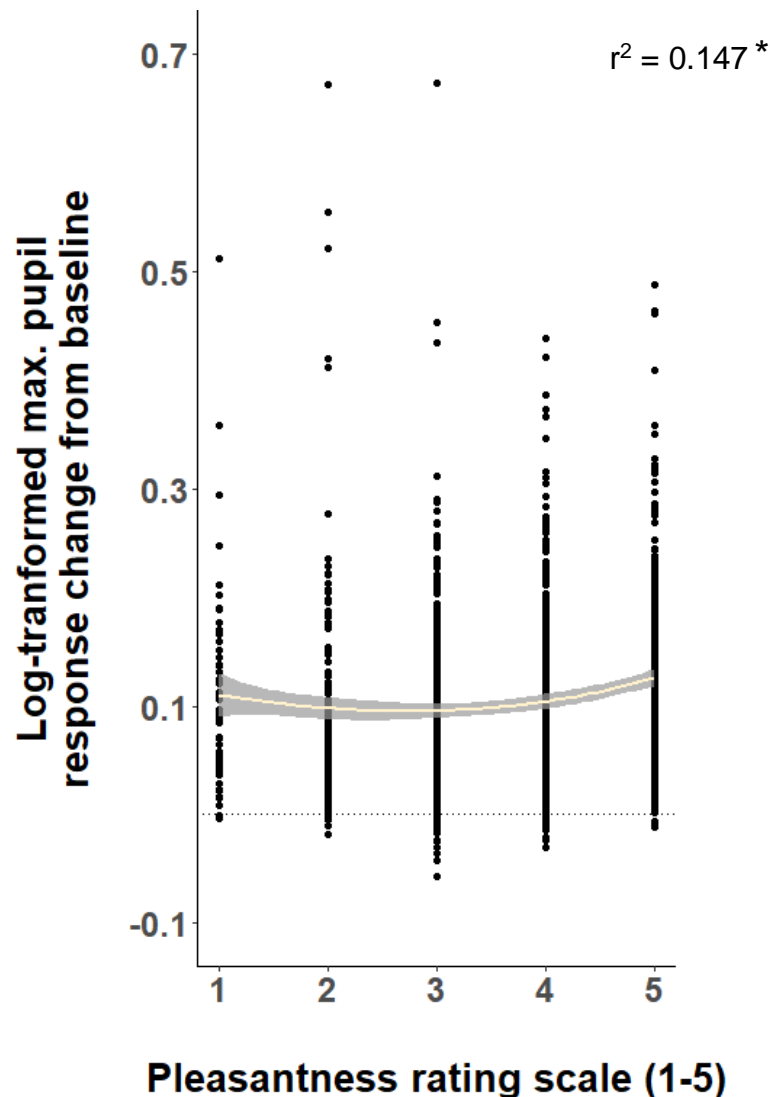
Figure 3.4 Older adults' maximum pupil response change from baseline plotted for music excerpts subjectively rated for pleasantness. Pleasantness ratings are on a Likert scale ranked: 1 = 'very unpleasant', 2 = 'unpleasant', 3 = 'neutral', 4 = 'pleasant', and 5 = 'very pleasant'. Quadratic regression lines of best fit with 95% confidence intervals (shaded in grey) and r^2 are indicated. Points (·) indicate individual trial maximum pupil size change from baseline. Dotted line indicates zero pupil size change from baseline.

3.4.2.2 Pleasantness responses predicted by a priori familiarity stimuli selection

CLMM modelling (Model$_{(rating)}$1) revealed strong evidence that the odds of rating a song with a higher pleasantness category were associated with *a priori familiarity*

prediction (see Table 3.6). Participants were nearly eight times more likely to report a higher pleasantness rating if the song had been independently predicted to be 'known' relative to songs predicted to be 'unknown'. *Genre* was also a significant predictor of pleasantness rating; when comparing each genre of music to nonsense songs, classical music was 37.6 times more likely to elicit a higher pleasantness rating, and

Table 3.5 Model$_{(pupil)}$1 regression findings of log maximum pupil response predicted by pleasantness ratings, setting, gender, age, average gaze position and trial number. Coefficients (exponentiated beta values) represent maximum percentage change in baselined pupil response.

Predictor variables	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.371 [1.277, 1.471]	<0.001**
Pleasantness Rating: Very unpleasant	1.010 [0.987, 1.034]	0.378
Pleasantness Rating: Unpleasant	1.010 [0.992, 1.028]	0.280
Pleasantness Rating: Pleasant	1.013 [1.002, 1.023]	0.017*
Pleasantness Rating: Very pleasant	1.033 [1.021, 1.046]	<0.001**
Setting: Wellcome	0.946 [0.929, 0.963]	<0.001**
Trial number	0.999 [0.999, 1.000]	0.025*
Age (unit: decades)	0.982 [0.976, 0.988]	<0.001**
Average gaze X (unit: 100 pixels)	0.994 [0.988, 1.000]	0.056
Average gaze Y (unit: 100 pixels)	0.998 [0.994, 1.002]	0.329
Gender: Male	1.024 [1.014, 1.035]	<0.001**
Number of observations	*1343*	

1960s and 1970s songs were equally (both ~10 times) more likely to elicit a higher pleasantness rating compared with nonsense songs. Neither the proximity of a song's release date to a participant's putative *reminiscence bump* nor *trial number* had a significant predictive effect upon pleasantness ratings.

Table 3.6 Model(rating)1 regression findings of pleasantness ratings predicted by a priori familiarity indicator, genre, trial number and reminiscence bump

Predictor variables	Odds Ratio [95% CI]	p-value
A priori familiarity indicator (odds of a higher pleasant rating when song was *a priori* predicted to be known)	7.720 [3.235, 18.423]	<0.001**
1960s songs (compared with nonsense songs)	10.126 [3.106, 33.010]	<0.001**
1970s songs (compared with nonsense songs)	10.116 [3.124, 32.763]	<0.001**
Classical songs (compared with nonsense songs)	37.592 [8.007, 176.476]	<0.001**
Reminiscence bump song release	1.212 [0.912, 1.612]	0.186
Trial number	0.992 [0.981, 1.004]	0.180
SD (Intercept Participant ID)	2.900	<0.001**
SD (Intercept Song ID)	2.572	<0.001**
Number of observations	*1486*	

These findings indicated that the nonsense songs (*a priori* predicted to be not familiar) may have inadvertently been rated as more unpleasant than their other-genre counterparts, and this could be driving the pupillometry results. Nonsense songs were removed from the sample, and the pupillometry model was re-run; the findings however remained unchanged from the original Model(pupil)1.

3.4.2.3 Pleasantness response associations with familiarity ratings

CLMM modelling (Model(rating)2) revealed that the odds of a song eliciting a higher rating on the pleasantness scale increased with each level increase in familiarity (indicated in Figure 3.5). The proximity of a song's release date to a participant's putative reminiscence bump did not have a significant predictive effect upon pleasantness ratings ($p = 0.318$), however this model revealed a trend towards an effect of trial number on pleasantness ratings ($p = 0.051$, see Table 3.7). The Kendall's Tau coefficient was 0.60 ($p < 0.001$), indicating a moderate-to-strong positive correlation between the two variables.

Table 3.7 Model$_{(rating)}$ 2 regression findings of pleasantness ratings predicted by self-reported familiarity, reminiscence bump and trial number

Predictor variables	Odds ratio [95% CI]	p-value
Songs self-reported as slightly familiar	2.176 [1.462, 3.237]	<0.001**
Songs self-reported as somewhat familiar	4.277 [2.682, 6.823]	<0.001**
Songs self-reported as moderately familiar	11.436 [6.946, 18.831]	<0.001**
Songs self-reported as extremely familiar	59.381 [35.616, 99.006]	<0.001**
Reminiscence bump song release	1.164 [0.864, 1.570]	0.318
Trial number	0.988 [0.977, 1.000]	0.051
SD (Intercept Participant ID)	2.900	<0.001**
SD (Intercept Song ID)	4.564	<0.001**
Number of observations	*1486*	
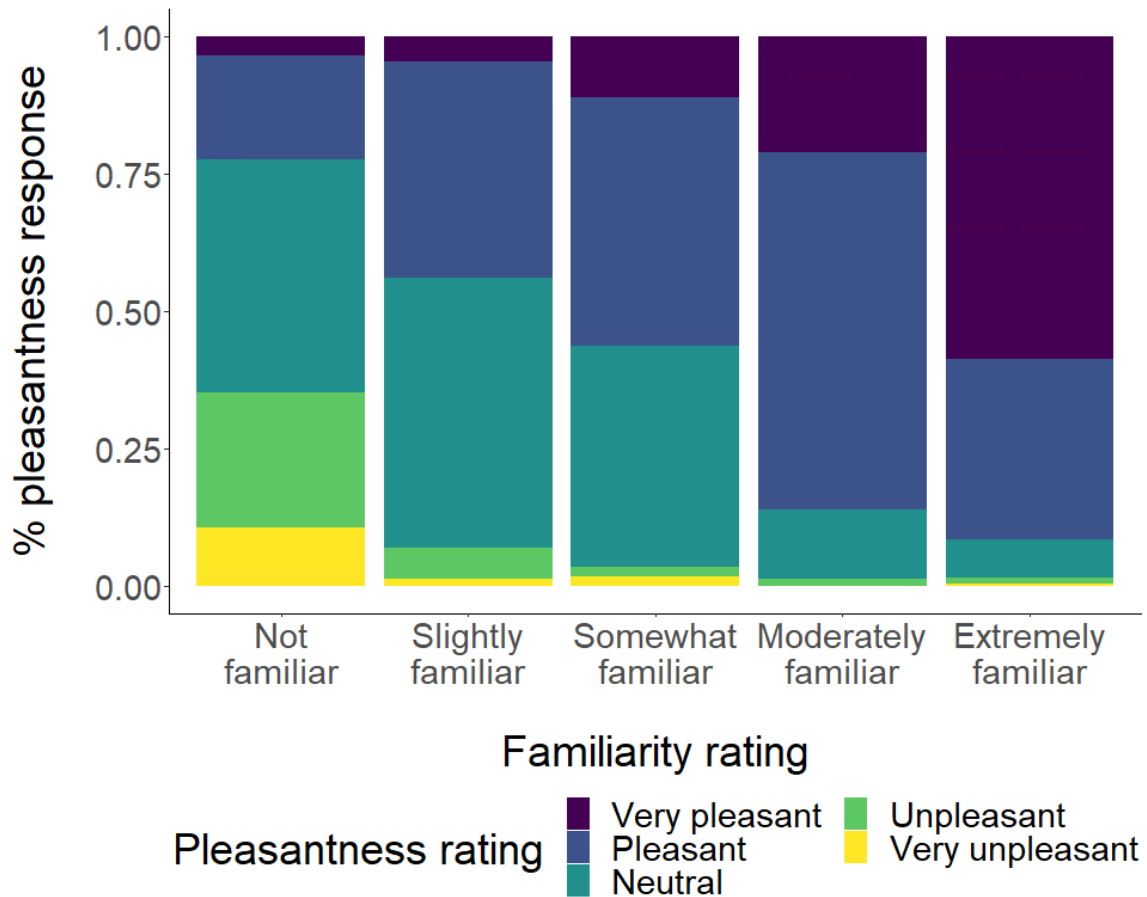
Figure 3.5 Stacked bar chart indicating the different profiles of pleasantness response at each level of familiarity rating in healthy older adults.

3.4.3 Familiarity responses to music

3.4.3.1 Pupil-familiarity responses based on a priori familiarity prediction
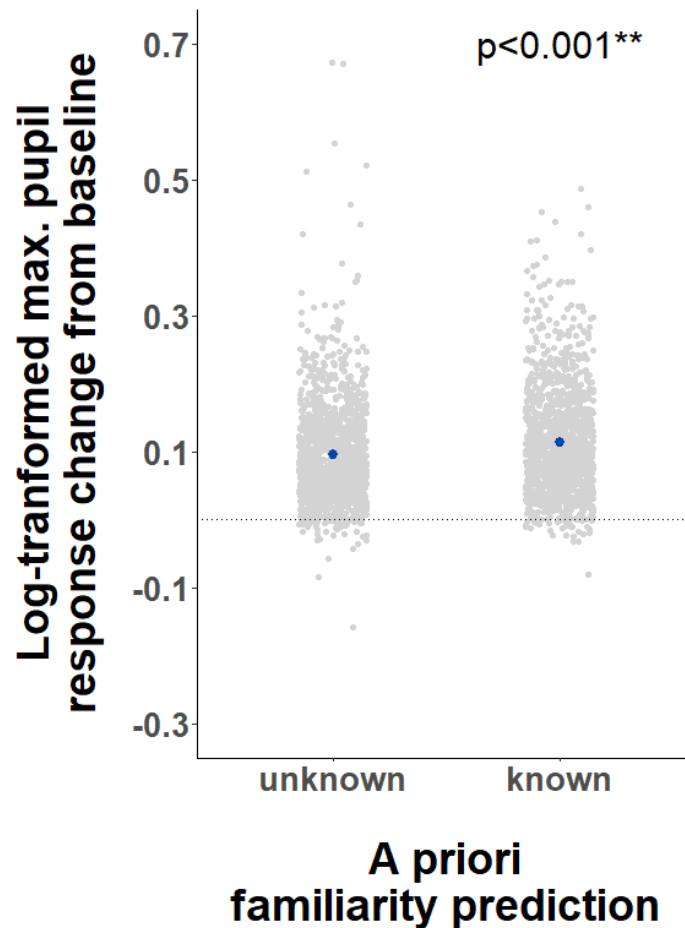
Figure 3.6 Older adults' maximum pupil response change from baseline plotted for music excerpts a priori predicted as either *known* or *unknown.* Points (·) indicate individual trial maximum pupil size change from baseline. Sample mean, lower and upper confidence intervals in blue. Dotted line indicates zero pupil size change from baseline. P-values indicate significance levels of the association between a priori prediction and maximum pupillometry response change from baseline, comparing a priori predicted *known* and *unknown* songs.

The $Model_{(pupil)}2$ intercept indicated a statistically significant larger $pupil_{max}$ response change from baseline (21.0[15.1, 27.1]%) elicited while participants listened to reference music excerpts (songs which were *a priori* predicted to be unknown, p<0.001, see Table 3.8). Relative to this, participants elicited a statistically significant larger $pupil_{max}$ response of 23.2[22.5, 24.0]% while listening to songs which were *a priori* predicted to be known (p<0.001; see Figure 3.6). *Setting, gender, trial number* and *age* were also significantly associated with $pupil_{max}$ response, in line with $Model_{(pupil)}1$ findings. *Pupil gaze position on the Y-axis* was found to be additionally

significantly associated with a larger pupil$_{max}$ response per unit increase (0.4[0.1, 0.8]%, p = 0.006).

Table 3.8 Model$_{(pupil)}$2 regression findings of log maximum pupil response predicted by *a priori* familiarity indicators, setting, gender, age, average gaze position and trial number. Coefficients represent maximum percentage change in baselined pupil response.

Predictor variables	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.210 [1.151, 1.271]	<0.001**
Songs *a priori* predicted as 'known'	1.018 [1.012, 1.025]	<0.001**
Setting: Wellcome	0.941 [0.925, 0.958]	<0.001**
Trial number	1.000 [0.999, 1.000]	0.021*
Age (unit: decades)	0.995 [0.991, 0.999]	0.008**
Average gaze X (unit: 100 pixels)	0.997 [0.993, 1.002]	0.274
Average gaze Y (unit: 100 pixels)	1.004 [1.001, 1.008]	0.006**
Gender: Male	1.007 [1.001, 1.014]	0.025**
Number of observations	*2542*	

The Model$_{(pupil)}$3 intercept revealed a significantly larger pupil$_{max}$ response change from baseline whilst healthy older adults listened to reference music excerpts (songs which were subjectively rated as not familiar, p<0.001, see Table 3.9). A linear familiarity-pupil relationship was observed; pupil$_{max}$ responses increased in accordance with each level of familiarity rating (see Figure 3.7). These larger pupil$_{max}$ responses reached statistical significance whilst participants listened to music excerpts subsequently rated as either 'moderately familiar' (21.6[20.4, 22.9]%, p = 0.006) or 'extremely familiar' (22.7[21.8, 23.6]%, p<0.001), relative to pupil$_{max}$ responses whilst listening to music excerpts subsequently rated 'not familiar' (19.9[14.0, 26.1]%). *Experimental setting*, *gender, age*, *trial number* and *pupil gaze positive on the Y-axis* were also significantly associated with pupil$_{max}$ response, in line with Model$_{(pupil)}$2 findings.

3.4.3.2 Pupil-familiarity responses based on memory evocation ratings

The Model$_{(pupil)}$4 intercept indicated a statistically significant larger pupil$_{max}$ response change from baseline (20.6[14.7, 26.8]%) elicited while participants listened to

reference excerpts (songs which participants rated as not evoking a memory) (p<0.001; see Table 3.10). Relative to this, participants elicited a statistically significant larger $pupil_{max}$ response of 22.8[21.9,23.5]% while listening to songs which evoked a memory (p<0.001; see Figure 3.8). Interestingly, regression coefficients revealed the

3.4.3.3 Pupil-familiarity responses based on subjective familiarity ratings
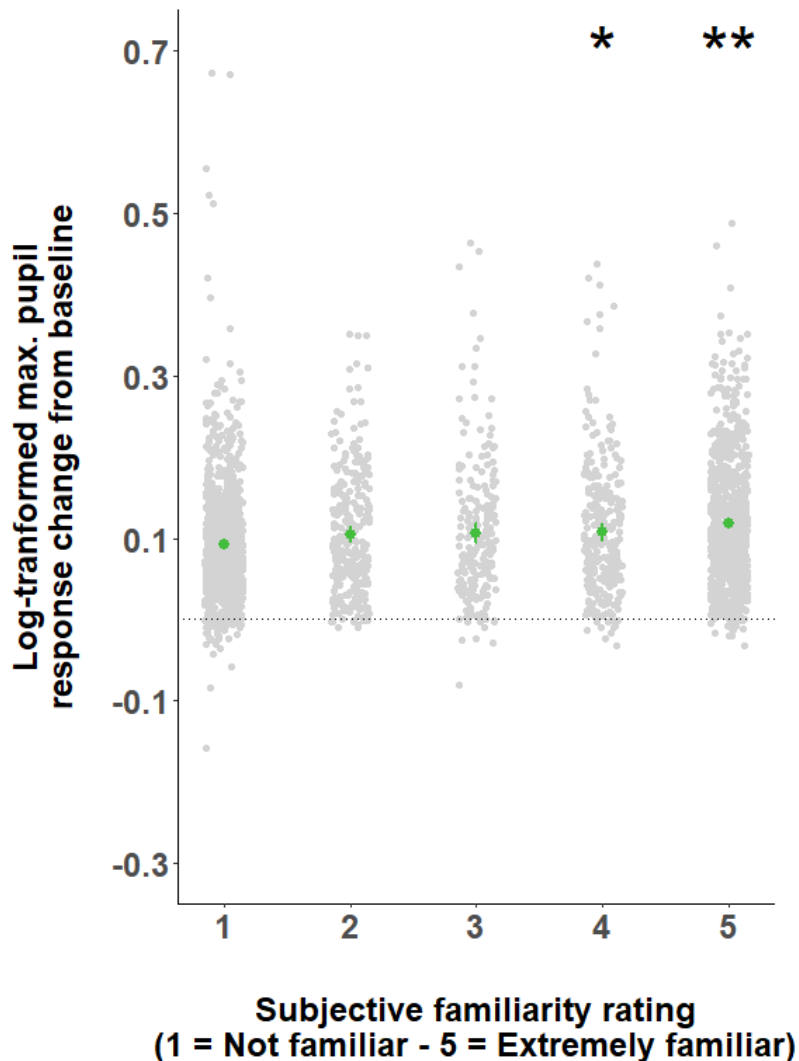
Figure 3.7. Maximum pupil response change from baseline plotted for music excerpts subjectively rated for familiarity (1[not familiar] to 5[extremely familiar]). Points (·) indicate individual trial maximum pupil size change from baseline. Sample mean, lower and upper confidence intervals in green. Dotted line indicates zero pupil size change from baseline. Asterisks indicate within-group significance level of the association between each subjective familiarity rating on maximum pupillometry response change from baseline, comparing familiarity levels with items rated *not familiar*.

Table 3.9 Model$_{(pupil)}$3 regression findings of log maximum pupil response predicted by familiarity ratings, gender, age, setting, average gaze position and trial number. Coefficients represent maximum percentage change in baselined pupil response.

Predictor variables	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.199 [1.140, 1.261]	<0.001**
Familiarity Rating: Slightly familiar	1.008 [0.998, 1.018]	0.124
Familiarity Rating: Somewhat familiar	1.010 [0.997, 1.023]	0.140
Familiarity Rating: Moderately familiar	1.015 [1.004, 1.025]	0.006*
Familiarity Rating: Extremely familiar	1.023 [1.016, 1.031]	<0.001**
Setting: Wellcome	0.943 [0.926, 0.959]	<0.001**
Trial number	1.000 [0.999, 1.000]	0.028*
Age (unit: decades)	0.996 [0.992, 1.000]	0.036*
Average gaze X (unit: 100 pixels)	0.997 [0.993, 1.002]	0.241
Average gaze Y (unit: 100 pixels)	1.004 [1.001, 1.007]	0.008*
Gender: Male	1.008 [1.002, 1.014]	0.015*
Number of observations	*2542*	

pupil$_{max}$ increase elicited whilst listening to songs which evoked a memory (1.8[1.1,2.4]%) were similar to the increased response observed for songs which were *a priori* predicted to be known (1.8[1.2,2.5]%, see Figure 3.8). *Experimental setting*, *gender, age*, *trial number* and *pupil gaze positive on the Y-axis* were also significantly associated with pupil$_{max}$ response, in line with Model$_{(pupil)}$2-3 findings.

3.4.4 Other findings of note

3.4.4.1 Self-report familiarity responses predicted by a priori familiarity indicators

An initial CLMM model (Model$_{(rating)}$3a) indicated strong evidence that the odds of a higher familiarity response were dependent on *a priori* familiarity prediction (p<0.001), but not trial number (p = 0.367). Additional covariates added in Model$_{(rating)}$3b of genre and reminiscence bump at song release marginally improved the AIC value, indicating a better model fit with the additional variables. This model, outlined in Table 3.11 below, demonstrated a similarly strong effect of *a priori* familiarity predictor that participants were 81.1[41.26,159.3] times more likely to rate a song as familiar if it had

been *a priori* predicted to be 'known'. In, addition a strong effect of genre on the familiarity rating was observed when comparing each genre to nonsense songs. No effect of reminiscence bump song release ($p = 0.169$) or trial ($p = 0.367$) was demonstrated.
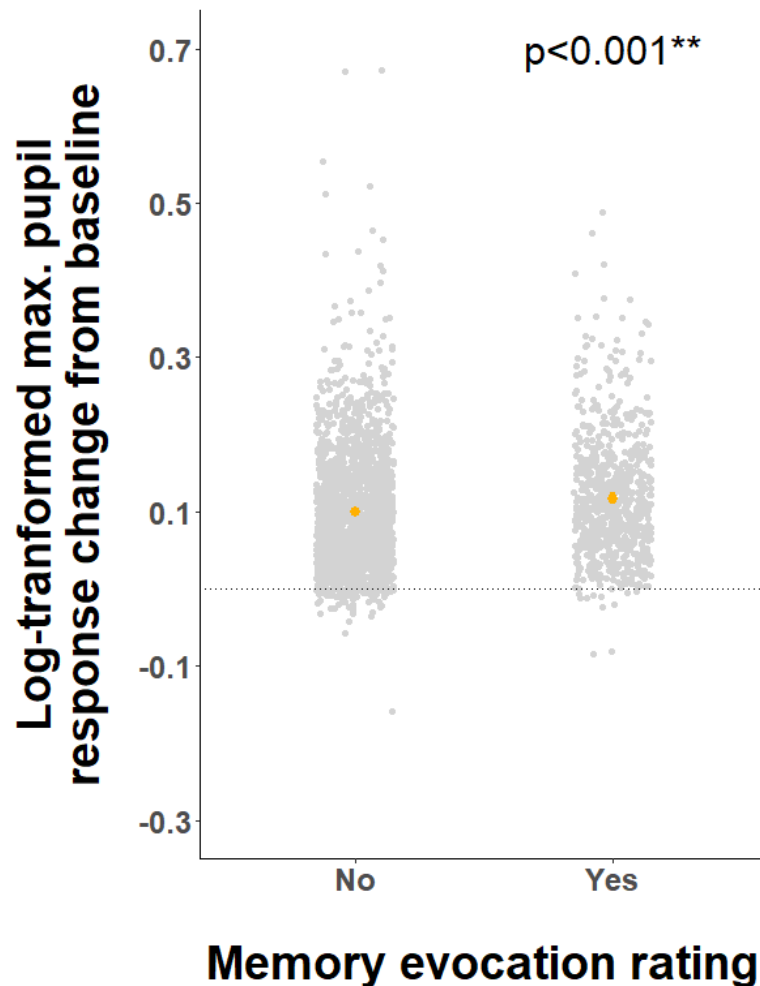
Figure 3.8. Maximum pupil response change from baseline for music excerpts subjectively rated as evoking a memory (*yes*) or not (*no*). Points (·) indicate individual trial maximum pupil size change from baseline. Sample mean, lower and upper confidence intervals in orange. Dotted line indicates zero pupil size change from baseline. P-values indicate within-group significance levels of the association between memory evocation rating on maximum pupillometry response, comparing *songs evoking memories* with *songs not evoking memories.*

Table 3.10 Model$_{(pupil)}$4 regression findings of log maximum pupil response predicted by memory evocation rating, gender, age, setting, average gaze co-ordinates and trial number. Coefficients represent maximum percentage change in baselined pupil response.

Predictor variables	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.206 [1.147, 1.268]	<0.001**
Songs evoking a memory	1.018 [1.011, 1.024]	<0.001**
Setting: Wellcome	0.942 [0.926, 0.959]	<0.001**
Trial number	1.000 [0.999, 1.000]	0.024*
Age (years)	0.995 [0.991, 0.999]	0.017*
Average gaze X (unit: 100 pixels)	0.998 [0.993, 1.002]	0.317
Average gaze Y (unit: 100 pixels)	1.005 [1.002, 1.008]	0.004*
Gender: Male	1.007 [1.001, 1.014]	0.025*
Number of observations	*2542*	

Table 3.**11** Model$_{(rating)}$ 3b regression findings of familiarity ratings as predicted by *a priori* familiarity indicators, genre, reminiscence bump song release and trial number

Predictor variables	Odds ratio [95% CI]	p-value
A priori familiarity indicator (odds of a higher pleasant rating when song was *a priori* predicted to be known)	81.081 [41.261, 159.332]	<0.001**
Classical songs (compared with nonsense songs)	10.134 [3.146, 32.642]	<0.001**
1960s songs (compared with nonsense songs)	8.610 [3.465, 21.397]	<0.001**
1970s songs (compared with nonsense songs)	4.406 [1.773, 10.945]	0.001*
Reminiscence bump (was the song released between 13 and 19 inclusive)	1.173 [0.935, 1.472]	0.169
Trial number	0.996 [0.986, 1.005]	0.367
SD (Intercept Participant ID)	3.371	<0.001**
SD (Intercept Song ID)	2.031	<0.001**
Number of observations	*2700*	

3.4.4.2   Self-report memory evocation responses: Associations with subjective familiarity ratings and genre

Modelling memory evocation ratings using subjective familiarity ratings in their original format (e.g. using the five levels of familiarity) was not recommended due to the low cell counts for songs which evoked a memory but were rated as not familiar (Table 3.12, Figure 3.9). Rescaling the variable did not reconcile this; nevertheless, the proportion of positive memory evocation responses increase with each level of familiarity, indicating the task was well comprehended. As a result, the variable was recoded, whereby songs rated as *not familiar* or *slightly familiar* were grouped into '*unfamiliar'* songs, and songs eliciting somewhat - extremely familiar songs were grouped as '*familiar'* songs. Recoding the variable revealed that songs rated as familiar were significantly associated with higher odds (19.72[12.18,31.93]) of songs being reported to have evoking a memory relative to songs recoded as *not familiar.*

Table 3.12 Songs reported to have evoked a memory, split by subjective familiarity rating levels

	Familiarity Rating						Total
	Not familiar	**Slightly familiar**	**Somewhat familiar**	**Moderately familiar**	**Extremely familiar**		**Total**
Song rated as evoking a memory	25 (3%)	28 (4%)	45 (6%)	116 (15%)	558	(72%)	**772(100%)**

To ascertain whether *genre* was associated with memory evocation ratings, a logistic regression predicting memory evocation from *genre, reminiscence bump* and *trial number* (with random effects of participant and song) was undertaken on a subset of data retaining the genres 1960s, 1970s and classical music. This revealed that, relative to classical songs, 1970s songs were associated with a lower likelihood of evoking a memory (β = -2.11 [-3.96, -0.25], p = 0.025); no association was demonstrated for 1960s songs (β = -1.562 [-3.41 0.28], p = 0.097).

3.4.4.3   Associations between behavioural/psychophysiological responses and musical properties

Spearman's rho correlation examining the relationship between BPM and pleasantness ratings revealed a significant but weak negative association between excerpt tempo and pleasantness ratings (r_s = -0.21, p <0.001). A simple linear regression indicated that older adults' pleasantness ratings were associated with musical excerpts which contained lyrics (β = 0.273, p = <0.001). Music key (major/minor) was not associated with pleasantness ratings (p = 0.550) (see Figure 3.10).
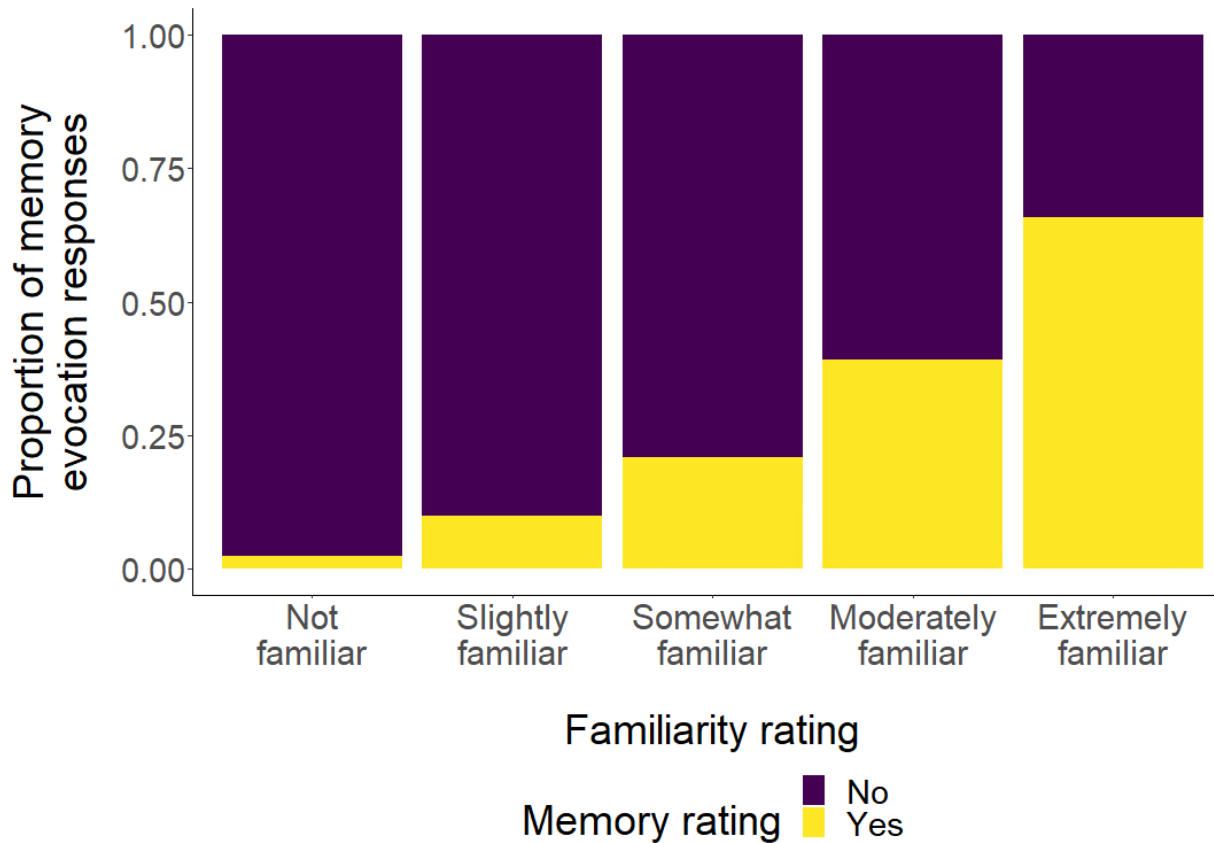
Figure 3.9 Proportion of memory evocation responses (yes/no) from healthy older adults, split by subjective familiarity rating levels.

101
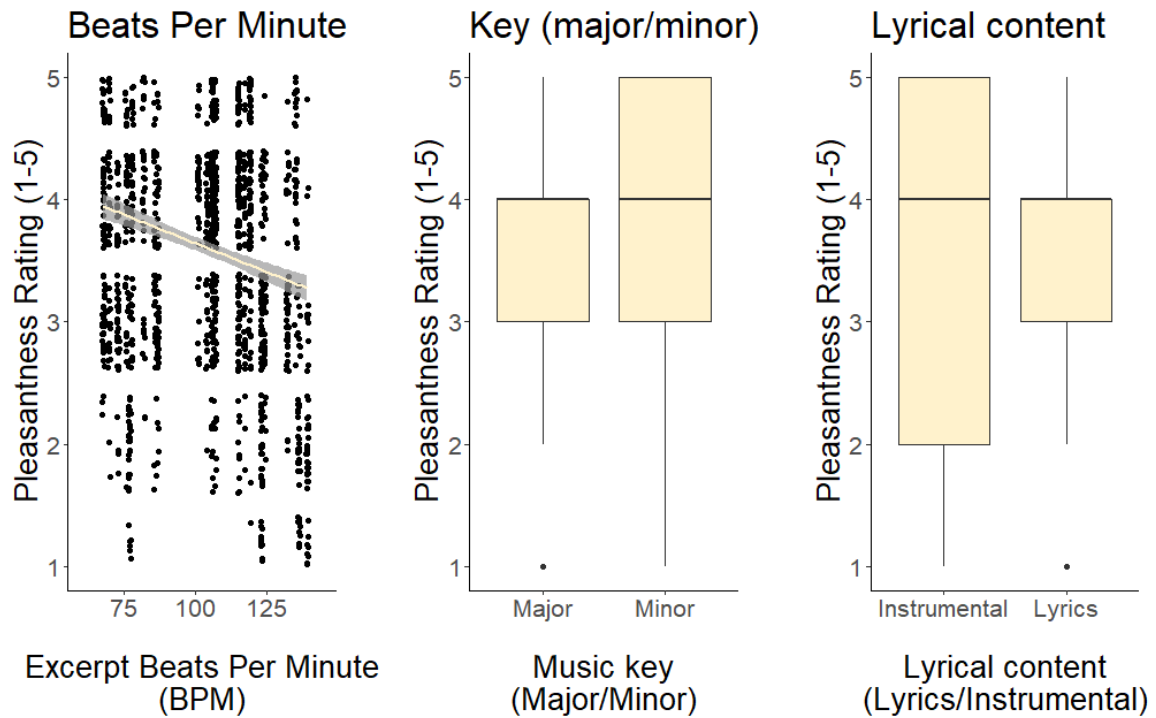
Figure 3.10. Representation of musical property influence on pleasantness responses. Left figure indicates BPM for each song split by pleasantness ratings in healthy older adults, with regression line indicating a negative association between pleasantness ratings and increased song tempo (shaded area represents the 95% confidence interval). Points (·) indicate individual pleasantness responses. The middle and right-hand figures represent box and whisker plots of pleasantness ratings split in terms of musical key (minor/major) and lyrical content (instrumental/lyrical).

Spearman's rho correlation examining the relationship between BPM and familiarity self-report ratings revealed a significant but weak negative association between excerpt tempo and familiarity ratings (r_s = -0.17, p <0.001). Neither music key (major/minor) (p = 0.103), nor lyrical content (p=0.115) were associated with familiarity ratings (see Figure 3.11). Separate linear regressions revealed no significant associations between excerpt tempo (p = 0.757) music key (major/minor; p = 0.182) or lyrical content (lyrics/instrumental) (p = 0.374) on pupillometry responses.
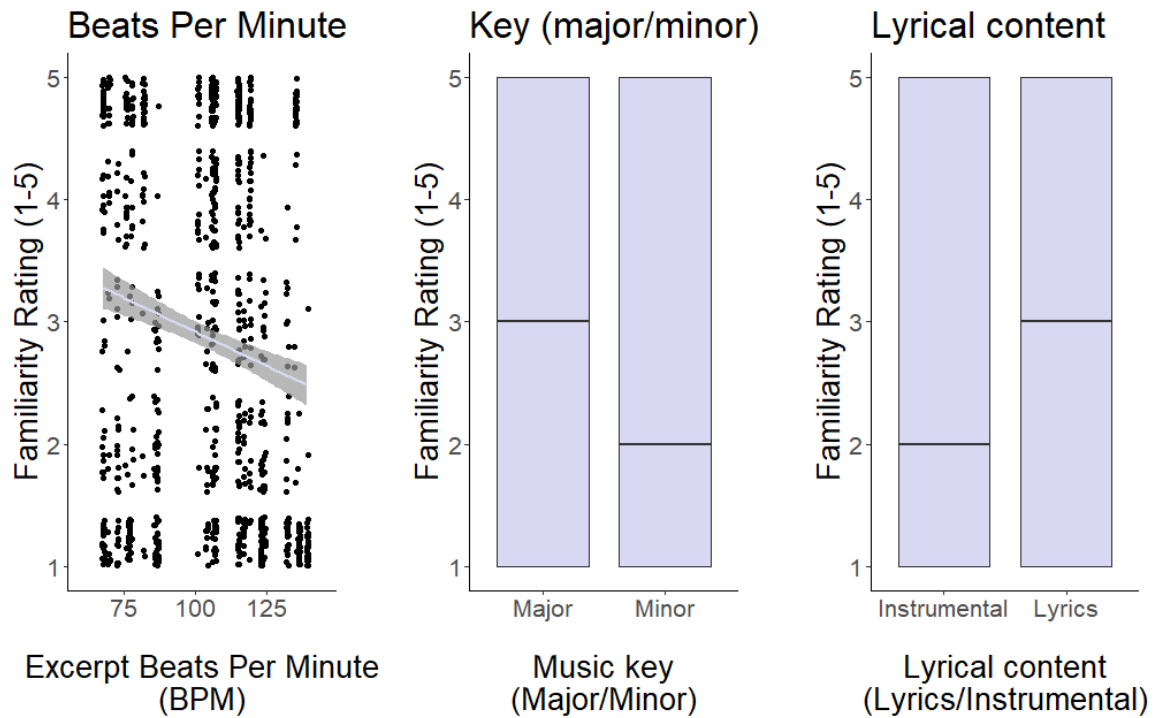
Figure 3.11 Representation of musical property influence on familiarity responses. Left figure indicates BPM for each song split by familiarity ratings in healthy older adults, with regression line indicating a negative association between familiarity ratings and increased song tempo (shaded area represents the 95% confidence interval). Points (·) indicate individual familiarity responses. The middle and right-hand figures represent box and whisker plots of familiarity ratings split in terms of musical key (minor/major) and lyrical content (instrumental/lyrical).

3.5   DISCUSSION

The current study findings demonstrate evidence of differential pupillometry responses to music in healthy older adults, observable in relation to song familiarity (regardless of whether predicted by a research team, subjectively rated, or classified in terms of evoking a memory) and song pleasantness. This is, to the author's knowledge, the largest pupil-auditory study in healthy older adults, and the first to demonstrate associations between pupil and pleasantness/familiarity responses of music in this population. The intercepts demonstrated in three of the four models indicated an approximate 20% change in pupil response from baseline, reflecting expected cognitive changes in pupil size, which are typically modest (20%/0.5mm) compared to illumination-related responses (up to 4mm) (J. Beatty & Lucero-Wagoner, 2000). Encouragingly, these within-normal-range findings were observed in maximum pupil responses relative to a pre-trial baseline, a standard pupillometry pre-processing

103

technique which has been previously argued may mask the possibility of observing psychosensory response effects in older adults (Piquado et al., 2010).

3.5.1 Interpretation of results

[RQ1] Are healthy older adults' self-reported pleasantness music ratings associated with pupillometry change from baseline?

RH1 was accepted; pupil responses were comparatively larger to songs rated as pleasant or unpleasant relative to pupil responses observed for songs participants rated as 'neutral'. Both regression modelling approaches to address this research question corroborated the curvilinear emotion-pupil relationship demonstrated in previous studies sampling from neurotypical younger adults (Partala & Surakka, 2003) and in a smaller sample of healthy older adults in response to emotional sounds (Fletcher, Nicholas, et al., 2015b). The current findings therefore extend the pupil-pleasant curvilinear phenomena (previously demonstrated using affective sounds) to pre-selected music, and add support to the evidence base that healthy older pupil responses retain a preserved profile of differential autonomic responses to emotionally arousing music. Two statistical approaches were selected; modelling the curvilinear relationship, and looking at each pleasantness level separately. This revealed a potential valence effect in addition to the curvilinear relationship, by virtue of the maximum pupil response being significantly increased for pleasant and very pleasant songs relative to neutral songs; whilst the pupil size increased for unpleasant and very unpleasant songs, this increase did not significantly differ to neutral song pupil responses. The finding of a pupil-pleasantness relationship in the current study shows promise for extending this paradigm to clinical populations.

[RQs 2-4] Is a song's familiarity associated with healthy older adults' pupil response change from baseline?

In line with previous evidence in younger neurotypical populations, the pupil old/new effect in healthy older adults was demonstrated by a larger maximum pupil response when listening to: songs which were *a priori* categorised as familiar by the research team [RQ2]; songs which participants subjectively rated as 'moderately' or 'extremely' familiar [RQ3] and/or songs which participants self-reported to have evoked a memory

[RQ4]. RHs 2-4 were therefore able to be accepted, in line with previous findings in younger neurotypical populations. Interestingly, the use of the Likert scale subjective rating of familiarity revealed a linear relationship between pupil response and level of reported familiarity. It may be that the typical approach of acquiring familiarity data through binary familiar/unfamiliar ratings, or bisecting Likert-scale data at the analysis stage has masked this effect in previous studies, however it appears as though an increasing feeling of familiarity with a piece of music creates a perhaps cumulative effect on pupil response. This, if replicable in clinical populations, would be a helpful proxy for establishing music familiarity in PLWD.

[RQs 5-6] Are independent classifications of a song's familiarity [RH5] and/or subjective ratings of a song's familiarity [RH6] associated with older adults' pleasantness ratings?

An exposure-affect link was demonstrated in subjective musical ratings; familiarity was a significant predictor of a higher pleasantness rating. This link was observed regardless of whether the songs had been independently rated for familiarity (indicating acceptance of RH5) or based on subjective familiarity ratings (indicating acceptance of RH6). Acceptance of RH5 indicates that, at least in healthy older adults, it is possible for an *a priori* predicted level of familiarity to predict how pleasant a person subjectively rates a song. Interestingly, for RH6, the odds of rating a song as higher in pleasantness increased with each additional level of familiarity rating, all reaching statistical significance when compared with pleasantness ratings for subjectively-rated unfamiliar songs. The latter of these findings is indicative of the *mere exposure effect* (MEE) in the current sample, in that the odds of rating a song as pleasant is significantly higher even if the person has presumably only been exposed to the song a handful of times. This has been supported by reports of positive correlations between the number of times a song is played and its subsequent preference ratings (Madison & Schiölde, 2017), and notably contributes to the limited literature on MEE demonstrated in older adults (Gaudreau & Peretz, 1999; Halpern & O'Connor, 2000).

[RQ7] Are independent classifications of a song's familiarity associated with older adult's likelihood of reporting that song as familiar [RH7]?

The indication that *a priori* 'familiar' songs were over 80 times more likely to be rated higher in familiarity compared to songs previously categorised as 'unknown' indicates

RH7 can be accepted. This result indicates that the stimuli selection in the experimental design achieved an accurate categorisation of a song being rated as familiar or unfamiliar in a healthy older population; we can be confident that in terms of the familiarity findings, the paradigm demonstrates face validity. The acceptance of RH7 indicates that that the paradigm could be utilised for further study of familiarity processing in pupillometry studies in clinical populations.

[RQ8] Are older adults' subjective judgements of a song's familiarity associated with their likelihood of reporting that song as evoking a memory?

The finding that songs which were subjectively rated as familiar were significantly more likely to be rated as evoking a memory increases confidence that the paradigm was well comprehended in the heathy older sample, owing to the low cell counts of songs which were deemed to be not familiar, but evoking a memory; RH8 was therefore accepted. Whilst any incidence of the latter observation ('not familiar' rating → positive memory evocation rating) may appear counterintuitive, anecdotal evidence from the data collection stage indicated that participants did tend to qualify these incidences with the fact that the songs "reminded them of something". This underscores both the ability for music to inadvertently elicit memories even by virtue of a similarity in melody or other properties to a known song, and the complexity in creating a music stimuli set which does not invoke feelings of familiarity that arise indirectly.

[RQ9] Are songs released within an individual's reminiscence bump associated with subjective music ratings?

There was no demonstrated effect of song release in reminiscence bump in any modelling of the subjective music ratings (and thus a rejection of RH9). This raises an important consideration for independently selecting music as familiar for an older adult, with a potential application for PLWD. The putative music reminiscence bump period, selected from the range outlined in Rao et al. (2021), was observed for songs released aged between 13 and 19 years of age, for older as well as younger adults within the sample. However, the authors concede that the musical stimuli comprised popular music only, and for those participants who did not report themselves to be interested in popular music, there was no pattern of age on reminiscence of the stimuli. We similarly attribute the lack of effect of reminiscence bump observed in the current study

to the inclusion of well-known classical songs, present throughout the lifespan and thus immune to any putative reminiscence bump time period. This finding supports the inclusion of musical stimuli outside the reminiscence bump in the current study, and adds to the evidence base that musical properties other than familiarity may be contributing to the appraisal of a musical stimulus.

3.5.2 Study strengths

This study and paradigm described sought to address some of the methodological limitations highlighted within the current literature; namely stimulus characterization and treatment. The findings contribute to the limited literature relating to psychosensory pupillometry responses in healthy older adults, and the replicated findings of well-documented psychosensory pupil response profiles justify the extension of the Play It Again paradigm to clinical populations. The participant group was well-characterised, with a large sample relative to other studies of age-related pupil psychosensory responses (Benhamou et al., 2021; Fletcher, Nicholas, et al., 2015b; Piquado et al., 2010; Zhao et al., 2019).

The reported music preferences seem to reflect predicted profiles of musical taste in accordance with the age profile of the current sample. A longitudinal ten-year study of the listening habits of over 250,000 people revealed that preferences for music dimensions such as 'sophistication' and 'unpretentiousness' increases with age, which factor loadings following exploratory structural equation modelling revealed contained significant correlations with classical music (associated with the sophisticated factor) and popular music (associated with the unpretentious factor) (Bonneville-Roussy et al., 2013).

Several findings in the study provide further assurances of the paradigm's suitability for application to clinical groups. The overall finding of no significant effect of trial number on subjective music ratings (noting the trend of trial length in Model$_{(rating)}$2) suggests that it is unlikely that participants were perseverating in their ratings responses owing to experiment length (i.e. participants remained engaged with the task throughout the procedure). This suggests the paradigm length is relatively low-burden to healthy older adults to complete. In addition, participants' PDR

characteristics had a similar influence on pupillometry as reported in Zhao et al. (2019), indicated by main effects of age and trial index which, for each unit increase, pupil response decreased. Trial index-related reductions in pupil size could reflect gradual innervation of the parasympathetic nervous system, reducing the pupil size as a reflection of returning to homeostasis over the course of the experiment (Mathôt, 2018).

3.5.3 Study limitations

The study has several limitations; the first relating to a comment on the specificity and validity of pupillometry in relation to the valence finding discussed in Section 3.5.1..While the curvilinear pleasant-pupil relationship was significant in the current study, modelling pupil response from each level of self-reported pleasantness rating indicated that a larger maximum pupil response was only significantly associated with songs rated as either 'pleasant' or 'very pleasant' relative to songs rated as 'neutral'. This indicates an effect of valence as well as arousal on the pupil responses, which contrasts with more recent findings (e.g. Gingras et al., 2015). One possible confound is the potential different levels of arousal induced by each song. It is possible that songs rated as 'very unpleasant' or 'unpleasant' induced less autonomic arousal than songs rated 'pleasant' or 'very pleasant', leading to a higher PDR for the latter songs. It is not possible to test this interpretation further, as the current paradigm did not include a specific arousal rating question to accompany each excerpt.

Direction from previous work (Fletcher, Nicholas, et al., 2015b) had indicated that any future pupil-emotion research should investigate the complex interrelationship between valence and arousal by additional rating questions subsequent to stimulus presentation. This was considered in the current paradigm; however, it is possible that the concepts of 'valence' and 'arousal' may be too challenging to distinguish and respond to for PLWD. It is important that any psychophysiological paradigm coupling stimulus presentation with contiguous rating questions reduces the extent to which cognitive appraisal needs to occur in order to respond to the questions to appraise the stimuli. This increase in task difficulty, owing to processing complex concepts, may lead to an overall increase in pupil responses reflecting cognitive load, which have the possibility of dominating any effects associated with the musical properties of interest (Van Gerven et al., 2004). The relatively simple pleasantness phrasing was therefore

established in the current paradigm in an effort to prospectively reduce the cognitive demands of PLWD undertaking the task. Previous psychophysiological studies adapted from neurotypical paradigms for PLWD (e.g. Barradas et al., 2021) have aimed to reduce cognitive load within similar psychophysiological paradigms by limiting ratings questions to measure 'felt' emotions, or reducing the number of points on proposed Likert scales. Asking the rating questions in this way also potentially extends the capacity to investigate potential physiological-behavioural coupling which has applications in a care context. Nevertheless, whilst this approach potentially has some methodological advantages, the relatively simplistic phrasing of the pleasantness question does make interpretation of the pleasantness-pupil coupling more challenging.

The unavoidable change in experimental setting towards the end of participant data collection was an initial cause for concern. While both rooms were uniformly lit, a luminance difference did exist between the two settings (University>Wellcome). This was reflected in the significant reduction in pupil response observed for participants in the lower luminance (Wellcome) setting compared to the university setting. This aligns with the principle that the pupils of participants experiencing the paradigm at Wellcome had already dilated to a larger extent by virtue of being in lower luminance, and therefore may have demonstrated lower reactivity to the musical stimuli as their baseline pupil size was closer to their physical maximum than for participants in the higher luminance (university) setting. Steps were taken in the pre-processing pipeline to minimise the influence of this factor (e.g. divisive baselining rather than subtractive), and setting was subsequently included as a covariate in each pupil model. The fact that effects of interest were still observable indicates that this was less of a cause for concern than initially thought, nevertheless in future studies this should evidently be avoided.

Despite care and attention paid to the stimuli selection for the current study, the paradigm may have been strengthened by further exploration and control of additional acoustic properties. While the stimulus set were interrogated in terms of tempo, key and lyrical content, it is a possibility that further candidate musical properties not explored in the current stimulus set were unbalanced between *a priori* familiarity music conditions, and therefore the resulting pupillary responses to familiar and unfamiliar

109

stimuli may in fact be owing to musical properties other than the intended experimental manipulation. The Music Information Retrieval (MIR) toolbox, developed by Lartillot and Toiviainen (2007) enables a comprehensive comparison of further acoustic properties, including: variability in sound amplitude, 'attack time' (time from initial note onset to the peak amplitude), pulse clarity (the stability of the beat), roughness (relating to timbre) and high-frequency energy (typically how 'bright' or 'dark' the music sounds). Future music paradigms developed would ideally subject potential stimuli to further testing of confounding properties using this or similar toolkits. Whilst the use of recorded, ecologically-valid music may not always make it possible to create a perfectly balanced musical corpus, this additional exploration of a proposed stimulus set would ensure that any potential confounding variables could be concurrently included in subsequent statistical modelling. Another potential confound relates to the volume of the stimuli set; whilst normalised for mean volume in line with recommendations (Mathôt, 2018), it is possible that the excerpts given a valence rating or higher familiarity rating contained different frequencies of peak sound amplitudes, which are likely to elicit pupil dilations in response to unexpected stimuli (Benhamou et al., 2021). Timecourse modelling of the data and establishing when maximum pupil responses were elicited could indicate the likelihood of whether or not these were in fact in response to different peak sound amplitudes rather than pleasantness and/or familiarity appraisals (Jagiello et al., 2019).

A final limitation is discussed in relation to the applicability of the response findings to potential clinical application. The pupil responses described in the current work have been explored and demonstrated separately, reflecting dissociations observed in the literature between music familiarity and emotional judgements. The fact remains however that outside of laboratory settings, our response to musical stimuli is a culmination of closely-linked emotional and cognitive processing, complex to disentangle (Hodges & Wilkins, 2015). The positive pleasantness/familiarity correlation of behavioural responses, and the significantly larger pupil responses to songs which were both subjectively rated as pleasant and familiar, indicates the possibility that the observed pupil old/new effect may be influenced by music pleasantness, or vice versa. Previous studies indicate that stimuli which are deemed to evoke emotional responses are more likely to be recalled (Emotional Enhancement of Memory; EEM), and in older adults this relates specifically to a 'positivity effect', in

contrast to younger adults who tend to encounter EEM in particular relation to negative stimuli (Joubert, Davidson, & Chainay, 2018).

The positivity effect of EEM in older populations appears to be reflected in the current findings, and this has ramifications for interpreting the relative contribution of emotional and familiarity appraisals observed in the pupillometry results. Any "interaction of both [emotional and familiarity] processes has been largely neglected" in prior pupillometry research (Võ et al. (2008), p.130). Võ et al. (2008) investigated this potential interaction by manipulating emotional content of words used to conduct a recognition memory task, whereby pupillometry responses were acquired during stimuli presentation and test phases. The observed pupil/old new effect was found to be mediated by the words' emotional valence (Võ et al., 2008). This complements findings outlined in van den Bosch et al. (2013), who demonstrated that familiarity may be in part modulating EDA in response to music that is subsequently rated as pleasurable. However, the old/new effect in Võ et al. (2008) was observed during the test phase only, which differs from the current findings observed during stimulus presentation. Therefore, the extent to which emotional appraisal mediated the observed pupil old/new effect remains unclear. Invoking a similar approach adopted by Gingras et al. (2015), whereby familiarity is controlled for and the stimulus set was selected specifically for its relatively low notoriety could elucidate the relative contributions of pleasantness and familiarity on pupil dilation in older adults during music listening.

In addition, whilst efforts were made in the current analysis to account for individual differences, it is possible that there may be individual differences in pupil response profiles (i.e. that some individuals may elicit larger pupil sizes in response to cognitive processing, and for others a larger pupil response may reflect emotional processing). From the perspective of future application of the work, it would be tempting to propose that a larger pupil size is likely to reflect a positive musical experience, regardless of whether due to higher familiarity, higher pleasantness, or both. However as indicated in previous literature (e.g. Barradas et al., 2021) familiarity is not always associated with higher liking ratings. An ideal outcome for applications to dementia care and support would be to demonstrate different physiological responses reflecting different and observable profiles of listening to a song that is familiar (liked), familiar (disliked), not familiar (liked) and not familiar (disliked). A future paradigm aiming to create these

different conditions would be helpful. However, while familiarity may be a relatively simple characteristic to manipulate for music paradigms, creating a playlist stimuli set whereby music is *a priori* selected to manipulate affect (e.g. replicating the approaches taken by Võ et al. (2008) for emotional words) is a challenge owing to the subjectivity of music preferences.

Acknowledging these caveats, the current findings suggest that it is possible, at least in healthy older adults, to observe pupillary effects reflecting familiarity and emotional music processing. Previous results of discernible physiological responses in PLWD indicates that there is merit in extending this paradigm to include clinical populations, particularly individuals with forms of Alzheimer's disease.

3.6   CONTRIBUTIONS & ACKNOWLEDGEMENTS

Using the Contributor Role Taxonomy (CRediT) as outlined in Holcombe (2019)

4   PLAY IT AGAIN FOR PEOPLE LIVING WITH TYPICAL AND ATYPICAL ALZHEIMER'S DISEASE

4.1   ABSTRACT

The second experiment in this thesis explored the Play It Again paradigm in patient populations following evidence of preserved pleasant-affect and pupil old/new effect responses to musical excerpts in healthy older controls. The study investigated whether these same differential pupillometry profiles could be established in people living with either early-mid stages of typical, memory-led Alzheimer's disease (tAD) or the atypical 'visual variant' of Alzheimer's disease (posterior cortical atrophy; PCA) for music rated in terms of pleasantness and familiarity. Patient groups [tAD=18; PCA=21] were recruited via the Specialist Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery, London, UK. Pupillometry responses were recorded whilst participants listened to the musical excerpts outlined in Section 3.3.3). The same background, basic clinical and music experience information was collected as in the previous chapter, alongside additional neuropsychological assessment. Separate within-group linear regression models obtaining cluster robust standard errors were run to predict pupil response from *a priori* predicted- and subjective-ratings of music familiarity and pleasantness, controlling for demographic factors including disease severity. Between-group differences and interactions were interrogated to explore potential differences in each clinical population vs. healthy overall pupil reactivity and psychophysiological responses. Results indicated significant within-group effects of pleasantness and familiarity on pupil response profiles in the tAD group in line with control findings, with larger pupil psychosensory responses relative to controls when listening to songs eliciting either a 'very pleasant', 'slightly familiar' or 'extremely familiar' subjective rating. Marginal within-group effects were observed within the PCA group in relation to larger pupil responses being associated with songs *a priori* predicted to be known, and songs which were subjectively rated as 'moderately familiar'; these effects were observed when disease severity was controlled for. Between-group analyses of pupil responses revealed that, for each music-pupillometry model, tAD overall pupil dilation responses did not differ significantly compared to controls (albeit one model elicited a marginal effect). In contrast, a significant difference was consistently observed between PCA and control overall

pupil dilation responses. The results are discussed in relation to the preliminary finding of relatively preserved differential pupillometry responses to valenced/neutral and familiar/unfamiliar music in the tAD group. Although the PCA results call into question the utility of pupillometry in individuals with significant oculomotor dysfunction, methodological adjustments in future work may reconcile this and enable further exploration of proposed neuroanatomical correlates for musical semantic memory. Pupillometry may therefore remain a useful non-invasive biosignal for establishing in-the-moment response to music in a variety of syndromes attributable to Alzheimer's disease pathology.

4.2   INTRODUCTION

4.2.1 Individualised music for people with a diagnosis of a dementia

With or without a dementia, optimal effects of music are observed when its presentation is tailored to our individual tastes and preferences (E. Johnston et al., 2017). The release of _Alive Inside,_ a feature-length documentary and testimonial to the positive effects of individualised music, sparked substantial public interest in receptive music activities for PLWD that do not require a qualified music therapist (Cohen & Rossato-Benett, 2014). Guidelines for use of music for PLWD champion the benefits of individualised music based on music preferences, and its particular application to reduce agitation by virtue of the music's provision to PLWD within a particular time-window when challenging behaviour is exhibited [the mid-range theory of individualised music intervention for agitation (IMIA); Gerdner (2012)]. This theory (and accompanying protocols) are widely cited, having been taken up by organisations such as Playlist for Life, a UK charity which promotes the creation, provision and use of personal music playlists for people with dementia and their families (Ranscombe, 2019).

Findings from numerous studies demonstrate benefits of individualised music listening in PLWD on a number of health and wellbeing outcomes. A recent study of 282 nursing home residents with dementia who listened to individualised playlists 1-3 times per

week for 6 months demonstrated significant associations with improved levels of agitation, reduction of depression, selective sundowning symptoms (disengagement) and general neuropsychiatric symptoms in the first three months of the intervention (Lineweaver et al., 2022). Following a systematic review and meta-analyses, Tsoi et al. (2018) asserted that the evidence supports music listening as being more effective for relieving behavioural and psychological symptoms of dementia BPSD than interactive music therapy; a stance upheld by Raglio et al. (2018). Whilst it is arguably time and cost-efficient to administer music listening activities for PLWD in a group setting in relation to 1:1 practice, music selection in this context cannot, for example, account for individual differences in music taste (North, 2010). Findings from the literature review and previous chapter suggest that different intra- and interpersonal emotional and cognitive appraisals of music may play a more important role in psychophysiological responses than perhaps other facets (e.g. music properties). It is therefore important we explore the possibility of extending the reach of the Play It Again paradigm to explore the extent to which psychophysiological response profiles of music can be demonstrated when differentially preferred or known in PLWD.

4.2.2   *Evidence for spared musical memory in Alzheimer's disease?*

An attractive clinical population to explore the potential for preserved physiological responses reflecting musical experience are individuals with Alzheimer's disease. The hallmark of this condition is impaired recollection, and yet, as highlighted in C. N. Clark and Warren (2015), "the power of music to unlock memories…in Alzheimer's disease is a cherished tenet of clinical neurology" (p. 2122). Preserved semantic musical memory has been identified in Alzheimer's disease relative to controls (Cuddy et al., 2012; Golden et al., 2017; Hsieh, Hornberger, Piguet, & Hodges, 2011; Vanstone et al., 2012). High-resolution functional magnetic resonance imaging (fMRI) studies - conducted firstly in younger healthy adults - revealed associations between long-known musical memory processing candidates, including higher regional activation of the caudal anterior cingulate and ventral pre-supplementary motor area, relative to activation when listening to unknown or recently-known music (excerpts played one hour before scanning) (Jacobsen et al., 2015). A second phase of this study investigated these cortical regions in 20 tAD participants using MRI and fluorodeoxyglucose positron emission tomography (PET) scanning, respectively

revealing minimal atrophy and disruption of glucose-metabolism in these areas relative to the rest of the brain (Jacobsen et al., 2015). A review of semantic musical memory by (Omar et al., 2012) however highlighted the challenges in assessing semantic musical memory in individuals with dementia, particularly those with language difficulties, which can co-occur in later-stage typical Alzheimer's disease. Given PCA is predominantly accounted for by underlying Alzheimer's disease pathology (Crutch et al., 2012), it may be informative to explore psychophysiological responses in both typical (memory-led) and atypical syndromic groups tested using the same musical paradigm. This investigation will elucidate whether syndromic differences give rise to any observable different within-group pupil response profiles to music, and/or in relation to healthy older adult findings.

4.2.3   Semantic musical memory preserved in individuals with posterior cortical atrophy and memory-led Alzheimer's disease?

Initial case studies investigating recognition of 15 familiar melodies revealed correct identification of all melodies in one individual with PCA [Patient MA; Polk and Kertesz (1993)]. More recently, one study sought to explore syndromic differences in familiar music processing in Alzheimer's disease at group level using functional neuroimaging. Slattery et al. (2019) recruited participants with young-onset memory-led Alzheimer's disease (YOAD, N = 24), PCA (N = 10) and healthy age-matched control volunteers (N = 19). The authors created a stimulus repository of familiar and unfamiliar melodies, to which participants listened during 3T MRI scanning. Selected melodies from the familiar and unfamiliar conditions were presented either only once or twice to modulate the presentation frequency and induce episodic memory processing. A post-scan task probing familiarity judgements demonstrated a spared ability to identify familiar from unfamiliar music in both groups relative to healthy controls, but impaired performance on an episodic musical memory task (selecting three songs presented 60 seconds prior and judged as pleasant or unpleasant among nine foil melodies; Slattery et al., 2019). The neuroimaging data indicated that in the semantic memory task, individuals in the memory-led AD group had significantly reduced activation of the right inferior frontal gyrus (IFG) compared with the healthy control participants, however this was not observed in the PCA group and nor were any further significant differences between groups in neuroimaging signal during musical tasks reported. The lateralisation of familiar melody processing particularly in IFG has mixed findings,

although another fMRI study has revealed associated activation localised to left IFG (Groussard et al., 2019). Nevertheless, in the knowledge of demonstrated spared semantic musical memory in Slattery et al. (2019), both patient cohorts appear to be candidates for further investigation. The Play It Again paradigm, incorporating familiarity metrics of both degree (not familiar - extremely familiar) and depth (Did the song evoke a memory?) will enable further exploration in relation to psychophysiological responses to known and unknown music in both groups.

4.2.4 Oculomotor changes in Alzheimer's disease

While neuroimaging studies with contiguous behavioural tasks have pointed to either retained activation relative to controls when processing familiar music (tAD group) or preserved musical semantic memory performance (tAD and PCA), it may be too great a stretch to anticipate any observable psychosensory response preservation in these clinical groups. In addition to the age-related pupil changes outlined in Chapter 3, pupillometry response profiles in people living with Alzheimer's disease are also thought to be compromised by virtue of the condition reflecting cholinergic deficiency (David & Malhotra, 2022). A recent scoping review of task-evoked pupil responses in individuals with dementia points to an overall reduction in pupil response profiles in tAD relative to healthy controls, likely owing to LC degeneration (as outlined in Section 2.1) (Zeeman et al., 2023). A reduced pupil flash response (PFR) profile is also observable for healthy older adults with neocortical amyloid burden (Frost et al., 2017). While the age-related physiological changes did not seem to deter psychosensory pupillometry effects in the healthy control group (both in the current findings and in previous work it is a possibility that the combination of physiological aging and autonomic dysfunction in the clinical groups may breach a threshold whereby all observable effects are attenuated (Fletcher, Nicholas, et al., 2015b). Fletcher, Nicholas, et al. (2015b) not only reported findings of a pupil-affect relationship in healthy older adults (as outlined in the previous chapter), but additionally explored the same paradigm in ten individuals with memory-led Alzheimer's disease, as well as individuals with FTD subtypes). Encouragingly, normal overall pupil reactivity was observed in the tAD group only, whereas this was reduced in other patient groups. The same curvilinear pupil-affect response observed in the healthy older control group was not however demonstrated in the AD group, despite mean valence ratings being

significantly correlated between the AD group and healthy age-matched controls ($p<0.001$; Fletcher, Nicholas, et al., 2015b). On the other hand, Benhamou et al. (2021) reported pupillometry results in the tAD group consistent with control findings (in relation to responses to musical expectancy). Between-group differences in tAD participants' pupil responses relative to controls have been observed when performing cognitively-demanding tasks (e.g. backwards digit span) (El Haj et al., 2022), and therefore effects in relation to music processing may be difficult to observe.

4.2.5 *Music familiarity neuroanatomical correlates: Links with pupillometry?*

The neuroanatomical findings from Slattery et al. (2019) may however give optimism to the notion of finding preserved psychosensory pupillometry responses in Alzheimer's disease cohorts. Despite the authors observing some between-group activation differences during familiar music listening, Slattery et al. (2019) reported that musical semantic processing (familiar - unfamiliar melody conditions) was associated in both tAD and control groups with bilateral supplementary motor cortex (SMA) activation. This finding is of particular interest when considering the pupil as a physiological candidate to measure subjective music familiarity in this clinical population. While precise activation localisation within SMA was not described in Slattery et al. (2019), a specific region within - implicated with oculomotor function - may suggest potential for preserved pupillometry responses to familiar music in individuals with Alzheimer's disease (if only for the memory-led group). The Supplementary Eye Fields (SEF) are located in the anterior part of the SMA in the posteromedial part of superior frontal gyrus (Pierrot-Deseilligny, Rivaud, Gaymard, Muri, & Vermersch, 1995), which was also identified in a neuroimaging meta-analysis of brain regions engaged during familiar music presentation, reflecting a more motor-related network then previously considered (Freitas et al., 2018). Whilst traditionally implicated in saccade preparation, recent findings in animal models have recently linked SEF to pupil dilation under conditions where cognitive processing is required (Claron et al., 2022).

4.2.6 The influence of disease severity on pupil response

There are mixed findings regarding the influence of disease severity on pupil responses in the studies undertaken to date. Bittner, Wieseler, Wilhelm, Riepe, and Muller (2014) reported marginal associations between lower pupil size increases and higher levels of tau (AD cerebrospinal fluid biomarkers); higher levels of tau were associated with a worse MMSE score. Fotiou, Kaltsatou, Tsiptsios, and Nakou (2015) explored time-based pupillometry metrics such as maximum constriction acceleration and velocity in individuals with AD, demonstrating a relationship with MMSE score. No correlations between disease severity (as indicated by MMSE scores) and pupil responses were however observed in Fletcher, Nicholas, et al. (2015b). It is likely that MMSE score will need to be included in any subsequent investigations regarding pupillometry responses in patient populations to avoid disease severity masking any potential effects of cognitive and emotional music processing, which has been demonstrated to persist in individuals with late-stage dementia indicated by severe MMSE scores [<10] (Baird & Samson, 2009).

4.2.7 The current study

The current study seeks to establish if the control group findings in the Play It Again study are observable within-group when administering the paradigm with people living with either typical or a non-typical variant of AD. Further analyses will then investigate between-group differences between pupillometry responses of each patient group relative to controls, and investigate between-all-group differences in subjective self-reports.

4.2.8 Research questions and hypotheses

Pupillometry responses

1. Are self-reported pleasantness music ratings associated with pupillometry change from baseline in all groups? [*Associated hypothesis (RH1)*]: *We predicted a curvilinear association of pleasantness ratings with pupil response change from baseline, with an interaction of group*

2. Are independent classifications of a song's familiarity associated with pupil response change from baseline in all groups? [*Associated hypotheses (RH2)*]:

We predicted an association of a priori familiarity predictions with pupil response change from baseline, with an interaction of group

3. Are subjective familiarity music ratings associated with pupillometry change from baseline? [*Associated hypotheses (RH3)*]: *We predicted an association of self-reported familiarity ratings with pupil response change from baseline, with an interaction of group*

4. Is memory evocation of a song associated with pupillometry change from baseline? [*Associated hypotheses (RH4)*]: *We predicted an association of binary memory ratings with pupil response change from baseline, with an interaction of group*

Secondary hypotheses related to associated subjective music ratings

5. Are independent classifications of a song's familiarity associated with pleasantness ratings in all groups? [*Associated hypothesis (RH5)*]*: We predicted an association of a priori familiarity ratings with pleasantness rating*

6. Are subjective judgements of a song's familiarity associated with their pleasantness ratings in all groups? [*Associated hypothesis (RH6)*]*: We predicted an association of subjective familiarity rating with pleasantness rating*

7. Are independent classifications of a song's familiarity associated with likelihoods of reporting that song as familiar in all groups? [*Associated hypothesis (RH7):*] *We predicted an association of a priori familiarity predictions with self-reported familiarity scores*

8. Are subjective judgements of a song's familiarity associated with their likelihood of reporting that song as evoking a memory in all groups? [*Associated hypothesis (RH8)*]: *We predicted an association of subjective familiarity scores with memory evocation ratings*

9. Are songs released within participants' putative 'reminiscence bump' (between 13 and 19 years old) associated with subjective music ratings in all groups? [*Associated hypothesis (RH9)*]*: We predicted an associated of a song being released between each participants' age of 13 and 19 with participants' subjective music ratings*

4.3 METHODS

4.3.1 Participants and characterization

42 participants with a diagnosis of either PCA (N = 23; see Box 4.1) or memory-led tAD (N = 19; see Box 4.2) were recruited from the Specialist Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery (NHNN), London, UK. Participants were identified by virtue of signing a departmental Data Protection form which indicates consent to be contacted about affiliated research opportunities.

Box 4.1 *Posterior cortical atrophy (PCA)*

PCA manifests itself by virtue of predominantly atrophying parietal, occipital and occipito-temporal regions, sporadically in individuals who are relatively young people of working age (<65) (Crutch et al., 2012). Its most common pathology is Alzheimer's disease, however PCA can arise from LBD and corticobasal syndrome (CBS) (Yong et al., 2023). Whilst PCA is relatively rare [suggested prevalence of 2.3/100,000 people aged between 30 and 64 years of age (Chiari et al., 2021)], it is the most common atypical form of Alzheimer's disease (Dubois et al., 2014). Although other dementia phenotypes can incur visual problems, PCA is termed the cardinal 'visual variant' of dementia, owing to its predominant and ubiquitous visual and visuospatial-related impairments (Schott & Crutch, 2019). It is however worth noting that impairments in other modalities (e.g. auditory scene analysis) have been reported in this group (Hardy, Yong, Goll, Crutch, & Warren, 2020). Healthcare professionals reviewing videos of people living with PCA and learning about these visuospatial deficits are often keen to learn more about the condition, however with some tendency to attribute these difficulties to more traditional perceptions of dementia (e.g. relating to memory problems or challenging behaviour) (McIntyre et al., 2019). Symptomatic relief using medication traditionally prescribed for more typical forms of Alzheimer's disease (e.g. donepezil) results in nightmares and vivid dreams, with no effect on MMSE as indicated in an RCT (Ridha et al., 2018). This condition is therefore important to consider in the current work to explore potential ways to measure responses to non-pharmacological activities, which may support emotional and cognitive wellbeing.

Three participants were subsequently excluded from analysis due to participant fatigue precluding commencement of the experimental paradigm (N_{PCA} = 2), and one

tAD participant accommodating only 52% of all pupillometry trials, resulting in a final sample of 39 participants living with a diagnosis (N_{PCA} = 21, N_{AD} = 18) (see Table 4.1).

Box 4.2 *memory-led 'typical' Alzheimer's disease (tAD)*

Alzheimer's disease, the most common underlying pathology causing dementia, accounts for 80% of all dementia cases. It manifests distinctive clinical presentations depending on the initial site of atrophy at disease onset (Crous-Bou, Minguillon, Gramunt, & Molinuevo, 2017). 'Typical' or memory-led Alzheimer's disease (mAD) is characterised by difficulties in memory and orientation (C. A. Gold & Budson, 2008). Most individuals with Alzheimer's disease develop the condition after the age of 65, however in approximately 4-5% of the total number of individuals with Alzheimer's disease, the disease will have an earlier onset, whereby symptoms tend to progress more rapidly (<65) (Mendez, 2012). Whilst this syndromic group are often selected by to fulfil a 'disease control group' position relative to rarer or non-memory-led dementias, the documentation of preserved semantic musical memory in this group is of particular interest for the current work (Benhamou & Warren, 2020).

Background participant information, health status (e.g. visual/hearing impairments, neurological comorbidities, family history of a neurological condition) was documented using the UCL Dementia Research Centre demographics questionnaire, also used in clinical assessments, and music engagement was established in line with the previous chapter (Section 3.3.1). At testing, 89% of PCA participants and 76% of memory-led Alzheimer's disease participants were receiving symptomatic treatment with acetylcholinesterase inhibitors. Putative reminiscence bump classification of each song was coded in line with the previous chapter. All participants attempted the Wechsler Abbreviated Scale of Intelligence (WASI Test) - Matrices (Shortened version) and participants additionally completed a comprehensive neuropsychological assessment, including the MMSE, recognition memory, general function, non-visual and visual parietal tasks.

4.3.2 *Ethical approval and funding*

Ethical approval and funding remain unchanged from the previous chapter.

4.3.3 Stimuli

Stimuli remains unchanged from the previous chapter.

4.3.4 Equipment

The equipment remains unchanged from the previous chapter.

4.3.5 Self-report measures

Subjective appraisal of each song remains unchanged from the previous chapter.

4.3.6 Procedure

The procedure remains unchanged from the previous chapter, with the exception of additional neuropsychological assessment undertaken in the patient population subsequent to the pupillometry paradigm (Figure 3.3 re-inserted below for reference). Participants were fully debriefed once the experimental procedure was completed.
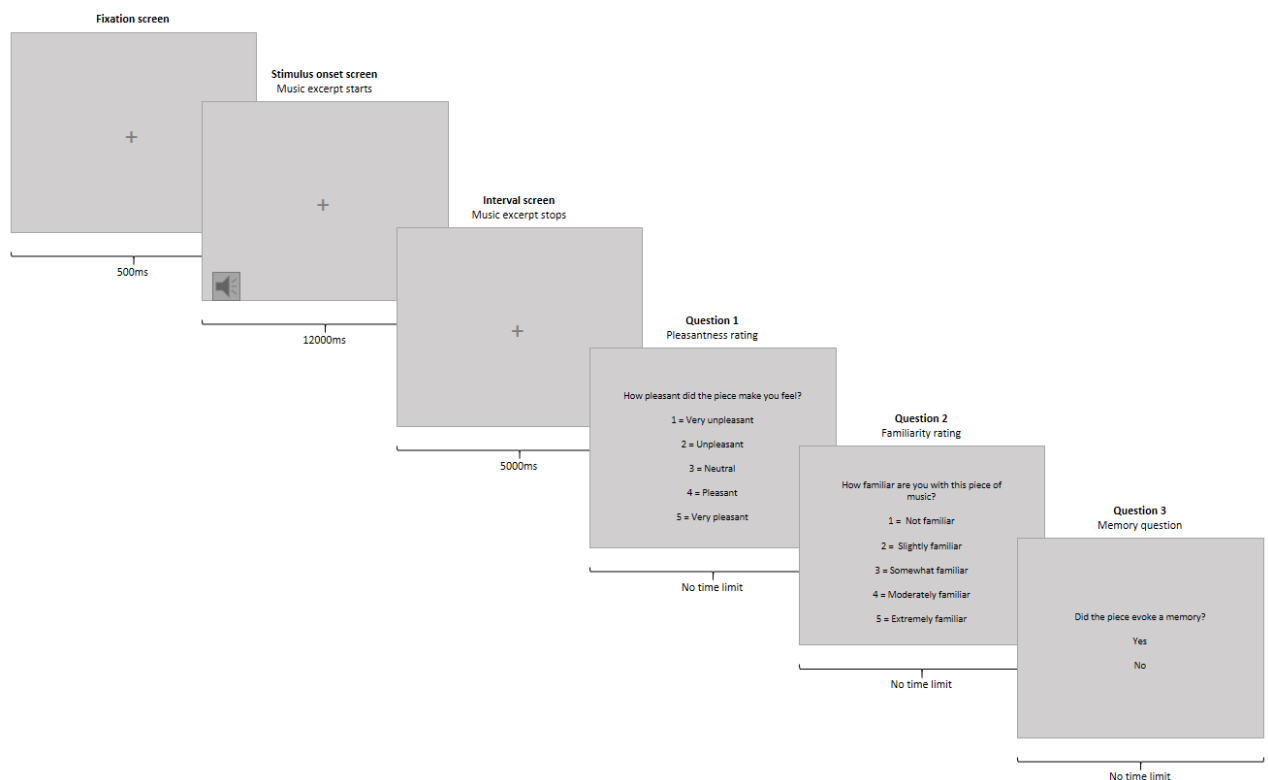
Figure 3.3 Sequence of one trial within the Play It Again experiment.

4.3.7   Pupillometry

Missing pupillometry data in collection phase

A software bug caused missing data for a small amount of trials (0.01%) across 6 participants at the point of data collection. Due to the stimuli being randomised, we could confidently conclude this resulted in data missing at random and therefore there was no reason to exclude these individuals from subsequent analysis. Other missing data attributable to artefacts such as eye blinks were also concluded to be missing at random, and therefore missing values were imputed using established pupillometry pre-processing techniques so that the final preprocessed datasets contained no missing entries.

Rating and pupillometry pre-processing

Self-rated: Rating variability was established by calculating the difference between each participant's rating and the group mean to explore between-group differences (rating$_{var}$).

Pupillometry: pre-processing steps remain unchanged from the previous chapter. Two participants were excluded owing to more than 50% of data missing, plus 4% of overall trials were excluded whereby more than 50% of data were missing within-trial. To align with the control outlier identification, trialwise responses greater than 6.5 times of each respective group mean were removed (PCA group N = 1; AD group, N = 1). Pupil variability was established by calculating the difference between each participant's pupil$_{max}$ response and the group pupil$_{max}$ mean to explore within- and between-group differences (pupil$_{var}$).

4.3.8   Statistical Methods

All statistical analyses were undertaken in R. All within-group statistical approaches in the current study replicate the previous chapter, with the additional variable of *MMSE* added to ascertain any effects of disease severity on self-report and/or pupillometry responses. Between-group analyses were undertaken. Additional considerations and approaches to between-group analyses comparing three participant groups are outlined in the sections below.

4.3.8.1 Demographic and neuropsychological data

Chi-square (or Fisher's Exact Test where expected frequencies were lower than 5) for categorical data, and Kruskal-Wallis rank sum test for numeric data (subsequent to ascertaining whether normality assumptions were upheld for each numeric variable) were used to compare across the three participant groups, with Bonferroni-corrected multiple pairwise comparisons conducted to elucidate specific differences. Neuropsychological performance in the two patient groups were compared for continuous data using independent-sample t-tests (or Wilcoxon rank sum where normality assumptions were violated) and Fisher's Exact Test to ascertain differences in the categorical pass/fail scoring within the Ishihara plate task.

4.3.8.2 Subjective music ratings

Within-group CLMM and logistic regressions were undertaken. Kendall's Tau correlations were performed to assess the relationship between self-reported pleasantness and familiarity ratings within each group. Replicating the previous chapter, exponential values of the model coefficients and confidence intervals were calculated to generate odds ratios (OR) and 95% confidence intervals, reported in square brackets. All significant effects are reported at the level of $p < 0.05$. For between-group analysis and graphical representation, the ordinal assumption for the rating scales was relaxed and the data were explored as continuous variables, as demonstrated in Fletcher, Nicholas, et al. (2015b). Data were interrogated between the three participant groups using Kruskal-Wallis tests, followed by post-hoc multiple pairwise comparisons (Wilcoxon rank-sum tests with Bonferroni correction) to ascertain specific drivers for between-group differences.

4.3.8.3 Physiological data

For the pupil outcome, the same modelling strategy was used as the previous chapter in the knowledge of the slightly skewed residuals in the control data. Cluster robust standard errors were run with the 'HC3' method when analysing within-group effects in the patient populations (correcting for smaller sample sizes). Between-group effects were interrogated by adding *participant group (control, tAD, PCA)* to the model composition, using the 'HC1' method, comparing each patient group to the control group. Additional interaction terms between diagnosis and (a) music rating; (b) trial

index and (c) average gaze position on the Y-axis and (d) average gaze position on the X-axis were explored.

4.4 RESULTS

Significant findings at the level of $p < 0.001$ are denoted by double asterisks; single asterisks denote p-values at the level of $0.001 < p < 0.05$.

4.4.1 Participant characteristics

All relative frequencies and positive responses to the demographics questions are reported in Table 4.1. No nuisance covariates in terms of age, mood, education, music experience or other background information were identified in the demographics data. Expected differences were observed between the patient groups and controls in the report of a neurological condition ($\chi2$ (2) = 127.0, $p < 0.001$) and acetylcholine medication prescription (p<0.001; Fisher's Exact Test, FET). A significant difference in the prevalence of reported dyslexia (p = 0.002; FET), was further investigated. The difference was, in line with prevailing symptoms, driven exclusively by the PCA group, who were 19.84 (1.82, 1023.88) times more likely to report dyslexia than controls (p = 0.004; FET). There was no significant difference in reported dyslexia between the tAD and PCA groups (p = 0.672; FET); or between reported dyslexia between the tAD and control groups (p = 0.07; FET).

4.4.2 Neuropsychological characteristics

Kruskal-Wallis tests revealed group differences in WASI scores between the three groups (H(2) = 74.44, $p < 0.001$), with post-hoc multiple comparisons demonstrating significant differences between all groups (Controls > AD > PCA, all p<0.05). Mean and confidence intervals of neuropsychological test scores comparing the detailed neuropsychological profiles captured within the PCA and AD participant groups, alongside significance levels for between-group test differences, are reported and discussed further in Chapter 5. For the purposes of the current chapter, scores reflected an impaired performance in visuospatial tasks in the PCA group and

impairment in mnemonic performance in the tAD group, with no significant difference in MMSE scores.

Table 4.1 Demographic, clinical and musical experience group characteristics of all Play It Again participant groups

	Participant group			P-value
	Control	**PCA**	**AD**	
No. in group	*90*	*21*	*18*	
Gender (female:male)	52:38	8:13	10:8	0.263 [c]
Age (years): $\bar{x}$ [IQR]	69.0 [61.0;74.0]	65.0 [61.0;66.0]	69.0 [65.2;70.8]	0.269 [a]
Min./Max.	51.0 / 80.0	56.0 / 75.0	50.0 / 77.0	
Handedness (left:right)	7:83	3:18	2:16	0.562 [b]
MMSE mean score ± sd	NA	20.9 ± 5.3	19.2 ± 5.2	0.405 [d]
UK-born	69 (77%)	18 (86%)	13 (76%)	0.792 [b]
Residing in a different country to birth for > 1 year	44 (49%)	7 (33%)	8 (47%)	0.434 [c]
Educated to at least undergraduate degree	73 (81%)	16 (76%)	10 (59%)	0.127 [b]
Native English speaker	83 (92%)	17 (81%)	17 (94%)	0.256 [b]
English fluency	89 (99%)[1]	21 (100%)	18 (100%)	1.000 [b]
Visual impairment present[2]	13 (15%)	6 (29%)	2 (12%)	0.276 [b]
Hearing impairment present[3]	6 (7%)	4 (19%)	2 (12%)	0.164 [b]
In a 'good' mood (>=3 on mood scale)	82 (93%)	21 (100%)	16 (100%)	0.524 [b]
Reporting a close (blood) relative with a neurological condition	28 (31%)	4 (21%)	7 (41%)	0.427 [c]
Dyslexia (self-reported)	1 (1%)	4 (19%)	2 (12%)	0.002 [b] **
Self-reporting a neurological condition	0 (0%)	21 (100%)	18 (100%)	<0.001 [c] **
No. receiving acetylcholinesterase medication	0 (0%)	16 (89%)	13 (76%)	<0.001 [b] **
Wechsler Adult Scale of Intelligence: Matrices (Short version) %: $\bar{x}$ [IQR]	0.8 [0.7;0.8]	0 [0;0]	0.3 [0.2;0.5]	<0.001 [a] **
Scored music experience: 0-4: $\bar{x}$ [IQR]	1.0 [0;3.0]	1.0 [1.0;3.0]	1.0 [0;1.0]	0.214 [a]

[1] As outlined in the Created Out of Mind study protocol (17/LO/0099 and 8545/002: Created Out of Mind), participants were required to be proficient in the English language to take part in the experiment. For all non-native English speakers, their English comprehension was deemed proficient by the research team to continue participation, with the exception of one fluent Italian speaker. For this

participant, a bilingual Italian-English academic collaborator on the project sought and obtained consent, and conducted the experiment.

[2] Visual impairments listed relating to the eyes to ascertain precautions around pupillometry recording (e.g. monocular recording). Visual impairments reported here any impairment could not be corrected-to-normal e.g. with the use of spectacles or contact lenses.

[3] Participants required to wear hearing aids were asked to bring them to the experiment. Hearing impairments reported here demonstrate any impairment which could not be corrected-to-normal e.g. with the use of hearings aids, or due to the participant forgetting to bring hearing aids to the experiment.

[a] = Kruskal-Wallis test; [b] = Fisher's Exact Test; [c] = Pearson's Chi-Squared test, NA = not administered

[d] As MMSE was only administered in the two patient groups, between-patient-group differences were interrogated for covariates; scored music experience was found to be different between the patient groups. A simple linear regression modelling MMSE score from diagnosis was run, with music experience included as a nuisance covariate to elicit the tabled results.

4.4.3 Participant music listening

PCA and AD groups were respectively 4.67 [1.49, 16.52] and 8.62 [2.13, 50.99] times more likely to listen to music for ten or more hours per week than control participants. Classical music was the most popular listened-to genre in all participant groups. The patient groups demonstrated a higher listening prevalence in each music genre compared with controls where a significant association was demonstrated (Table 4.2). With regards to music settings (e.g. where a participant played their instrument(s) and/or engaged in singing), individuals with typical Alzheimer's disease were 5.33 [1.39, 20.15] times more likely to go to listen to music at a festival compared with control participants (p = 0.009, FET).

Table 4.2 Odds ratios for music listening genres with significant association of *group*

Music genre	PCA vs Controls	AD vs Controls	PCA vs AD
Pop	ns	3.76 [1.11, 14.89]	ns
Easy Listening	3.26 [1.08, 9.97]	5.09 [1.52, 18.10]	ns
Comedy	20.52 [3.46,221.29]	12.68 [1.64,153.34]	ns
Country	7.49 [2.38,24.91]	ns	ns
Electronic	4.31 [1.16,15.65]	1.87 [0.29, 8.81]	ns
Folk	3.46[1.16,10.57]	ns	ns
Hip-Hop	6.50 [1.00,48.43]	8.43 [1.27, 64.39]	ns
Reggae	4.19 [0.90, 18.85]	5.55 [1.15, 25.95]	ns
African	6.93 [1.86, 26.96]	ns	ns
Asian	5.34 [1.25, 23.12]	ns	<0.001[<0.001,0.91]
Latin	3.93 [0.98, 15.22]	ns	ns

4.4.4 Self-report responses: Overview

Proportional responses for, and mean ratings of *pleasantness* and *familiarity* ratings for each participant group are respectively plotted in Figure 4.1; further graphical representation of the self-report ratings are available in Figure 4.2.

Between-group pleasantness rating comparisons

All groups elicited the same median rating ($\tilde{x}$ = 4 [*pleasant*]), however Kruskal-Wallis tests revealed a significant group difference in pleasantness subjective responses ($H(2)$ = 9.43, p = 0.009). Post-hoc multiple pairwise comparisons revealed that this was driven solely by differences in the PCA and control group; PCA participants attributed significantly higher pleasantness scores to the playlist compared to control participants (p=0.006) (Figure 4.1). Variability of pleasantness ratings differed significantly between groups ($H(2)$=79.603, $p < 0.001$). Bonferroni-corrected multiple pairwise comparisons indicated this was driven by significant differences in the variability of pleasantness ratings between each participant group (largest variability: PCA>AD>Control group, all p <0.001). While no further between-group differences in pleasantness ratings were demonstrated, patient group-control comparisons for each song revealed several significant differences in pleasantness ratings (see Table 4.4).

Between-group familiarity rating comparisons

No between-group differences were observed for subjective familiarity responses ($H(2)$ = 0.827, p = 0.661). Significant differences in between-group familiarity rating variability was observed ($H(2)$ = 194.03, p <0.001). Bonferroni-corrected multiple pairwise comparisons revealed that this was driven by significant differences in the variability of familiarity ratings between each participant group (largest variability: PCA>AD>Control group, all p <0.001). While no further between-group differences in familiarity ratings were demonstrated, patient group-control comparisons for each song revealed several significant differences in familiarity ratings (see Table 4.4).

Between-group memory evocation comparisons

All groups elicited the same median rating ($\tilde{x}$ = 0 [*no memory evocation*]), however Kruskal-Wallis tests revealed a significant group difference in memory evocation responses ($H(2)$ = 14.046, $p < 0.001$). Post-hoc multiple pairwise comparisons revealed that this was driven solely by differences between the PCA and control group;

129
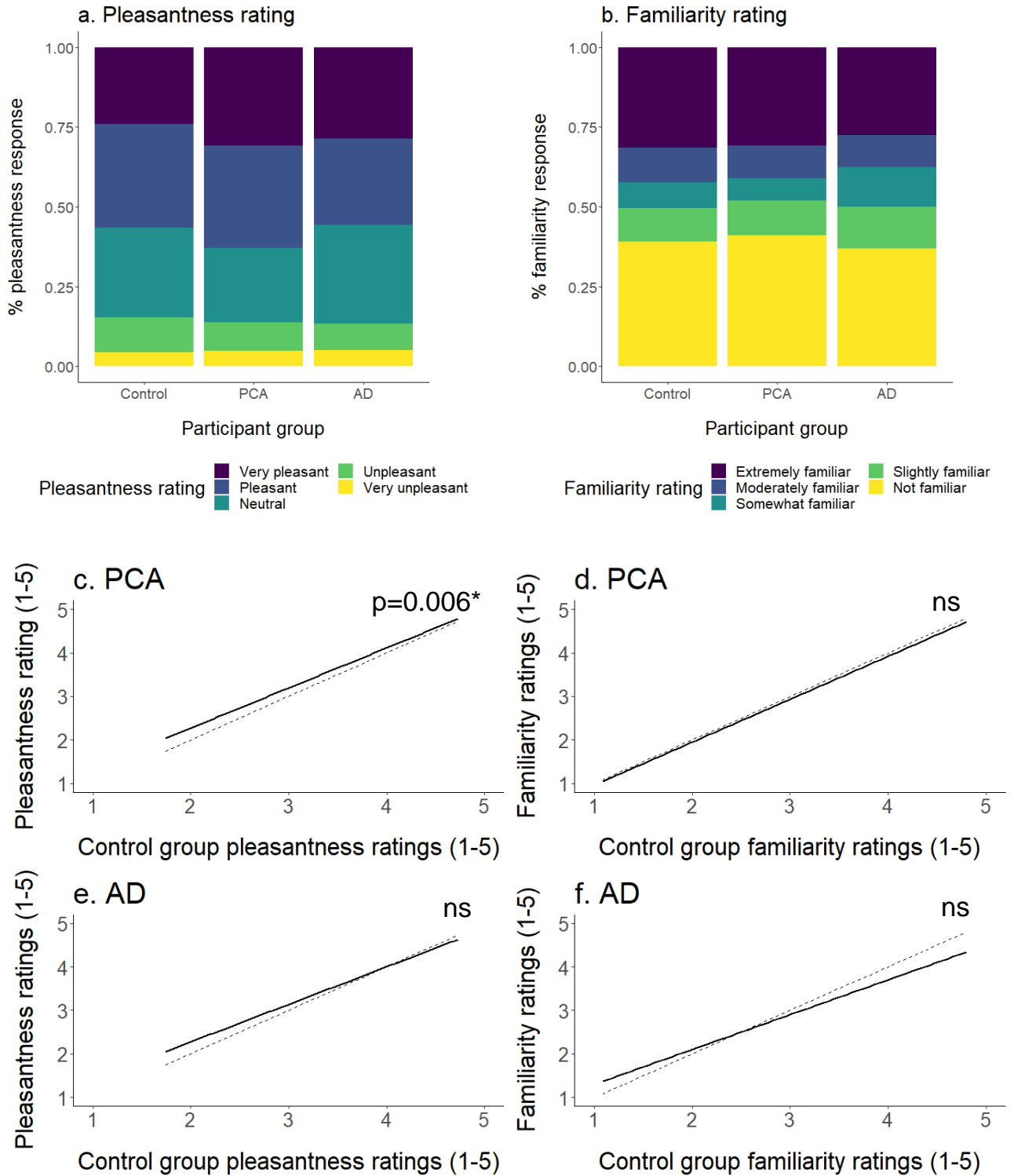
Figure 4.1 Within- and between-group graphical representations of the pleasantness and familiarity subjective ratings. Stacked bar charts (top) indicate the within-group proportional responses to *pleasantness* [**A**] and *familiarity* [**B**] ratings and facilitating visual inspection for potential between-group differences; **C-F:** mean group pleasantness and familiarity rating for each stimulus against the healthy control group mean subjective ratings, for each patient group. Best fit lines demonstrate control group

responses (dotted lines) and patient group responses (solid line). AD = Alzheimer's disease; PCA = Posterior Cortical Atrophy.

PCA participants attributed significantly higher memory evocation scores to the playlist compared to control participants (p< 0.001). While no further group differences were demonstrated, patient group-control comparisons for each song revealed several significant differences in memory evocation ratings (see Table 4.4).

4.4.5   Pupillometry responses: Overview

Kruskal-Wallis tests revealed significant differences in baseline pupil size (denoted in Eyelink arbitrary units) between the participant groups (H(2) = 1392.8, p <0.001). Post-hoc Bonferroni-corrected multiple pairwise comparisons revealed that this was driven by significant differences between all groups (baseline size: Control > AD > PCA, all p < 0.001) (Table 4.3). Similarly, significant differences between the maximum trialwise pupil size (in Eyelink arbitrary units) were demonstrated (H(2) = 1300.4, p<0.001). Post-hoc Bonferroni-corrected multiple pairwise comparisons revealed the same profile of differences as with the baseline pupil size (control-PCA difference: p<0.001; control-AD difference: p<0.001), and between the patient groups (p=0.006) (Table 4.3). Kruskal-Wallis tests revealed a significant effect of *participant group* on pupil$_{max}$ variability (H(2) = 8.507, p = 0.014), with Bonferroni-corrected multiple pairwise comparisons indicating that this difference was being driven solely by larger variability in the PCA group pupil$_{max}$ responses compared to the control group (p = 0.034).

Table 4.3 Mean baseline pupil size and maximum trialwise pupil size (arbitrary Eyelink units [95% CI]).

	Participant group		
	Control	**PCA**	**AD**
Absolute mean baseline pupil size	1525.7 [1505.2;1546.1]	783.3 [763.4;803.2]	881.8 [855.7;907.9]
Maximum trialwise pupil size	1689.9 [1667.6;1712.2]	918.7 [895.5;941.9]	1003.6 [974.9;1032.2]
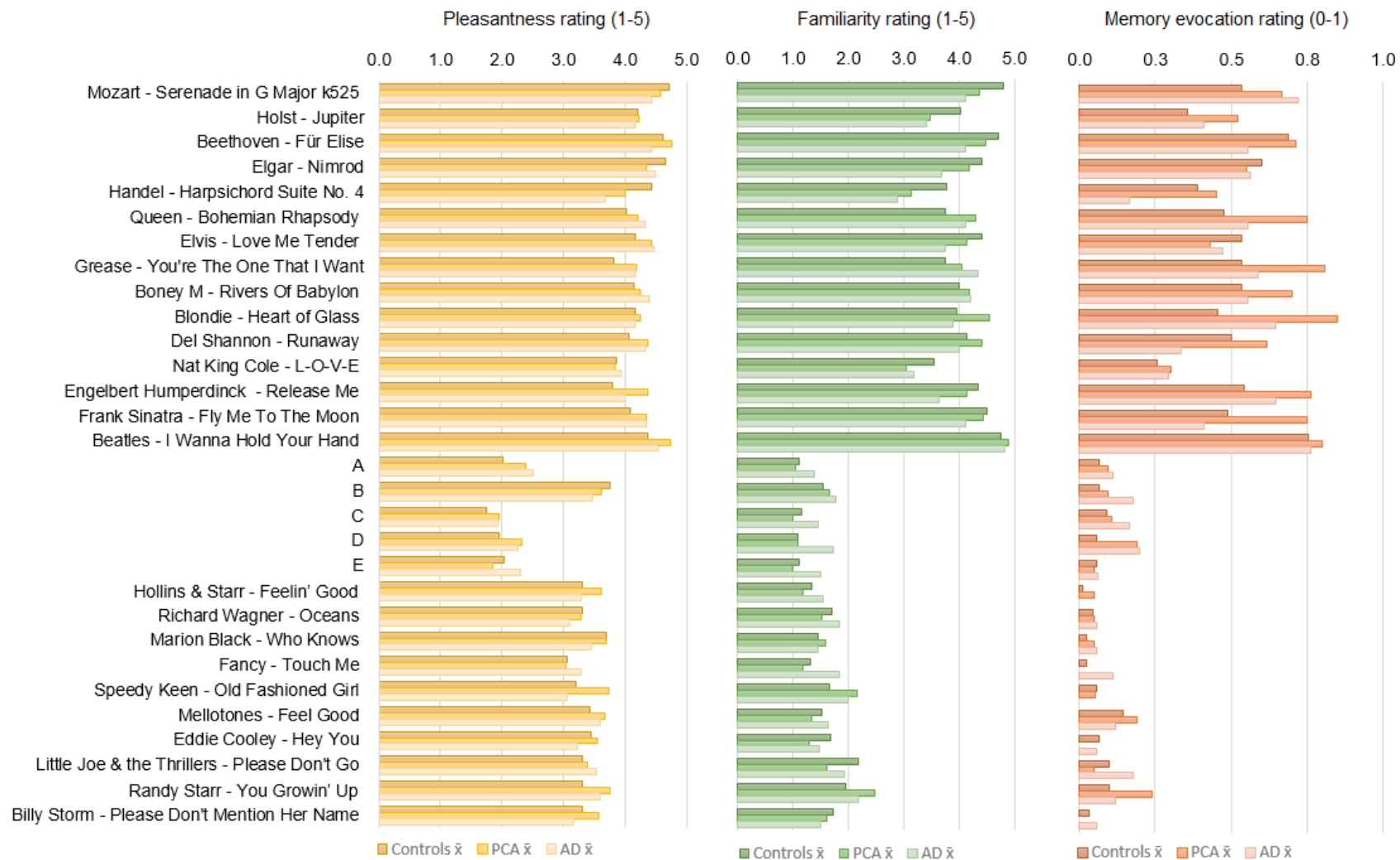
Figure 4.2   Graphical representation of the mean pleasantness, familiarity and memory evocation ratings broadly categorised into *a priori* known songs (top 15 songs) and unknown (bottom 15 songs) comprising the experimental playlist. **PCA:** Posterior Cortical Atrophy; **AD:** typical (memory-led) Alzheimer's disease

Table 4.4 Pleasantness, familiarity and memory evocation self-report ratings of experimental playlist by healthy older adult control participants, PCA and AD participants. PLWD group ratings which differ significantly from the control group are in bold; arrows indicate whether ratings were significantly higher (↑) or lower (↓) than the control participants.

A priori prediction	Stimuli set	Pleasantness ratings						Familiarity rating						Memory evocation rating					
		Controls		PCA		AD		Controls		PCA		AD		Controls		PCA		AD	
	Artist - Song name	x̄	sd	x̄	sd	x̄	sd	x̄	sd	x̄	sd	x̄	sd	x̄	sd	x̄	sd	x̄	sd
Known	Mozart - Serenade in G Major k525	4.7	0.4	4.6	0.7	4.4	0.6	4.8	0.5	**4.4↓**	0.9	**4.1↓**	0.9	0.5	0.5	0.7	0.5	0.7	0.5
	Holst - Jupiter	4.2	0.7	4.2	0.8	4.2	1.0	4.0	1.2	3.5	1.5	3.4	1.5	0.4	0.5	0.5	0.5	0.4	0.5
	Beethoven - Für Elise	4.6	0.5	4.8	0.4	4.4	1.0	4.7	0.6	4.5	0.9	4.1	1.2	0.7	0.5	0.7	0.5	0.6	0.5
	Elgar - Nimrod	4.7	0.6	4.4	0.7	4.5	0.6	4.4	1.1	4.2	1.3	3.7	1.6	0.6	0.5	0.6	0.5	0.6	0.5
	Handel - Harpsichord Suite No. 4	4.4	0.6	4.0	1.2	**3.7↓**	0.8	3.8	1.4	3.2	1.7	2.9	1.6	0.4	0.5	0.5	0.5	0.2	0.4
	Queen - Bohemian Rhapsody	4.0	0.9	4.2	1.0	4.3	0.6	3.8	1.7	4.3	1.5	4.1	1.4	0.5	0.5	0.8	0.4	0.6	0.5
	Elvis - Love Me Tender	4.2	0.8	4.4	0.8	4.5	0.8	4.4	1.1	4.1	1.3	3.8	1.4	0.5	0.5	0.4	0.5	0.5	0.5
	Grease - You're The One That I Want	3.8	1.0	4.2	0.9	4.2	0.8	3.8	1.5	4.0	1.3	4.4	1.1	0.5	0.5	0.8	0.4	0.6	0.5
	Boney M - Rivers Of Babylon	4.2	0.7	4.3	0.9	4.4	0.8	4.0	1.3	4.2	1.4	4.2	1.2	0.5	0.5	0.7	0.5	0.6	0.5
	Blondie - Heart of Glass	4.2	0.9	4.3	1.1	4.2	1.0	4.0	1.5	4.6	0.9	3.9	1.3	0.5	0.5	**0.9↑**	0.4	0.6	0.5
	Del Shannon - Runaway	4.1	0.8	4.4	0.9	4.3	0.7	4.1	1.2	4.4	1.0	4.0	1.1	0.5	0.5	0.6	0.5	0.3	0.5
	Nat King Cole - L-O-V-E	3.9	0.9	3.9	0.8	3.9	0.9	3.6	1.6	3.1	1.8	3.2	1.5	0.3	0.4	0.3	0.5	0.3	0.5
	Engelbert Humperdinck - Release Me	3.8	0.7	**4.4↑**	0.7	4.0	1.3	4.4	1.0	4.1	1.2	3.6	1.5	0.5	0.5	0.8	0.4	0.6	0.5
	Frank Sinatra - Fly Me To The Moon	4.1	0.9	4.4	0.8	4.4	0.9	4.5	0.9	4.5	0.9	4.1	1.2	0.5	0.5	0.8	0.4	0.4	0.5
	Beatles - I Wanna Hold Your Hand	4.4	0.8	4.8	0.6	4.5	0.7	4.8	0.8	4.9	0.3	4.8	0.4	0.8	0.4	0.8	0.4	0.8	0.4
Unknown	Nonsense (A)	2.0	0.8	2.4	1.3	2.5	1.1	1.1	0.5	1.0	0.2	1.4	1.0	0.1	0.3	0.1	0.3	0.1	0.3
	Nonsense (B)	3.8	0.6	3.6	0.9	3.5	0.7	1.5	0.9	1.7	1.1	1.8	1.3	0.1	0.3	0.1	0.3	0.2	0.4
	Nonsense (C)	1.7	0.8	1.9	1.0	1.9	1.3	1.2	0.6	1.0	0.0	1.4	1.0	0.1	0.3	0.1	0.3	0.2	0.4
	Nonsense (D)	2.0	0.8	2.3	1.1	2.3	1.0	1.1	0.4	1.1	0.3	**1.7↑**	1.2	0.1	0.2	0.2	0.4	0.2	0.4
	Nonsense (E)	2.0	0.9	1.9	0.9	2.3	1.0	1.1	0.5	1.0	0.0	1.5	1.0	0.1	0.2	0.1	0.2	0.1	0.3
	Hollins & Starr - Feelin' Good	3.3	0.7	3.6	0.8	3.3	1.0	1.3	0.7	1.2	0.4	1.6	1.1	0.0	0.1	0.0	0.2	0.0	0.0
	Richard Wagner - Oceans	3.3	0.6	3.3	0.8	3.1	0.8	1.7	1.0	1.5	1.0	1.8	1.2	0.0	0.2	0.0	0.2	0.1	0.2
	Marion Black - Who Knows	3.7	0.8	3.7	0.7	3.4	0.8	1.5	0.9	1.6	0.8	1.4	0.7	0.0	0.1	0.1	0.2	0.1	0.2
	Fancy - Touch Me	3.1	0.7	3.0	0.7	3.3	0.8	1.3	0.6	1.2	0.4	1.8	1.3	0.0	0.1	0.0	0.0	0.1	0.3
	Speedy Keen - Old Fashioned Girl	3.2	0.6	**3.7↑**	0.8	3.1	0.8	1.7	1.0	2.2	1.3	2.0	1.4	0.1	0.2	0.1	0.2	0.0	0.0
	Mellotones - Feel Good	3.4	0.8	3.7	0.7	3.6	0.6	1.5	0.9	1.3	0.7	1.6	0.9	0.1	0.4	0.2	0.4	0.1	0.3
	Eddie Cooley - Hey You	3.4	0.8	3.6	0.8	3.2	0.7	1.7	1.0	1.3	0.5	1.5	0.9	0.1	0.3	0.0	0.0	0.1	0.2
	Little Joe & the Thrillers - Please Don't Go	3.3	0.8	3.4	0.7	3.5	0.7	2.2	1.2	1.6	0.9	1.9	1.1	0.1	0.3	0.0	0.2	0.2	0.4
	Randy Starr - You Growin' Up	3.3	0.8	3.8	0.8	3.6	1.3	2.0	1.1	2.5	1.5	2.2	1.6	0.1	0.3	0.2	0.4	0.1	0.3
	Billy Storm - Please Don't Mention Her Name	3.3	0.7	3.6	0.7	3.2	0.7	1.7	1.0	1.6	0.9	1.5	1.0	0.0	0.2	0.0	0.0	0.1	0.2

Separate within-group linear regressions revealed no associations in the PCA group between musical properties and $pupil_{max}$ response (excerpt tempo [p = 0.981; music key [major/minor; p = 0.435]; lyrical content [p=0.832]). In the AD group, $pupil_{max}$ response was similarly not associated with either excerpt tempo (p=0.093) or lyrical content (p=0.144). However a reduced $pupil_{max}$ response was significantly associated with listening to songs in a minor key relative to songs in a major key (β = -0.026, p=0.013). Greater disease severity (lower MMSE score) was associated with an increased maximum pupil response in both the PCA (β = -0.006 [-0.008, -0.004], p<0.001) and AD (β = -0.003[-0.005, -0.001], p=0.005) groups.

4.4.6 Pleasantness responses

4.4.6.1 Pupil-pleasantness responses

In contrast with controls, the separate intercepts of each within-group linear regressions obtaining cluster robust standard errors ($Model_{(pupil)}1$) revealed no significant difference in $pupil_{max}$ response relative to baseline whilst listening to music excerpts which were subsequently rated as neutral in either patient group (PCA: p = 0.069; tAD: p = 0.210). However, in line with control findings, a curvilinear valence-pupil relationship was observed in the AD group; $pupil_{max}$ responses were larger whilst AD participants listened to songs which subsequently elicited any pleasantness rating compared to 'neutral' (see Table 4.5). These $pupil_{max}$ increases reached statistical significance relative to songs rated as 'neutral' whilst participants with AD listened to music excerpts they subsequently rated as either 'pleasant' (13.5[10.8, 16.2]%, p < 0.001) or 'very pleasant' (16.2[13.3, 19.1]%, p < 0.001), relative to $pupil_{max}$ responses whilst listening to music excerpts which were subsequently rated 'neutral'. Recoding pleasantness ratings as a continuous variable enabled a formal interrogation of the curvilinear relationship including all the terms below as both linear and quadratic terms (var^2), revealing a significant but weak pupil-pleasantness relationship in the AD group (r^2 = 0.122, p = 0.003) (see Figure 4.3). These findings were not observed in the PCA
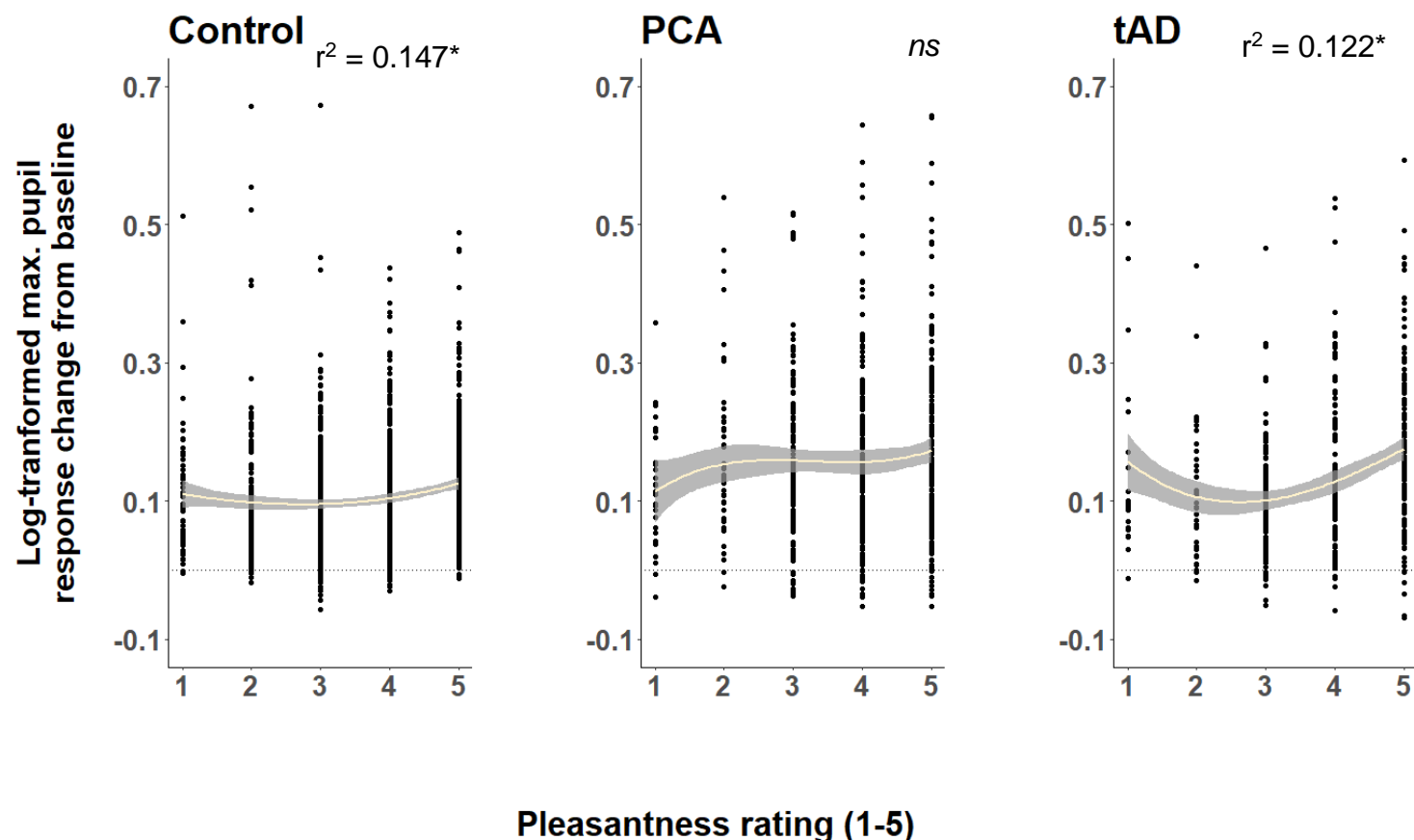
Figure 4.3 Log ratio of pupilmax response change from baseline, plotted against pleasantness ratings, split by participant group. Pleasantness ratings are on a Likert scale ranked: 1 = 'very unpleasant', 2 = 'unpleasant', 3 = 'neutral', 4 = 'pleasant', and 5 = 'very pleasant'. Points (·) indicate individual trial maximum pupil size change from baseline. Quadratic regression lines of best fit with 95% confidence intervals (shaded in grey) and r^2 are indicated. Dotted line indicates pre-trial baseline for reference. *Significant ($p < 0.05$) correlations between pupil response and song pleasantness rating; tAD: typical Alzheimer's disease; PCA: posterior cortical atrophy.

135

group, for which pupil$_{max}$ was predicted only by disease severity (*MMSE*) and average *gaze position on the Y-axis* (Table 4.5). *Trial number* was also significantly associated with pupil$_{max}$ responses in the AD group; as trial number increased, pupil responses became larger, in contrast to the pupil response observations in the control group. These effects were observed while controlling for disease severity (*MMSE*; $p < 0.001$), which was significantly associated with an overall reduced pupil$_{max}$ response in both patient groups.

Table 4.5 Separate within-group Model$_{(pupil)}$1 regression analyses undertaken predicting log maximum pupil response by pleasantness ratings, setting, gender, age, average gaze co-ordinates and trial number in PCA and tAD. Coefficients represent maximum percentage change in pupil response compared to baseline.

Model$_{[pupil]}$1	PCA		AD	
Predictor variable	Coefficient (exp)β [95% CI]	p-value	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.190 [0.987, 1.436]	0.069	1.095 [0.950, 1.263]	0.210
Pleasantness Rating: Very unpleasant	0.970 [0.935, 1.007]	0.111	1.047 [0.987, 1.111]	0.128
Pleasantness Rating: Unpleasant	1.024 [0.987, 1.062]	0.203	1.009 [0.974, 1.045]	0.632
Pleasantness Rating: Pleasant	0.999 [0.973, 1.027]	0.965	1.042 [1.017, 1.067]	<0.001**
Pleasantness Rating: Very pleasant	1.008 [0.980, 1.036]	0.581	1.067 [1.041, 1.094]	<0.001**
Setting: Wellcome	0.975 [0.935, 1.017]	0.240	-	-
Trial index	1.001 [0.999, 1.002]	0.369	1.002 [1.001, 1.003]	0.004*
Age (decade)	0.997 [0.971, 1.023]	0.799	1.009 [0.995, 1.023]	0.191
Average gaze X (unit:100 pixels)	0.992 [0.980, 1.003]	0.144	0.995 [0.984, 1.006]	0.372
Average gaze Y (unit: 100 pixels)	1.033 [1.024, 1.043]	<0.001**	1.007 [0.993, 1.022]	0.316
Gender: Male	0.997 [0.975, 1.019]	0.755	1.015 [0.992, 1.038]	0.197
MMSE	0.995 [0.993, 0.997]	<0.001**	0.995 [0.993, 0.998]	0.001*
Number of observations	*542*		*465*	

Following the observation that music key (major/minor) was a significant predictor of pupil$_{max}$ response in the AD group, data were modelled including *key* alongside the other Model$_{(pupil)}$1 predictor variables to ascertain any mediating effect on the impact

Table 4.6 Model$_{(pupil)}$1 (between-group) regression findings of log maximum pupil response predicted by pleasantness ratings, diagnosis setting, age, average gaze co-ordinates, trial number. Coefficients represent maximum percentage change in pupil response change from baseline, all relative to control group responses. Interactions (where explored) are indicated by the addition of an interaction term adjacent to the predictor variable column.

Predictor variable (relative to control group)	Interaction term	Coefficient (exp)β [95% CI]	p-value
(Intercept)	-	1.305 [1.221, 1.394]	<0.001**
Pleasantness Rating: Very unpleasant	AD	1.036 [0.975, 1.102]	0.252
	PCA	0.961 [0.921, 1.004]	0.072
Pleasantness Rating: Unpleasant	AD	0.992 [0.955, 1.032]	0.703
	PCA	1.011 [0.972, 1.052]	0.579
Pleasantness Rating: Pleasant	AD	1.026 [1.000, 1.053]	0.053
	PCA	0.990 [0.961, 1.020]	0.520
Pleasantness Rating: Very pleasant	AD	1.033 [1.006, 1.062]	0.018*
	PCA	0.983 [0.954, 1.012]	0.247
Diagnosis: AD (compared to controls)	-	0.896 [0.801, 1.003]	0.056
Diagnosis: PCA (compared to controls)	-	0.822 [0.749, 0.901]	<0.001**
Trial index	AD	1.003 [1.001, 1.004]	<0.001**
	PCA	1.001 [1.000, 1.002]	0.180
Average gaze Y (unit: 100 pixels)	AD	0.999 [0.986, 1.011]	0.843
	PCA	1.038 [1.028, 1.048]	<0.001**
Number of observations	*2350*		

of the main variables of interest. Running this model did not alter the initial findings reported above.

Between-group analyses exploring each PLWD group's pupil responses relative to controls revealed an association between *diagnosis* and overall pupil$_{max}$ response, indicated by 17.8[9.9,25.1]% smaller responses in the PCA participants relative to controls (see Table 4.6). No AD-control pupil$_{max}$ differences were observed. There was an interaction effect of *diagnosis x pleasantness* demonstrated by a 3.3[0.6,6.2]% larger pupil response in the AD group whilst listening to songs which were

subsequently rated as 'very pleasant' relative to control responses to similarly-rated songs (p=0.018). An interaction effect of *gaze position x diagnosis* was revealed (in the *PCA* group compared to controls' pupil size when positioned on the Y-axis). A *trial number x diagnosis* interaction was observed (with AD pupils increasing with trial number in contrast to control pupils decreasing).

4.4.6.2 Pleasantness responses predicted by a priori familiarity stimuli selection

Separate within-group CLMM modelling (Model$_{(rating)}$1) run for PCA and tAD responses revealed that, similarly to controls, both patient groups were significantly more likely to rate a song as pleasant if the song had been predicted to be 'known' compared to 'unknown' songs (see Table 4.7). All genres had higher odds of a higher pleasantness rating compared to nonsense songs, albeit with different odds within the patient groups. MMSE score was not a predictor of pleasantness ratings in either patient group. In both groups, the Participant and Song ID intercepts were significant (p<0.001).

Table 4.7 Separate within-group cumulative link mixed model (CLMM) Model$_{(rating)}$1 outcomes for the PCA and tAD groups; significant predictive factors indicated within shading, main predictive variable in bold.

Model$_{(rating)}$1 (within-group)	PCA		AD	
Fixed effect	*Odds ratio [95% CI]*	*p-value*	*Odds ratio [95% CI]*	*p-value*
A priori familiarity indicator (odds of a higher pleasant rating when song was *a priori* predicted to be known)	7.091 [3.577, 14.054]	<0.001**	9.546 [5.072, 17.966]	<0.001**
1960s songs (compared with nonsense songs)	9.418 [3.788, 23.418]	<0.001**	8.083 [3.435, 19.021]	<0.001**
1970s songs (compared with nonsense songs)	7.895 [3.141, 19.844]	<0.001**	6.795 [2.949, 15.654]	<0.001**
Classical songs (compared with nonsense songs)	10.587 [3.216, 34.849]	<0.001**	6.508 [2.212, 19.142]	<0.001**
Reminiscence bump song release	1.039 [0.684, 1.578]	0.859	0.754 [0.467, 1.218]	0.249
Trial number	0.997 [0.980, 1.015]	0.733	0.989 [0.970, 1.009]	0.282
MMSE	0.945 [0.884, 1.010]	0.096	0.955 [0.886, 1.028]	0.221
SD (Intercept Participant ID)	1.881	<0.001	1.664	<0.001**
SD (Intercept Song ID)	2.062	<0.001	2.031	<0.001**
Number of observations	*615*		*515*	

The same response profile of preferable responses being associated with all songs over the nonsense songs prompted a re-analysis of Model$_{(pupil)}$1 in each group without the nonsense songs included, to establish whether the nonsense subset was driving the curvilinear response. However, in line with the control findings, this did not notably alter the findings relating to the pleasantness-pupil responses in either group.

4.4.6.3 Pleasantness response associations with subjective ratings of familiarity

Separate within-group CLMM modelling (Model$_{(rating)}$2) run for PCA and tAD responses revealed a similar profile of odds ratios in the PCA as observed in control findings; the odds ratios of a higher pleasantness rating increased linearly with each level of familiarity (see Table 4.8). In the tAD group, participants' odd ratios of a higher pleasantness rating showed an S-curve, with songs reported as somewhat familiar and moderately familiar having similar odds of a higher pleasantness rating. Neither *MMSE score* nor *trial number* were predictors of pleasantness ratings in either patient group. In both groups, the Participant and Song ID intercept SDs were significant ($p<0.001$).

Table 4.8 Separate within-group CLMM Model$_{(rating)}$2 outcomes for the PCA and tAD groups; significant predictive factors indicated within shading, main predictive variable of interest in bold.

Model$_{(rating)}$2 (within-group)	PCA		AD	
Predictor variable	*Odds ratio [95% CI]*	*p-value*	*Odds ratio [95% CI]*	*p-value*
Songs self-reported as slightly familiar	3.849 [2.185, 6.780]	<0.001**	2.490 [1.336, 4.644]	0.004*
Songs self-reported as somewhat familiar	4.417 [2.184, 8.934]	<0.001**	7.583 [3.743, 15.362]	<0.001**
Songs self-reported as moderately familiar	10.523 [5.326, 20.789]	<0.001**	7.393 [3.507, 15.584]	<0.001**
Songs self-reported as extremely familiar	80.766 [40.321, 161.778]	<0.001**	82.144 [38.893, 173.492]	<0.001**
Reminiscence bump song release	1.224 [0.791, 1.894]	0.365	0.720 [0.431, 1.204]	0.211
Trial number	0.998 [0.980, 1.017]	0.842	0.980 [0.960, 1.001]	0.065
MMSE	0.972 [0.904, 1.045]	0.445	0.968 [0.880, 1.065]	0.501
SD (Intercept Participant ID)	2.374	<0.001**	2.603	<0.001**
SD (Intercept Song ID)	2.199	<0.001**	2.608	<0.001**
Number of observations	*615*		*515*	

Both the PCA and tAD groups reflected the positive correlation between pleasantness and familiarity ratings observed in the control group, albeit with a slightly lower Kendall's Tau coefficient in both patient groups relative to the control group (PCA = 0.59, p < 0.001; tAD = 0.54, p < 0.001). Both Kendall's Tau coefficients indicated a moderate positive correlation between the two variables.

4.4.7 Familiarity responses to music

4.4.7.1 Pupil-familiarity responses based on a priori familiarity prediction

In line with controls, the intercept in the AD group Model$_{(pupil)}$2 revealed a significantly larger pupil$_{max}$ response was elicited relative to baseline while AD participants listened to reference excerpts (music *a priori* predicted to be unknown: 17.5[2.2,35.2]%, p=0.023; see Table 4.9). Relative to this, a statistically significant larger pupil$_{max}$ response was elicited while participants with AD listened to songs which were *a priori* predicted to be known (20.5[18.2, 22.9]%) relative to pupil$_{max}$ response whilst listening to songs which were a priori predicted to be unknown (p = 0.012; see Figure 4.4). In the PCA group, the intercept in the PCA group was not significant (p=0.089) indicating only a 17[9.8,40.2]% change from baseline whilst listening to reference excerpts (songs which were *a priori* predicted to be unknown). However, in line with both control and AD findings, listening to songs *a priori* predicted to be known was associated with a larger PCA pupil$_{max}$ response (2.0[<0.001, 4.0]%, p = 0.048) relative to pupil$_{max}$ response elicited by the reference group. In both within-group models, similar pupil$_{max}$ coefficients associating pupil size and *gender, age*, *trial number and MMSE* were observed as were demonstrated in each group's respective Model$_{(pupil)}$1. Post-hoc testing removing *MMSE* from the model composition and re-running the analyses attenuated the marginal effect of *a priori* prediction on pupil responses in the PCA group only.

Between-group linear analyses revealed no evidence of an *a priori prediction x diagnosis* interaction on pupillometry responses (Table 4.10). The same profile of significant interaction effects was observed in this between-groups model as in the previous between-groups model.

140
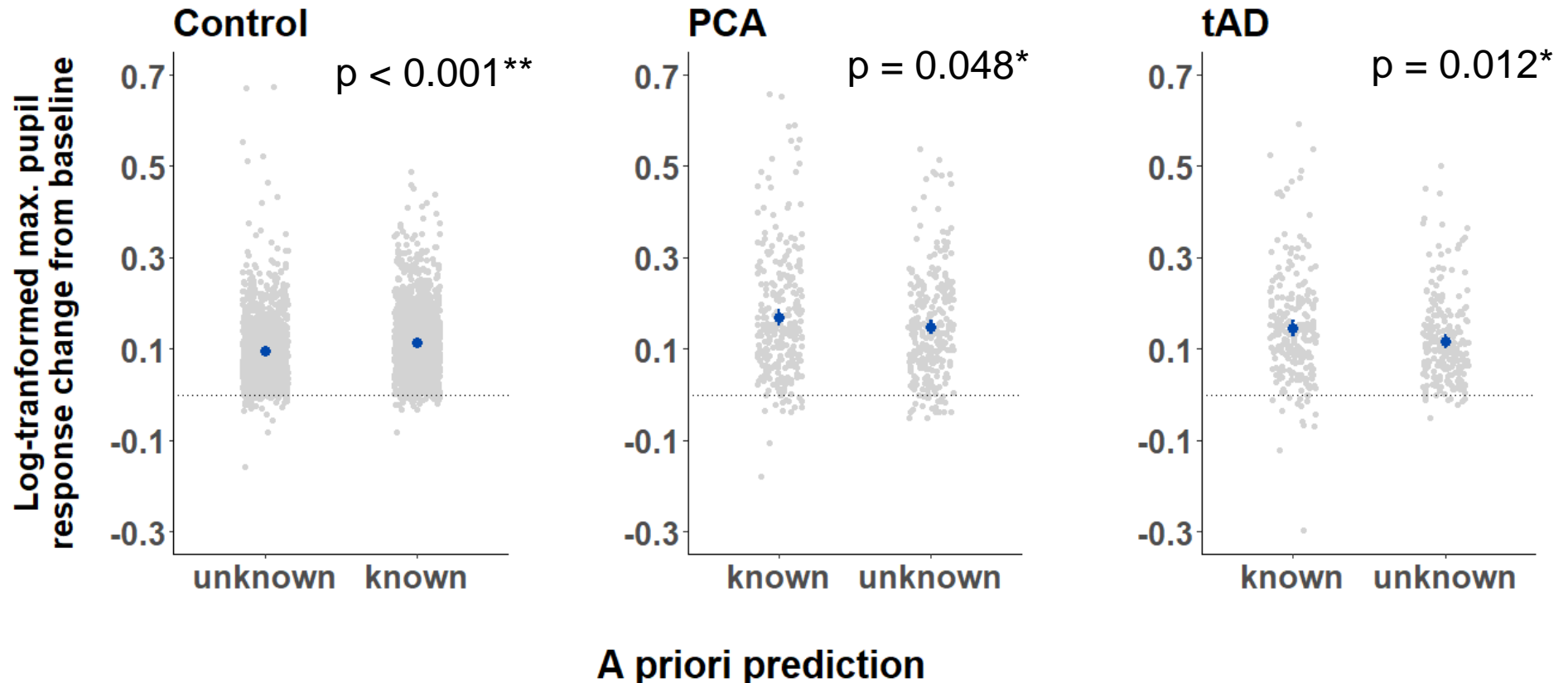
Figure 4.4 Log ratio of pupil$_{max}$ response change from baseline, plotted for music excerpts *a priori* predicted as either *known* or *unknown*, split by participant group. Points (·) indicate individual trial pupil$_{max}$ response change from baseline. Sample mean and 95% confidence intervals indicated in blue. Dotted line indicates pre-trial baseline for reference. P-values indicate within-group significance levels of the association between *a priori* predicted and pupilmax response, comparing *a priori* predicted *known* and *unknown* songs in each participant group. tAD: typical Alzheimer's disease; PCA: posterior cortical atrophy.

Table 4.9 Separate within-group Model$_{(pupil)}$2 regression analyses predicting log maximum pupil response by *a priori* familiarity indicator, setting, gender, age, average gaze co-ordinates and trial number in PCA and AD. Coefficients represent maximum percentage change in pupil response compared to baseline.

Model$_{(pupil)}$2 (within-group)	PCA		AD	
Predictor variable	Coefficient (exp)β [95% CI]	p-value	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.170 [0.976, 1.402]	0.089	1.175 [1.022, 1.352]	0.023*
Songs *a priori* predicted as known	1.020 [1.000, 1.040]	0.048*	1.026 [1.006, 1.046]	0.012*
Setting: Wellcome	0.980 [0.941, 1.021]	0.344	-	-
Trial index	1.001 [0.999, 1.002]	0.314	1.002 [1.001, 1.003]	0.002*
Age (decade)	0.998 [0.972, 1.025]	0.889	1.005 [0.991, 1.018]	0.522
Average gaze X (units: 100 pixels)	0.992 [0.981, 1.003]	0.150	0.992 [0.981, 1.003]	0.174
Average gaze Y (units: 100 pixels)	1.033 [1.024, 1.043]	<0.001**	1.008 [0.994, 1.023]	0.267
Gender: male	0.998 [0.976, 1.020]	0.825	1.006 [0.983, 1.029]	0.630
MMSE	0.995 [0.993, 0.997]	<0.001**	0.995 [0.993, 0.998]	0.001*
Number of observations.	*542*		*465*	

Table 4.10 Model$_{(pupil)}$2 (between-group) regression outcomes predicting change in pupil size from baseline from *a priori* predicted song familiarity. Interactions (where explored) are indicated by the addition of an interaction term adjacent to the predictor variable column.

Predictor variable (relative to control group)	Interaction term	Coefficient (exp)β [95% CI]	p-value
(Intercept)	-	1.209 [1.153, 1.267]	<0.001**
Songs *a priori* predicted as known	AD	1.007 [0.986, 1.028]	0.516
	PCA	1.001 [0.981, 1.022]	0.913
Diagnosis: AD	-	0.959 [0.862, 1.067]	0.445
Diagnosis: PCA	-	0.858 [0.790, 0.932]	<0.001**
Trial index	AD	1.003 [1.001, 1.004]	<0.001**
	PCA	1.001 [0.999, 1.002]	0.218
Average gaze Y (units: 100 pixels)	AD	0.995 [0.982, 1.007]	0.408
	PCA	1.031 [1.021, 1.041]	<0.001**
Number of observations	-	*3549*	

4.4.7.2 Pupil-familiarity responses based on subjective familiarity ratings

In line with controls, the intercept in the tAD group Model$_{(pupil)}$3 revealed a significantly larger pupil$_{max}$ response relative to baseline (14.8[0.2,31.6]%) while tAD participants listened to music in the reference group (excerpts subsequently rated as not familiar; see Table 4.11). This was not demonstrated in the PCA group (p = 0.112). In both patient groups, larger pupil sizes were significantly associated with familiarity levels, however the profile of these findings differed to the linear relationship observed in the control findings (see Figure 4.5). In the tAD group, a significantly larger pupil size was associated with songs subsequently rated as 'slightly familiar' (21.9[18.4, 25.6]%, p = 0.009) or 'extremely familiar' (23.6[20.7,26.5]%, p < 0.001) relative to songs rated as 'not familiar' (17.3[3.1,33.4]%). In the two within-group models, similar pupil$_{max}$ coefficients demonstrated an association between pupil size and *gender, age, trial number and MMSE* as were observed in each group's respective Model$_{(pupil)}$1-2.

Table 4.11 Separate within-group Model$_{(pupil)}$3 regression analyses predicting log maximum pupil response by familiarity ratings; significant predictive factors indicated within shading, main predictive variable of interest in bold.

Predictive factors	PCA Coefficient (exp)β [95% CI]	p-value	AD Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.160 [0.966, 1.392]	0.112	1.148 [1.002, 1.316]	0.047*
Familiarity Rating: Slightly familiar	1.022 [0.991, 1.055]	0.172	1.039 [1.009, 1.070]	0.011*
Familiarity Rating: Somewhat familiar	1.008 [0.964, 1.054]	0.727	1.032 [0.998, 1.066]	0.063
Familiarity Rating: Moderately familiar	1.032 [1.001, 1.064]	0.046*	1.023 [0.982, 1.066]	0.275
Familiarity Rating: Extremely familiar	1.020 [0.995, 1.045]	0.124	1.050 [1.024, 1.076]	<0.001**
Setting: Wellcome	0.986 [0.946, 1.027]	0.485	-	-
Trial index	1.001 [0.999, 1.002]	0.297	1.002 [1.001, 1.003]	0.002*
Age (decade)	0.999 [0.973, 1.026]	0.946	1.005 [0.991, 1.019]	0.510
Average gaze X (units: 100 pixels)	0.991 [0.980, 1.003]	0.136	0.993 [0.982, 1.003]	0.174
Average gaze Y (units: 100 pixels)	1.034 [1.025, 1.043]	<0.001**	1.008 [0.994, 1.023]	0.269
Gender: male	0.998 [0.976, 1.021]	0.870	1.001 [0.977, 1.026]	0.925
MMSE	0.995 [0.993, 0.997]	<0.001**	0.996 [0.993, 0.999]	0.005*
Number of observations	*542*		*465*	
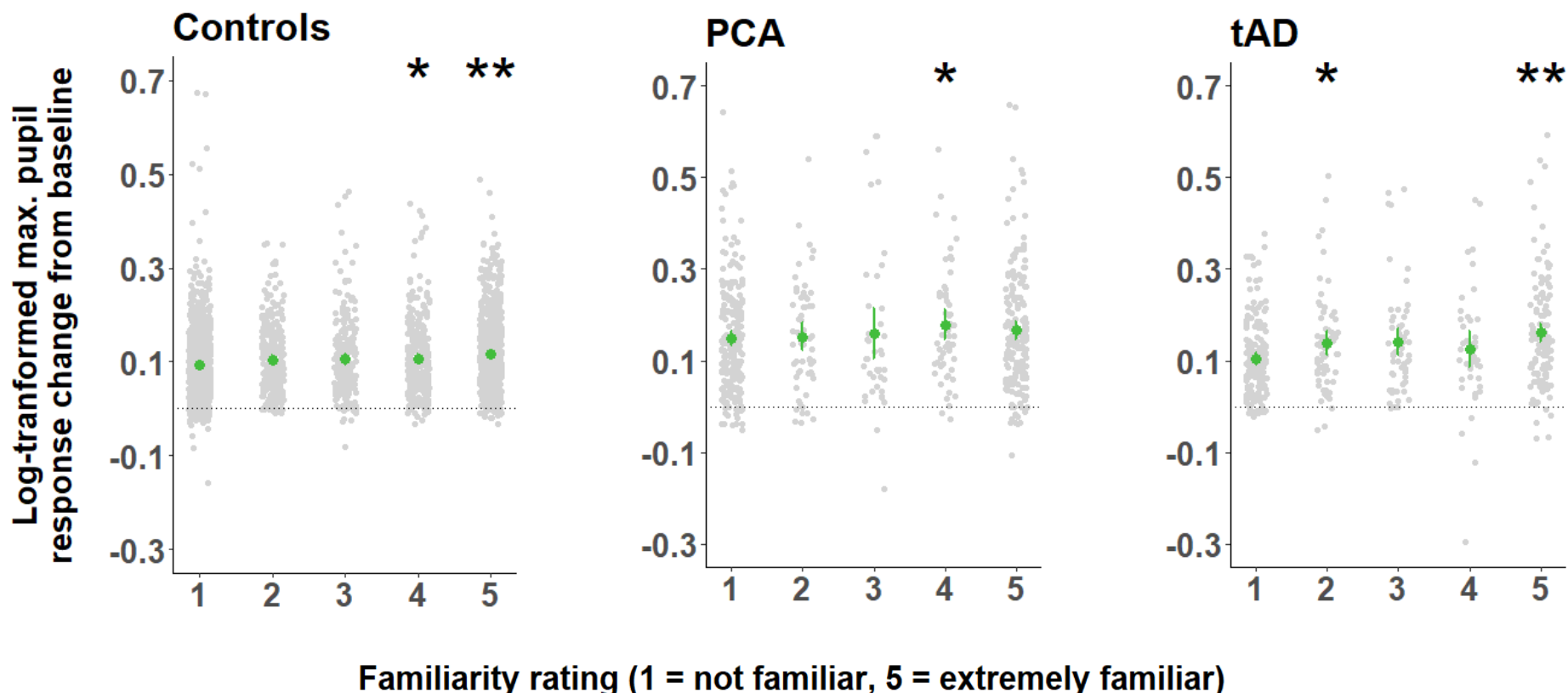
Figure 4.5 Log ratio of pupil$_{max}$ response change from baseline, plotted against subjective familiarity ratings, split by participant group. Points (·) indicate individual trial pupil$_{max}$ response change from baseline. Sample mean and 95% confidence intervals indicated in green. Dotted line indicates zero pupil response change from baseline, provided for reference. P-values indicate within-group significance levels of the association between pupil$_{max}$ response relative to songs rated as not familiar, in each participant group. tAD: typical Alzheimer's disease; PCA: posterior cortical atrophy.

As observed in Model$_{(pupil)}$2, post-hoc testing removing *MMSE* from the model composition and re-running the analyses attenuated the marginal effect of songs rated as 'moderately familiar' on pupil responses in the PCA group only.

Between-group linear regressions obtaining cluster robust standard errors revealed significantly larger pupil$_{max}$ responses in the AD participant group for music excerpts rated as 'slightly familiar' (3.4[0.3,6.6]%, p = 0.030) or 'extremely familiar' (2.7[0.1,5.4]%, p = 0.040) compared to control pupil$_{max}$ responses for songs which received the same ratings (see Table 4.12). Other interaction effects were consistent with the previous two between-group models

Table 4.12 Model$_{(pupil)}$3 (between-group) regression findings of log maximum pupil response predicted by familiarity ratings, diagnosis, age, setting, average gaze co-ordinates and trial number. Coefficients represent maximum percentage change in pupil response compared to baseline. Interactions indicated by the addition of an interaction term adjacent to the predictor variable column.

Predictor variable (relative to control group)	Interaction term	Coefficient (exp)β [95% CI]	p-value
(Intercept)	-	1.197 [1.141, 1.256]	<0.001**
Familiarity Rating: Slightly familiar	AD	1.034 [1.003, 1.066]	0.030*
	PCA	1.009 [0.976, 1.043]	0.597
Familiarity Rating: Somewhat familiar	AD	1.024 [0.990, 1.060]	0.169
	PCA	0.991 [0.947, 1.038]	0.704
Familiarity Rating: Moderately familiar	AD	1.012 [0.972, 1.053]	0.570
	PCA	1.013 [0.980, 1.047]	0.454
Familiarity Rating: Extremely familiar	AD	1.027 [1.001, 1.054]	0.040*
	PCA	0.999 [0.975, 1.025]	0.956
Diagnosis: AD	-	0.945 [0.854, 1.046]	0.277
Diagnosis: PCA	-	0.858 [0.789, 0.931]	<0.001**
Trial index	AD	1.003 [1.001, 1.004]	<0.001**
	PCA	1.001 [0.999, 1.002]	0.221
Average gaze Y (100 pixels)	AD	0.997 [0.984, 1.009]	0.585
	PCA	1.032 [1.022, 1.042]	<0.001**
Number of observations	*3549*		

4.4.7.3 Pupil-familiarity responses based on memory evocation ratings

Neither Model$_{(pupil)}$4 intercepts revealed a significantly larger pupil$_{max}$ response relative to baseline whilst PCA or AD participants listened to reference excerpts (music subsequently rated as not evoking a memory (PCA: p = 0.152; AD: p = 0.056; Table 4.13). However, a significantly larger pupil$_{max}$ response in the AD group was associated with songs subsequently rated as evoking a memory (21.6[18.8,24.5]%) relative to the pupil$_{max}$ response to the reference music (see Figure 4.6). In the two within-group models, similar pupil$_{max}$ coefficients demonstrated an association between pupil size and *gender, age*, *trial number and MMSE* as were observed in each group's respective Model$_{(pupil)}$1-3. Between-group linear regressions obtaining cluster robust standard errors revealed that, unlike the other models, no significantly different pupil$_{max}$ responses were elicited in the AD participant group for music excerpts rated as evoking a memory compared to control pupil$_{max}$ responses for songs which received the same ratings (p = 0.123; see Table 4.14). Other interaction effects were consistent with the previous between- group models. One difference in this model is the marginal finding of an overall pupil$_{max}$ response difference between the AD and control group just above the conventional level of statistical significance; the AD group elicited an overall 5.4[<0.001,10.5]% lower pupil$_{max}$ response relative to controls (p = 0.050).

Other findings of note:

4.4.7.4 Self-report familiarity responses predicted by a priori familiarity indicators

Initial within-group CLMM models were run separately for the PCA and AD group, including only *a priori familiarity prediction* and *trial number.* In this model, *a priori familiarity prediction* was associated with higher familiarity ratings in the patient groups (both p<0.001) with no main effect of *trial number* in the PCA group (p = 0.603) or in the AD group (p = 0.684). Running a second model with additional covariates (Model$_{(rating)}$3) did not change the significance level of the *prior familiarity prediction* or *trial number* in either group (see Table 4.15)*.* Classical, 1960s and 1970s songs had higher odds of elicited a greater familiarity response in PCA participants than nonsense songs; this pattern was not observed in the AD group. There was no effect of MMSE on familiarity response in either patient group.
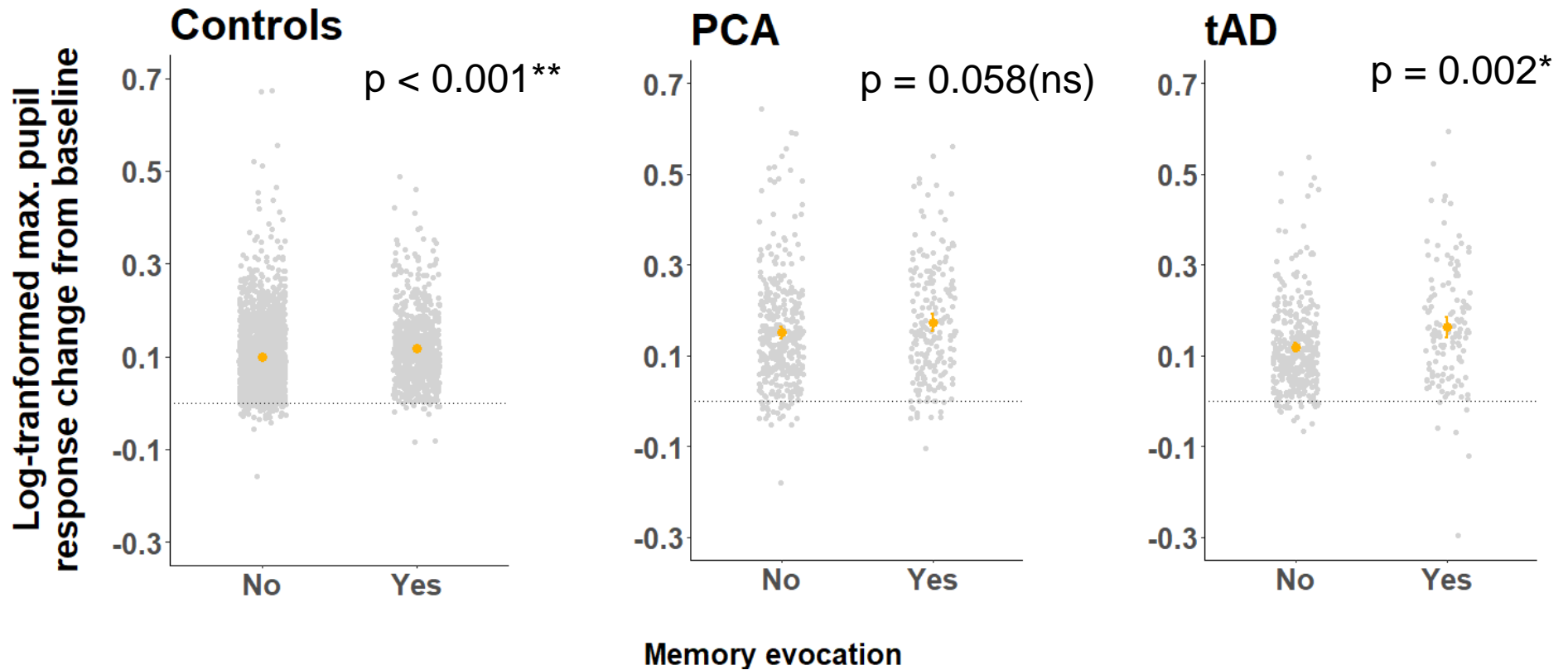
Figure 4.6 Log ratio of pupil_max response change from baseline, plotted against subjective memory evocation ratings, split by participant group. Points (·) indicate individual trial pupil_max response change from baseline. Sample mean and 95% confidence intervals indicated in orange. Dotted line indicates zero pupil response change from baseline, provided for reference. P-values indicate within-group significance levels of the association between pupil_max response relative to songs rated as not evoking a memory, in each participant group. tAD: typical Alzheimer's disease; PCA: posterior cortical atrophy.

Table 4.13 Separate within-group Model$_{(pupil)}$4 regression analyses undertaken predicting log maximum pupil response by memory evocation ratings; significant predictive factors indicated within shading, main predictive variable of interest in bold.

Predictive factors	PCA		AD	
	Coefficient (exp)β [95% CI]	p-value	Coefficient (exp)β [95% CI]	p-value
(Intercept)	1.145 [0.951, 1.379]	0.152	1.145 [0.997, 1.316]	0.056
Songs evoking a memory	1.022 [0.999, 1.045]	0.058	1.038 [1.014, 1.063]	0.002*
Setting: Wellcome	0.986 [0.947, 1.027]	0.506	-	-
Trial index	1.001 [0.999, 1.002]	0.304	1.002 [1.001, 1.003]	0.003*
Age (decade)	1.001 [0.975, 1.029]	0.924	1.006 [0.992, 1.020]	0.410
Average gaze X (units: 100 pixels)	0.992 [0.981, 1.004]	0.178	0.994 [0.983, 1.006]	0.319
Average gaze Y (units: 100 pixels)	1.033 [1.024, 1.043]	<0.001**	1.010 [0.996, 1.025]	0.165
Gender: male	0.997 [0.976, 1.019]	0.804	1.012 [0.989, 1.036]	0.302
MMSE	0.995 [0.993, 0.997]	<0.001**	0.995 [0.992, 0.998]	<0.001**
Number of observations.	*542*		*465*	

Table 4.14 Model$_{(pupil)}$4 (between-group) regression findings of log maximum pupil response predicted by memory evocation ratings, diagnosis, age, setting, average gaze co-ordinates and trial number. Coefficients represent maximum percentage change in pupil response compared to baseline. Interactions (where explored) are indicated by the addition of an interaction term adjacent to the predictor variable column.

Predictor variable (relative to control group)	Interaction term	Coefficient (exp)β [95% CI]	p-value
(Intercept)	-	1.206 [1.151, 1.264]	<0.001**
Songs evoking a memory	AD	1.019 [0.995, 1.044]	0.123
	PCA	1.000 [0.979, 1.023]	0.973
Diagnosis: AD	-	0.946 [0.895, 1.000]	0.050
Diagnosis: PCA	-	0.849 [0.808, 0.892]	<0.001**
Trial number	AD	1.003 [1.001, 1.004]	<0.001**
	PCA	1.001 [0.999, 1.002]	0.223
Average gaze Y	AD	0.995 [0.983, 1.008]	0.461
	PCA	1.030 [1.021, 1.040]	<0.001**
Number of observations.	*3549*		

Table 4.15 Separate within-group CLMM Model$_{(rating)}$3 outcomes for the PCA and tAD groups; significant predictive factors indicated within shading, main predictive variable of interest in bold.

	PCA		AD	
Predictor variable	*Odds ratio [95% CI]*	*p-value*	*Odds ratio [95% CI]*	*p-value*
A priori familiarity indicator (odds of a higher pleasant rating when song was *a priori* predicted to be known)	81.136 [33.497, 196.523]	<0.001**	25.835 [15.192, 43.935]	<0.001**
1960s songs (compared with nonsense songs)	7.581 [2.160, 26.604]	0.002*	1.694 [0.851, 3.373]	0.133
1970s songs (compared with nonsense songs)	7.144 [2.013, 25.350]	0.002*	1.728 [0.889, 3.362]	0.107
Classical songs (compared with nonsense songs)	3.684 [0.796, 17.060]	0.095	0.936 [0.411, 2.130]	0.874
Reminiscence bump song release	0.725 [0.439, 1.196]	0.208	0.871 [0.516, 1.471]	0.606
Trial number	0.993 [0.973, 1.014]	0.516	1.004 [0.984, 1.025]	0.684
MMSE	0.933 [0.859, 1.012]	0.096	0.969 [0.908, 1.034]	0.341
SD (Intercept Participant ID)	2.193	<0.001**	1.245	<0.001**
SD (Intercept Song ID)	2.442	<0.001**	1.792	<0.001**
Number of observations.	*615*		*515*	

4.4.7.5 *Self-report memory evocation responses: Associations with subjective familiarity ratings and genre*

Table 4.16 Frequencies and proportions of song evoking a memory, split by familiarity rating levels and participant group

Participant group	Familiarity Rating					
	Not familiar	Slightly familiar	Somewhat familiar	Moderately familiar	Extremely familiar	Total
Controls	25 (3%)	28 (4%)	45 (6%)	116 (15%)	558 (72%)	**772 (100%)**
PCA	17 (8%)	6 (3%)	10 (4%)	37 (16%)	153 (69%)	**223 (100%)**
AD	7 (4%)	8 (5%)	23 (15%)	31 (20%)	88 (56%)	**157 (100%)**

Given the low cell counts for songs which were reported to evoke a memory which were rated as not familiar or slightly familiar (see Table 4.16) the variable was recoded for the PCA and AD group ratings to reflect the control group recoding (whereby songs

rated as *not familiar* or *slightly familiar* were grouped into '*unfamiliar*' songs, and songs eliciting somewhat - extremely familiar songs were grouped as '*familiar*' songs). Running a logistic regression predicting memory evocation from *familiarity recoded, reminiscence bump and trial number* as separate within-group models for the PCA and AD groups indicated that only songs rated as *familiar* were significantly associated with higher odds of evoking a memory relative to songs recoded as *not familiar* (PCA: 34.69 [15.69, 76.72], p <0.001; AD: 58.65[24.18, 142.26], p <0.001).

To ascertain whether *genre* was associated with memory evocation ratings, separate within-group logistic regressions predicting memory evocation from *genre, reminiscence bump* and *trial number* (with random effects of participant and song) was undertaken on a subset of data retaining the genres 1960s, 1970s and classical music. No significant association of genre with memory evocation ratings was demonstrated in either the PCA or AD group.

4.4.7.6 Associations between self-report responses and musical properties/disease severity

Spearman's rho correlation examining the relationship between BPM and pleasantness ratings revealed a significant but weak negative association between excerpt tempo and pleasantness ratings in both the PCA (r_s = -0.18, p <0.001) and tAD (r_s = -0.19, p <0.001) groups. A simple regression within each group revealed that music key (major/minor) was not a significant predictor of pleasantness ratings in either the PCA (p = 0.517) or tAD (p=0.955) groups. Simple within-group linear regressions revealed an association of lyrical content (lyrics>instrumental) with pleasantness ratings in the PCA group (β = 0.495, p < 0.001) and tAD group (β = 0.408, p < 0.001)

Spearman's rho correlation examining the relationship between BPM and familiarity ratings revealed a significant but weak negative association between excerpt tempo and familiarity ratings in both the PCA (r_s = -0.13, p = 0.001) and AD (r_s = -0.11, p = 0.014) groups. A simple regression within each group revealed that music key (major/minor) was not a significant predictor of familiarity ratings in either the PCA (p = 0.120) or AD (p=0.150) groups. A simple regression within each group revealed an association of lyrical content (lyrics>instrumental) with familiarity ratings in the PCA group (β = 0.342, p = 0.022) but not in the AD group (p=0.122). The influence of

disease severity on self-report ratings was variable. In the PCA group, lower MMSE score was associated with higher pleasantness (β = -0.034, p < 0.001) and familiarity ratings (β = -0.029, p = 0.044). In the AD group, lower MMSE score was associated with higher pleasantness (β = -0.022, p = 0.049); but not familiarity (β = -0.009, p = 0.613) ratings.

4.5   DISCUSSION

The current study findings demonstrate evidence of differential pupillometry responses to music in people living with typical Alzheimer's disease in relation to song pleasantness and familiarity. This was demonstrated more selectively in the PCA group by marginal effects, evident in relation only to *a priori* predicted music excerpts. This is, to our knowledge, the first study demonstrating a coupling between psychosensory pupillometry responses and emotional/familiarity processing of music in people living with typical and atypical Alzheimer's disease. Despite the model intercepts indicating a lower overall level of pupil reactivity when listening to reference excerpts relative to controls, encouragingly the between-group models indicated no significant association of tAD diagnosis with maximum pupil response (noting the marginal finding in Model$_{(pupil)}$4).

4.5.1   Interpretation of results

[RQ1] Are self-reported pleasantness music ratings associated with pupillometry change from baseline in all groups?

RH1 was partially accepted. No associations between subjective pleasantness ratings and pupil responses were observed in the PCA group, however a significant curvilinear pupil-affect relationship was demonstrated in the tAD group in both modelling approaches, similar in profile to the healthy control group findings (Section 3.4.2). The dynamic indicates that tAD pupil responses reflect differential autonomic processing in response to emotionally-arousing music. This finding pervaded in a subsequent model (run post-hoc including *key* as a predictor variable, whereby major key-songs had been associated with increased pupil responses in the tAD group). As well as the

overall pupil-affect dynamic, the tAD group also aligned with control findings in that the only maximum pupil responses which reached statistical significance relative to the reference (neutral) songs were associated with excerpts subjectively rated as either 'pleasant' or 'very pleasant'. Noting the limitations outlined in relation to this in the previous chapter, it is encouraging that, whether these ratings truly reflect feelings of pleasantness or in fact differentiated levels of 'felt' arousal (see Section 3.5.3), it appears that in this clinical cohort at least, there is evidence for preservation of autonomic emotional processing of music.

The current study findings contrast with the decoupling of valence ratings and pupil responses observed in the tAD group in Fletcher, Nicholas, et al. (2015b). This finding was explained in the previous study as reflecting degradation of the ability to integrate information, owing to the compromise of the putative 'default mode network' (DMN) in tAD populations (Fletcher, Nicholas, et al., 2015b). The DMN is a resting-state network of disparate but interconnecting brain regions [i.e. medial prefrontal cortex; posterior cingulate cortex; bilateral inferior parietal lobes and hippocampi] which neuroimaging studies have shown co- and de-activate in tandem (Wilkins, Hodges, Laurienti, Steen, & Burdette, 2014). DiNuzzo et al. (2019) demonstrated via functional connectivity maps from neurotypical scans that pupil diameter change engages numerous cortical and subcortical regions. These include posterior cingulate cortex[PCC]-linked brain regions, which in-part form the DMN (e.g. (medial) superior frontal gyrus; Di Nuzzo et al., 2019). Importantly, disease severity in Alzheimer's disease has been shown to correlate with reduced functional connectivity in the DMN relative to healthy age-matched controls (Zhou et al., 2010).

The specific use of music in the current study - rather than environmental pleasant/unpleasant sound stimuli - may account for the contrasting pupil-affect findings. DMN activity is functionally associated with autobiographical memory and introspection (DiNuzzo et al., 2019). Listening to preferred music can act as a conduit to inducing these processes, as indicated by functional neuroimaging studies demonstrating that preferred music is associated with increased DMN connectivity (Wilkins et al., 2014). While the exact mechanism underlying the preserved pupil-affect profile in the tAD group remains elusive, the finding in response to musical stimuli implicates part-preservation of the DMN in this patient cohort at a relatively early

disease stage. Potential brain region candidates are those implicated in both music-related emotional processing and task-evoked psychosensory pupil responses.

The likelihood of differential emotional processing for music versus environmental sounds is increased when we consider the different between-group pleasantness rating findings in Fletcher, Nicholas, et al. (2015b) compared with the current work. Fletcher, Nicholas, et al. (2015b) reported lower ratings of sound pleasantness in the tAD group relative to the control and other dementia subtype groups. This was in contrast to the current study findings, whereby tAD participants rated the excerpts' pleasantness similarly to the control group. The self-report pleasantness findings in Fletcher, Nicholas, et al. (2015b) align with other study results, whereby all individuals with tAD with auditory hedonic changes since diagnosis (N = 7/17) demonstrated a uniformly decreased liking for environmental sounds. However, findings from this second study also highlighted a difference between aversion to environmental sounds, and music aversion, which was reported only for two out of the seven participants in Fletcher, Nicholas, et al. (2015b). This seems to be reflective both of the current self-report findings, whereby the pleasantness ratings (overall and for each individual song, bar one), and pupillometry-affect curvilinear dynamic were similar in the tAD group as observed in the controls. This evidence aligns with previous findings of preserved music reward valuation (pleasantness ratings) in individuals with tAD (C. N. Clark et al., 2018) and supports the notion that individuals with tAD may experience more of a disease-related change in their emotional responses to environmental sounds, while the emotional appraisal of music appears to be relatively retained.

[RQ2-4] Is a song's familiarity associated with pupil response change from baseline in all groups?

In line with younger and healthy older adults, modelling pupil response from *a priori* predicted levels of familiarity revealed a preservation of the pupil old/new effect profile in both patient groups (songs *a priori* predicted to be known > a priori predicted unknown songs) [RH2]. While this effect in the PCA group was marginal, the relatively larger pupil size for *a priori* known songs was in line with the tAD and control findings, evidenced in the between-groups model by a lack of diagnosis-'familiarity prediction' interaction effect. This adds to the evidence base for preserved semantic musical memory in typical and atypical Alzheimer's disease (e.g. Slattery et al., 2019), and extends these findings to demonstrate preservation of psychophysiological processing

153

in the tAD group, and some evidence to support preservation in the PCA group in relation to predicted familiar and unfamiliar songs.

RH4 was partially accepted by virtue of differential pupil responses to memory ratings being observed in the tAD group only. Episodic musical memory is widely-reported to be impaired in individuals with Alzheimer's disease (Slattery et al., 2019); therefore, the finding of preserved tAD pupillometry psychosensory responses to excerpts subjectively identified as memory-evoking was somewhat unexpected. These findings must however be discussed in the context of querying the face validity of the affiliated prompt (i.e. that the memory evocation question was interpreted in a similar way by each participant, in the way intended by the research team). Episodic musical memory is challenging to assess for songs that are repeatedly played over the life-course, as repeated exposure distorts the particular time-space context of when it was first experienced. As a result, the episodic memory for music can become blurred or lost even in the absence of disease. Groussard et al. (2019) termed this the 'semantisation' of musical episodic memories; it is therefore possible that the memory evocation question simply delineated responses reflecting semantic musical memory processing.

RH3 was tentatively accepted owing to the marginal pupil response-familiarity finding in the PCA group and the enlarged pupil responses to both slightly- and extremely-familiar-rated songs in the tAD group. While the tAD group did not elicit the same cumulative increase per familiarity level as demonstrated in the controls, the largest pupil response in this model was observed for songs subsequently reported as 'extremely familiar', which was similarly observed in the control group findings. The marginal results specific to the pupillometry findings in the PCA group (in the presence of relatively spared self-report ratings; see below) may allude to disruption in the association cortices with which an individual's supplementary eye field (SEF) is able to engage to elicit unconscious differential pupil responses while appraising music. Goodwin (2007) highlights an association of SEF activity and implication of posterior parietal cortex, a hallmark target of PCA (Crutch et al., 2012).

[RQs 5-6] Are independent classifications of a song's familiarity [RH5] and/or subjective ratings of a song's familiarity [RH6] associated with patient groups' pleasantness ratings?

RH5 was accepted in both patient groups; songs were more likely to be rated higher in pleasantness if they had been *a priori* predicted to be known (with similar odds ratios in the PCA group as observed in the control group). RH6 was partially accepted; the PCA group observations of a linear increase in odds of rating an excerpt as pleasant with increasing feeling of familiarity reflected control findings, and is therefore likely to be interpretable in the context of the mere exposure effect (*MEE*; Section 3.5.1). MEE is reported to be preserved in individuals with mild-moderate AD for presentation of faces (Willems, Adam, & Van der Linden, 2002); evidence of the effect in the auditory modality, however, remains less clear. Quoniam et al. (2003) reported preserved MEE in tAD following re-presentation of unfamiliar musical melodies either 1, 5, or 10 times. However, individuals with tAD did not elicit the MEE when re-exposed to previously-unfamiliar songs only once (Halpern & O'Connor, 2000), and individuals with tAD and age-matched controls scored poorly (but above chance) in terms of demonstrating the MEE when the same stimulus set was re-presented three times (Vanstone et al., 2012). A recent study involving 26 individuals with mild-moderate dementia (cf. 20 healthy older controls) repeatedly presented avant-garde music clips three times a week demonstrated no MEE effect was observed in either group for aggregated (vocal/instrumental) pleasantness ratings (Felisberti, 2021). Post-hoc latent growth modelling however revealed an increase in affect for the excerpts after repeated exposure at the individual level (Felisberti, 2021). The attenuation of effect in the aggregated findings was attributed to repeated exposure dampening hedonic responses by the third week of repetition (and this can be observed particularly in the dementia population in response to vocal music). This indicates that a careful balance of timed exposure must be struck in terms of playing familiar songs to individuals with dementia in order to elicit hedonic responses. Notably, none of these studies have sought to differentiate MEE between typical and atypical Alzheimer's disease groups. A similar positive correlation between pleasantness and familiarity behavioural responses was observed in the PCA and AD group, in line with healthy older adult findings. Given the similarity of direction and magnitude of correlation in both patient groups relative to healthy older adults, it stands to reason a similar interpretation can

be made as in the previous chapter concerning potential implications for the pupillometry results.

[RQ7] Are independent classifications of a song's familiarity associated with likelihoods of reporting that song as familiar in all groups?

RH7 was accepted; both patient groups reflected the control findings and demonstrated higher odds of rating a song higher in familiarity if it had been *a priori* predicted to be known (noting the remarkably similar odds ratio elicited in the PCA group as observed in the control Model$_{(rating)}$3b). While the self-rated responses were not collected to be scored in terms of a hit/false alarm design, the preservation of this finding in both the PCA and tAD groups is promising in terms of this paradigm's ability to demonstrate face validity.

[RQ8] Are subjective judgements of a song's familiarity associated with their likelihood of reporting that song as evoking a memory in all groups?

RH8 was accepted; both patient groups reflected the control findings, demonstrating higher odds of reporting a song had evoked a memory if they had subjectively rated it as familiar. Noting that the five-point familiarity scale was recoded (collapsing the levels into a binary outcome), this nevertheless indicates retained task comprehension in the patient groups, in that few 'illogical scores' (i.e. 'not familiar' rating → positive memory evocation rating) were made in these clinical groups. Examining the contingency table comparing familiarity ratings ascribed to songs which were reported to have evoked a memory, the tAD group rates in particular were aligned with control proportions of illogical ratings. The PCA group proportion of illogical ratings were slightly higher than both the tAD and control groups, however this was not statistically interrogated.

[RQ9] Are songs released within participants' putative 'reminiscence bump' (between 13 and 19 years old) associated with subjective music ratings in all groups?

The findings in all within-group rating models revealed no association in either patient group of reminiscence bump release on pleasantness or familiarity. These findings, in keeping with the control group observations, uphold the interpretations outlined in the previous chapter discussion (Section 3.5.1.) and strengthen the considerations around selecting music for people living with typical or atypical Alzheimer's disease that may not necessarily focus on release date within a person's adolescence.

156

4.5.2   Study strengths

The study design considered and in-part addressed criticisms of previous studies in the dementia-music-physiology field. Firstly, the in-depth characterisation of the participant groups, and furthermore the lack of evident covariates between them - strengthens the likelihood of experimental findings relating to participant group rather than another candidate demographic variable.

The music experience questionnaire and self-reported ratings revealed interesting findings, which may be helpful to further elucidate the role of music in supporting people living with tAD or PCA. The observation of lyrical content being associated with higher pleasantness ratings in both patient groups (and with higher familiarity ratings in the PCA group) reflects previous reported gathering liking ratings comparing vocal and instrumental music in dementia populations (Felisberti, 2021). Particularly pertinent to the tAD group is the finding that music with lyrical content aids musical memory by virtue of engaging additional networks involved in processing language (Baird & Samson, 2009). Findings from the music experience questionnaire indicated that intense music listening of 10+ hours per week to be increased in both patient groups, which could be interpreted perhaps of people with dementia either 'using' music more than neurotypical peers who are not living with the condition and/or replacing lost or relinquished activities with music listening. This increase in time dedicated to music listening may account for the increased propensity of the PCA group to report a song as evoking a memory relative to the control group. Longer time dedicated to music listening may have increased the chances of a more recent rehearsal of a memory linked to a song in the stimulus set. Whilst both patient groups listened to music more often than healthy older adults, for people living with PCA, episodic memory impairments are not a primary clinical feature (Schott & Crutch, 2019). Accurate reporting of memory evocation may however have been impacted in the tAD group owing to episodic memory difficulties encountered even in the early stages (Schott & Crutch, 2019). Complementary to this is the current study finding that individuals with PCA demonstrated more diverse music listening habits relative to control participants than Alzheimer's disease, which – coupled with the increase of music listening time per week relative to healthy older adults – may account for the

higher pleasantness ratings demonstrated in the PCA group in comparison to the other participant groups. Auditory content may prove less problematic for individuals with PCA to interact with, relative to the challenges they face when presented with visuospatial information. A recently-published video of one person living with PCA asserts that favourite music can help a person with this condition stay connected to the world (Alzheimer's Research UK, 2022). The expanded repertoire in the PCA group could either reflect pre-morbid group differences, or hint at this group's desire to capitalise on a modality that is relatively spared at least in earlier stages of their condition. Assessing pre-morbid music preferences and subsequent changes using a similar approach to (Fletcher, Downey, et al., 2015) in future work with PCA participants may elucidate the factors contributing to these findings.

The fact that disease severity (indicated by MMSE score) was incorporated in the data collection enabled us to investigate the impact of this potential confound on the current findings. This was particularly important in the PCA group, whereby post-hoc testing revealed that marginally significant findings (e.g. in Models$_{(pupil)}$2 and 3) were attenuated when MMSE was not controlled for. The tAD results however appeared resilient to model change. The changeable pupillometry findings in the PCA group with or without MMSE incorporated into modelling warrants the inclusion of this metric within psychophysiological studies, in line with recommendations from the literature review to ascertain the influence of disease severity (Chapter 2). An increased pupil size associated with lower MMSE score perhaps reflects the higher cognitive effort needing to be exerted by individuals with relatively more severe symptoms in order to complete the categorisation task.

Trial number was also significantly associated with significant increases in pupil size in the tAD group over the course of the experiment, in direct contrast to controls where a decrease was observed. The increase in maximum pupil response could indicate a cumulative effect of music listening on autonomic arousal (thus inhibiting pupil constriction which is innervated by the parasympathetic nervous system). This may however reflect a greater amount of cognitive effort required in the tAD group to continue engaging with the task at hand, relative to the cognitive effort required by the healthy age-matched controls to continue appraising the musical excerpts. The cognitive effort exerted (if this is indeed what accounts for the trial number association

finding) was not sufficient to mask the pupillary response indexed by the emotional and cognitive appraisals of the music, and therefore the task can be regarded as relatively low-burden for individuals with dementia to undertake (in line with assertions from Gingras et al., 2015). This interpretation is supported by findings of increased pupil size in AD participants relative to MCI and healthy control participants during a digit span task whereby AD task performance was normal (Granholm et al., 2017). It must be noted however that in Slattery et al. (2019) whereby a similar familiar/unfamiliar music paradigm was employed, the preserved behavioural performance in both patient groups did not correlate with the observed profiles of brain activation, indicating the complexity of the brain-pupil relationship in response to music which may only be elucidated by concurrent pupillometry-fMRI paradigms.

While formal hypotheses were not generated regarding each acoustic property ascertained in the current study, exploratory analyses indicated that musical properties did not appear to be as influential on patient groups' pupillometry or self-report ratings as perhaps initially envisaged given previous studies in this clinical population (Garrido, Stevens, et al., 2019; Walker et al., 2021). There was also no effect of lyrical content on pupillometry responses in either patient group, in contrast to findings reported in neurotypical adults (Weiss et al., 2016). It was however interesting that an effect of music key in the tAD group pupillometry responses was demonstrated, which was not evidenced in the self-report ratings. This may reflect an unconscious cognitivist processing being undertaken (e.g. differentiating between sad and 'happy' music) rather than the emotivist intention behind the current paradigm to establish a person's felt emotion. Raglio, Bai, et al. (2022) and Särkämö (2018) highlight that music can enhance cognitive performance in healthy older adults by virtue of selecting various musical properties (e.g. fast tempo), and that this could also be applied to PLWD to target desired functional ADL outcomes (e.g. to ease the process of dressing). Raglio, Bai, et al. (2022) asserts that incorporating music in dementia care and support this way is preferable with a trained MT only. The fact that tAD pupillometry-affect effects remained the same regardless of the inclusion of *key* in Model$_{(pupil)}$1 is promising in the potential use of using candidate biosignals in future therapeutic application, in that these (and the familiarity-associated) physiological changes seem genuinely to be reflecting 'felt' appraisals of music, rather than simply reflecting differences in acoustic properties. This interpretation is also mirrored in

neuroimaging network scientific findings, whereby 21 young adults listening to subjectively-rated liked and disliked music demonstrated differential brain network connectivity patterns in the DMN regardless of acoustic properties such as lyrics (Wilkins et al., 2014).

Neuroanatomical correlates identified in individuals with and without Alzheimer's disease whilst listening to novel music indicates that the processing of unfamiliar excerpts is not simply the absence or even obverse activation in the brain regions and networks engaged when listening to music we know (Slattery et al., 2019). The nonsense tracks in the current paradigm were included to provide truly novel stimuli which remained musically sensible. However, it is possible that these *de novo* excerpts may have been driving some of the marked effects seen by virtue not of being unfamiliar, but by being devoid of 'meaning', as indicated by preserved tAD pupillometry responses to meaningful versus meaningless sounds which were created in a similar way (Fletcher et al., 2016). Within-group and between-group analyses were rerun to investigate whether the de novo excerpts were in fact driving the results (by removing the nonsense tracks from each model). Encouragingly, there were no remarkable changes to the overall findings in the control or patient groups. This indicates that even using exclusively ecologically valid stimuli - provided it is well-selected - it is possible to observe differential pupillometry responses to emotional and familiarity appraisals of music, in individuals living with tAD.

4.5.3 Study limitations

In addition to limitations previously discussed (see Section 3.5.3), there are caveats to the current between-group findings by virtue of two experimental sites. The fact that controls were mostly tested in one site, and patients in another, indicates that any patient group-control pupillometry interaction effects observed here must be interpreted with caution. The application of the study does however largely rely on reported within-group effects, reducing the cause for concern here. It was also reassuring to see group-by-pupil-size interactions observed not at the syndromic group-control level (which may have supported the argument that experimental site was more influential), but between only the PCA and control groups, whereas tAD

overall pupil responses were comparable to the control group. These differential group-pupil interactions were observed in each model. Whilst entirely clear that any pupillometry work should avoid confounds of between-group luminance differences in experimental settings, where this cannot be avoided (as in the current work), interspersed - rather than contiguous group-by-group - recruitment of control and patient participants may have mitigated this.

The between group differences in each patient group relative to controls may result from a number of factors. One proposal is that the different pupil-affect responses in the PCA group relative to the tAD and healthy older adult pupil profiles could be accounted for by the more diverse music taste exhibited by the PCA group relative to the healthy older adults. Wider exposure to a variety of music genres may give rise to greater capacity to identify differences in musical properties, and therefore the pupil response in the PCA group may be mediated to a greater extent by the unconscious processing of these properties relative to subjective or 'felt' appraisal of each excerpt. However, the tAD group also elicited a similar music listening frequency and diversity to the PCA group (i.e., more frequent music listening per week than healthy older adults, and a more diverse music selection), and still elicited similar pupil profiles relative to healthy older adults. It therefore appears as though other factors may be contributing to the different pupil profiles demonstrated in the PCA group which relate to the syndrome itself. For example, between-syndromic-group profiles exist in early auditory perceptual impairments (e.g. auditory scene analysis, impaired in PCA relative to tAD and healthy control participants) (Hardy et al., 2020). Other methodological considerations may however account for the current findings. For one, the PCA findings may reflect the complexity of implementing physiological measures of in-the-moment responses (e.g. pupillometry) in individuals with oculomotor challenges. A relatively commonplace limitation in pupillometry studies is the challenge of reconciling deviations of fixation from the central point on the screen. This occurrence, known as the Pupil Foreshortening Effect (PFE; whereby the pupil size diminishes as gaze deviates from a central fixation point by virtue of the eye rotating), was partially attenuated in the current study by pupillometry pre-processing steps, replicating the approaches in Benhamou et al. (2021). These steps aimed to reduce the prevalence of extreme pupil responses, which are more likely to reflect deviations from the central fixation point than influence of experimental manipulation. Other

studies investigating pupillometry in younger neurotypical populations (e.g. Jagiello et al., 2019) have been more stringent and applied an exclusion of any pupil response fixating more than one degree of visual angle from the central fixation point. This was attempted in Play It Again, but was not however a tenable approach owing to large proportion of fixations outside this perimeter, particularly in the clinical populations.

Despite participants' pupil sizes reducing by virtue of position on the X-axis, X-axis co-ordinates were not significantly associated with maximum pupil response. This indicates that horizontal saccades were not particularly prevalent in the patient groups. Position on the Y-axis however, reflecting vertical eye movements, were significantly associated with maximum pupil response in the PCA group. In addition, the interaction of diagnosis and average Y-axis gaze position was demonstrated in all between-group models between the PCA and control group. This indicates that vertical fixations elicited in the PCA group influenced the recordable pupil size acquired in the current dataset. This perhaps reflects previously-reported PCA-control group differences in fixation stability owing to increased saccadic intrusions (involuntary deviations from a fixation point) (Shakespeare, Kaski, et al., 2015) and other between-group differences in eye movement and fixation behaviour in PCA compared to tAD and healthy controls (Pavisic et al., 2017; Shakespeare, Pertzov, Yong, Nicholas, & Crutch, 2015; Shakespeare et al., 2013).

However, this association of pupil response and Y-axis position was in an unexpected direction; that is, position on the Y-axis reflected an *increase* in pupil responses as a function of increased position of the Y-axis, in direct contrast to the expected Pupil Foreshortening Effect. Given this finding was consistent only in the Y-axis, it seems sensible to propose that this may be related to within-trial blink behaviours leading to an overall difference in trialwise averaged Y-axis co-ordinate position. For example, the findings may reflect the increase in pupil area recordable by the infrared camera as the eyelid opens after a blink and it searches for the fixation cross on the vertical axis. Despite pre-processing steps being taken to remove data after blinks to avoid this, perhaps the 100ms selected was too small a timeframe to remove all blink-related pupil response phenomena in the PCA group, leading to some change in pupil response due to Y-axis gaze position being unintentionally retained. If this

interpretation is correct, it at least serves to highlight the complexity of pre-processing pipeline decisions in application to older and/or clinical populations.

A further limitation relates to the reliability with which individuals with dementia (and in particular those with PCA) are able to interact with Likert scale prompts, owing to a combination of spatial, visual imagery and working memory deficits. This may account for the current findings of greatest variability within the PCA group in pleasantness and familiarity response ratings, and has ramifications for the consideration of how researchers can be confident they are establishing 'ground truth' music response experiences in this group of individuals, rather than responses reflecting particular challenges in interacting with the data collection tools.

Music is not the only stimulus to elicit the pupil old/new effect; this PDR profile has been demonstrated in neurotypical populations in other types of auditory stimuli via aural presentation of words/nonsense words (Otero et al., 2011) and across modalities (visually-presented words/nonsense words (R. M. Gardner, Mo, & Borrego, 1974); sensible/novel consonant trigrams (R. M. Gardner, Mo, & Krinsky, 1974); famous/novel faces (Maw & Pomplun, 2004)). As observed previously, the pupil-affect relationship has been demonstrated thus far using environmental sounds (Partala & Surraka, 2003; Fletcher et al., 2015b). One limitation of the current paradigm design relates to the extent to which the paradigm included sufficient control or reference clips to establish the specificity of the effects of music on the pupil and behavioural responses. The Play It Again paradigm incorporated nonsense excerpts created from each group of the stimuli set (1960s known/unknown, 1970s known/unknown and classical) to create a reference group of clips used in the behavioural modelling when establishing the association between music genre and pleasantness/familiarity ratings. However, these so-called 'nonsense' clips were still classifiable as musical material (e.g. incorporating pitch and rhythm) and therefore it is difficult to establish the specificity of music relative to other stimuli (e.g. soundscapes) on the observed pupil and behavioural profiles. Manipulating the stimuli excerpt properties to produce a non-musical sound may have served as a more ecologically valid pre-trial baseline period than the silent baseline period incorporated into the paradigm. An analogous approach has been taken up in visual pupillometry paradigms whereby visual stimuli

are pixelated and presented as the pre-trial baseline to establish stimulus specificity on pupil results (Franzen, Cabugao, Grohmann, Elalouf, & Johnson, 2022).

4.5.4 Future direction

Subject to replication and extension of the current work in larger and more varied samples (e.g., with other forms of dementia), there is a potential for future application of this work in individuals with more severe forms of dementia, seeking to establish physiological responses as a proxy for self-reported subjective musical experiences. It must be entertained that, whilst individuals with Alzheimer's disease appear to have relatively preserved psychophysiological responses in early-mid stages of the disease, this may not be replicable in individuals with more severe symptoms. Jacobsen et al. (2015) cited that, whilst they observed retained structural and glucose-metabolism integrity in cortical regions involved in musical memory, the observation of amyloid-B deposition increases in these same regions (by virtue of florbetapir-PET scanning) indicated likelihood of these areas succumbing to eventual cortical atrophy as the disease progressed. Large-scale studies of musical semantic memory in 50 individuals with AD have however revealed spared abilities in individuals living in the later stages of the condition (Cuddy et al., 2012). In the current findings, familiarity ratings were overall not associated with MMSE score in the tAD group, indicating that the ratings were not influenced by factors such as disease severity and supporting the notion of relative preservation. Concurrent important work is underway to extend our understanding of the impact of personalised music listening interventions for people living with later-stage disease, for example in people living with cancer in hospice settings (B. Johnston et al., 2022). Wider-scale explorations are however required in people living with severe neurodegenerative conditions (noting the methodological challenges such a study would encounter).

Overall, the Play It Again paradigm has demonstrated relatively preserved pupil response profile couplings with affective and familiarity appraisals of music in people living with typical Alzheimer's disease, and to a lesser extent in individuals with the visual variant of the same disease. Whilst these initial findings are promising, methodological limitations of testing individuals with oculomotor challenges must be

reconciled in future iterations of the paradigm in order to ascertain whether evidence for these effects in people with PCA can be strengthened.

4.6   CONTRIBUTIONS & ACKNOWLEDGEMENTS

As before (see Section 3.6)

5   LOW-BURDEN PUPILLOMETRY MUSIC PARADIGM DISTINGUISHES BETWEEN TYPICAL AND ATYPICAL ALZHEIMER'S DISEASE: ARTIFICIAL INTELLIGENCE (AI) IN DEMENTIA CARE COLLABORATION

5.1   *ABSTRACT*

Interest in, and motivation to explore, data-driven approaches to differentiating dementia diagnoses has expanded rapidly in recent years (Ranson et al., 2023). While the majority of work to-date relies on neuroimaging data, we propose that pupillometry may serve as a potential biomarker candidate, which machine-learning methods may be able to harness to discriminate typical and atypical Alzheimer's disease. The work herein describes a secondary analysis of the Play It Again patient pupillometry dataset and automatically-captured latency metrics using machine-learning approaches. Eighteen pupil features, alongside latency metrics of the familiarity ratings (e.g., two reaction times) were acquired in 18 individuals with typical Alzheimer's disease and 21 individuals with PCA whilst listening and rating music excerpts, and subjected to algorithm development using a random forest classifier. Discrimination performance was ascertained for four models. Two different thresholds of missing data acceptance were explored; for each threshold, the model's sensitivity and specificity were established with (fusion model) and without (pupil-only model) the inclusion of the latency metrics. Area Under the Curve Receiver Operating Characteristics (AUROC) analyses of the four models revealed excellent performance discrimination achieved by the fusion models in both missing data threshold levels (AUROC = 0.81), and acceptable comparable discrimination performance in the pupil-only models (relaxed threshold AUROC = 0.78; stricter threshold AUROC = 0.76). Findings are discussed in the context of utilising passively-collected physiological data in dementia populations to serve a dual-purpose of algorithm training and intended experimental manipulation to reduce participant burden. Results are considered as proof-of-principle, with the current sample size limiting more confident conclusions.

5.2   INTRODUCTION

5.2.1 Differentiating syndromic groups using artificial intelligence and machine-learning methods: An overview

It is more important than ever to explore (and exploit) the capacity of different approaches which can help to make sense of the often-puzzling symptom profiles the different dementia syndromes can elicit. A sense of urgency has increased in recent years due to a number of important factors relying on timely dementia diagnoses, including: (a) being given the opportunity to enter clinical drug trials; (b) receiving any future-licensed disease-modifying treatment, or (c) (for those whereby symptom management is the only option) being provided with the most appropriate and relevant condition-specific information and support. Data-driven methods, such as machine-learning (ML) and artificial intelligence (AI) may be harnessed to accompany clinical acumen, by virtue of their capacity to 'learn' or classify different health conditions based on a large number of observations from different types of data (Johnson, McWhirter, et al., 2021; Qiu et al., 2020). A recent review by Ranson et al. (2023) highlighted that, to date, much of the work undertaken to classify different dementia syndromic groups from healthy controls has focused on multimodal datasets; in over 250 of these, the predominant dataset is acquired via neuroimaging (alongside accompanying cross-sectional participant demographic and clinical information). Neuroimaging data are incredibly complex, large datasets; for these reasons, they are suited to machine-learning methods by virtue of the large observation n which provides sufficient training and testing information for algorithm development (Ranson et al., 2023). Furthermore, the intricacy of the information acquired during scanning outstrips the insights that traditional statistical approaches can provide (Ranson et al., 2023). Multimodal deep learning frameworks recently used full volumetric T_1-weighted MRI scans, participant's age, gender and MMSE scores to discriminate not only patients from controls, but between cognitive impairment etiologies [MCI, AD, and non-AD dementia such as LBD and FTD] (Qiu et al., 2022). Discrimination performance of this 'fusion model' in classifying individuals with non-Alzheimer's disease dementia was acceptable in its ability to conduct LBD-tAD classification by virtue of the AUC value being between 0.70 and <0.80 (AUC ~ 0.75 for both datasets used) and ranged from 'acceptable' to 'excellent' (0.80 < AUC <0.90) in its ability to conduct an FTD-tAD classification

depending on the dataset used (Qui et al., 2022). Lampe et al. (2023) recently used a similar approach in an attempt to discriminate individuals with bvFTD, tAD, and the three PPA subtypes (lvPPA, svPPA and nfvPPA) in a multiclass classification, with relative success (prediction accuracies of 71-95%).

Whilst these endeavours hint at the promise of artificial intelligence to elucidate dementia diagnoses, it is important to note that a range of potential individual contraindications prevent a person from being able to undergo structural neuroimaging [e.g. metal in the body, claustrophobia]) (Bonifacio & Zamboni, 2016). Interestingly, Qiu et al. (2022) explored a 'non-imaging' model comprising features such as MMSE, past medical history and neuropsychological test scores, which boasted acceptable-to-excellent discrimination performance between individuals with non-Alzheimer's dementia (FTD, LBD, VD) and those with tAD (accuracy = 0.786 [Dataset 1]; 0.806 [Dataset 2]) (Qiu et al., 2022). The authors conceded that, while the non-imaging data model approached a fusion (MRI + non-imaging data) model accuracy, the MRI scan was integral to achieving a consistent 'excellent' discrimination performance in classifying AD from nonAD subtypes. Neuropsychological test batteries, which formed part of this model, are designed to selectively and systematically probe a range of cognitive tasks. In the context of the current work, they typically hone in on performance in visuospatial and mnemonic domains to elucidate differences between PCA and tAD. However, whilst neuropsychological testing is no doubt an extremely useful tool, they are lengthy to administer, and some commonly-used screening tasks induce anxiety for individuals living dementia, influencing performance (Keady & Gilliard, 2002). Given the contraindications for MRI, and the suggestion of lower model accuracy using non-imaging data, it is therefore important to consider other types of multimodal datasets which can be acquired in different syndromic dementia groups with relative ease and efficiency, but which also yield high numbers of observations that fulfil requirements for algorithm testing and development.

5.2.2 *Machine-learning: Applications to pupillometry*

Pupillometry data, particularly those acquired at a high sampling rate (up to 1000Hz) may provide an alternative to neuroimaging data in differentiating dementia etiologies. Indeed, a number of studies have used pupillometry data to discriminate participants

with dementia from healthy controls. Receiver Operating Characteristics (ROC) analyses of regression models applied to pupillometry and eye-tracking data acquired in people with Alzheimer's disease and healthy controls during a short-term memory binding task revealed 100% sensitivity and specificity in discriminating the two groups (Fernández & Parra, 2021). Importantly, normalised pupil size outperformed the classification prowess of both behavioural responses to the cognitive task [sensitivity: 83%; specificity: 81%] and other automated eye-tracking data acquisition such as gaze duration [sensitivity: 73%; specificity: 74%], indicated by the significantly higher area under the curve (AUC) relating to pupil size relative to memory scores or gaze duration (Fernández & Parra, 2021). Applying machine-learning methods to chromatic-luminance-induced pupil light reflex dynamics (specifically; pupil response latency) in healthy older adults with and without family history of AD also revealed a capacity for algorithms to classify these two groups with a high level of accuracy (AUROC range = 0.87-0.90), attainable even in the absence of any between-group differences in cognitive function or volumetric MRI scan data (Lustig-Barzelay et al., 2022). This latter finding in particular indicates not only the potential sensitivity of pupil dynamics in classifying dementia subtypes in preclinical stages, but also alludes to capacity for pupil dynamics and machine-learning methods to work in tandem to potentially improve on the insights neuroimaging data can provide when subjected to a similar process.

5.2.3 Differentiating typical from atypical ('visual') Alzheimer's disease: An overview of the current science

Due to great deal of comorbidity with visual problems, which often co-occur in normal physiological ageing - and further still in PLWD - it can be a challenge to differentiate between tAD and PCA in clinic without invoking further exploratory methods (Yong et al., 2023). Furthermore, evidence points to some corticovisual dysfunction even in individuals with tAD at the early stages of the disease, which undoubtedly complicates the differential diagnostic process (Almkvist, 1996). Differences between individuals with tAD and PCA are however observable in neuroimaging studies by virtue of marked atrophy (arising respectively in the hippocampal and posterior-parietal regions); however, other more fine-tune neuroimaging biomarkers have been identified, including: cortical thickness (Lehmann et al., 2011) local tissue magnetic susceptibility properties (Singh et al., 2022), and localised white matter hyperintensities (Pham et

al., 2022). There are relatively few oculomotor studies specifically involving individuals with PCA, perhaps by virtue of presuming the methodology to be counterintuitive to apply to a population with such marked basic and higher visual order deficits. Eye-tracking studies which have, however, focused on between-group dynamics have demonstrated notable differences in fixation stability and profiles, which shows promise for a pupillometry-focused machine-learning classification attempt between PCA and its typical memory-led counterpart.

Limited attempts have been made to investigate machine-learning classification on eye-tracking metrics in individuals with PCA. Pavisic et al. (2017) explored classification performance of discriminating individuals with young-onset Alzheimer's disease (YOAD, onset age of <65 and PCA) from healthy age-matched controls based on oculomotor profiles. Modelling was undertaken using smooth pursuit as the feature of interest; exploratory analyses did not extend to PCA-tAD classification (Pavisic et al., 2017). Between-syndromic-group differences in fixation were however established [for maximum fixation duration; tAD>PCA Pavisic et al., 2017] indicating the low-order visual tasks may elucidate different syndromic group differences with relative ease. This proposal is supported by previous work firstly demonstrating PCA-control differences in fixation profiles, with PCA participants exhibiting lower frequency of fixations within a pre-specified region of interest during a two-dimensional naturalistic scene perception task, in the absence of any between-group differences in overall fixation duration being observed (Shakespeare et al., 2013). A subsequent study including different Alzheimer's disease phenotypes (individuals with PCA, typical Alzheimer's disease) and age-matched controls confirmed between-syndromic-group differences in fixation profiles during a simple fixation task (Shakespeare, Kaski, et al., 2015). While PCA participants' elicited a higher frequency of intrusive saccades and lower periods of fixation relative to age-matched controls, distinct between-patient-group patterns of fixation were also observed by virtue of increased frequency of square-wave jerks; (noting tAD>PCA) in the absence of any significant between-patient-group differences in intrusive saccade frequency or longest period of fixation (Shakespeare, Kaski, et al., 2015). Crucially, this study employed AUROC analyses to interrogate the ability for each selected feature in the study (which also included saccade and smooth pursuit assessment) to discriminate patient groups. The highest area under the ROC curve performance was moderate for saccade metrics (0.80-0.87)

but low for fixation metrics, in some cases just above chance (0.51-0.66). It may be important to note that the fixation task in this study was simply to maintain eye contact with a central fixation cross without blinking. Any paradigm wishing to explore tAD-PCA classification prowess based on low-level visual dynamics (e.g. such as pupillometry profiles during constant fixation) may improve a model's capacity to discriminate syndromic groups by combining this type of task with a simultaneous low-burden cognitive task designed to elicit additional processing. This approach may optimise model performance and improve on previous AUROC findings, owing to: (a) the potential increase in cognitive processing underlying task completion, which may yield more dynamic between-group pupil profiles; and (b) the capacity to undertake a multimodal classification, which has been shown to increase model performance in previous machine-learning investigations (Jang et al., 2021).

5.2.4   Repurposing an existing dataset for machine-learning classification: The Play It Again music pupillometry paradigm

The existing Play It Again paradigm and data acquisition was selected as a candidate dataset with which to test the capacity for machine-learning methods to classify tAD and PCA. It fulfilled the required criteria as outlined above, in that it (a) primarily involved a fixation task whilst invoking additional processing by virtue of making emotional and cognitive appraisals of music; (b) automated metrics aside from pupillometry features had been collected, enabling multimodal classification to take place; and (c) total experimental stimulus presentation time was short (6 minutes) but yielded up to 360,000 within-trial observations per participant due to the high eye-tracker sampling rate. The music paradigm itself, relative to performance tests of fixation and/or oculomotor function, is arguably more engaging as it incorporates an activity that the previous chapter has shown to be widely accessed in this patient population. The notion of no correct/incorrect response for the prompts (in that they prompt a response about a felt emotion or feeling of familiarity) could serve as a less stress-inducing paradigm to engage with relative to standardised neuropsychological test batteries. Exploration of the current paradigm's ability to classify PCA from tAD reflects attempts to improve patient experiences in other clinical testing contexts; for example, moving away from audiometric pure tone detection test towards soundscapes, with higher ecological validity to real-world listening experiences

(Johnson, Marshall, et al., 2021). Music listening is one such example of a low-burden task with which participants may wish to engage, but it also holds the advantage of invoking clinically relevant complex cognitive processes which can be reflected in differential low-level visual profiles as shown in the previous chapter.

5.2.5 The current study

The previous chapter indicated that, at least for individuals with tAD, there is a potential music-pupillometry link between emotional and cognitive appraisals of individual songs, with relative preservation of overall pupil response in line with controls. This was, however, observed in the PCA cohort to a far lesser extent. Initial between-group comparisons of trialwise features (maximum pupil response and baseline size) revealed not only patient-control differences, but differences between the typical and atypical Alzheimer's disease groups. The current chapter seeks to capitalise on these initial reported syndromic group differences in a secondary analysis of the Play It Again data undertaken in collaboration with the UCL Department of Computer Science, as part of the 'AI in Dementia Care UCL-University of Toronto' collaboration. This secondary exploration involved the application of machine-learning (ML) methods, which attempted to classify tAD and PCA using several pupillometry features and automated latency response times to the self-reported familiarity ratings during the Play It Again paradigm. The current chapter also sought to compare the sensitivity and specificity of the model to the AUC metrics generated for each neuropsychological test which accompanied the Play It Again paradigm.

5.2.6 Research questions and hypotheses

Owing to the lack of pupillometry studies undertaken in individuals with PCA to date, and moreover lack of specific between-syndromic-group comparisons in the previous chapters beyond elicited pupil size, it was not appropriate to generate confirmatory hypotheses for these secondary analyses, which were exploratory in nature.

5.3   METHODS

The work outlined in this chapter was a secondary analysis of the existing Play It Again syndromic dataset, therefore all methods (unless outlined here) remain the same as outlined in Chapters 3 and 4 unless specified below.

5.3.1   Ethical approval for transfer of existing data

The existing ethical approval (17/LO/0099 and 8545/002: Created Out of Mind) incorporated permissions for data sharing to the UCL Department of Computer Science for further data processing for research purposes; this was verified and re-confirmed by the UCL/ULCH Joint Research Office before data sharing took place. Subsequent funding to the original funding was secured to enable this additional analysis to take place (University College London-University of Toronto AI in Dementia Care collaboration grant).

5.3.2   Pupillometry: Pre-processing

For this collaboration, pupil preprocessing steps were undertaken by firstly by EB and subsequently by the UCL Department of Computer Science collaborator YZ (see Section 5.6. for individual acknowledgements). All pre-processing steps took place on mean pupil area data collected within the timebins in line with the previous chapter, with few exceptions. For computational purposes, data were downsampled to 200Hz (5ms) timebins rather than the more global 50ms output utilised in Chapter 4. This increase in sampling rate was requested by UCL Department of Computer Science collaborators for increased number of observations, imperative for algorithm training and testing. The preprocessing pipeline was adapted further in that data were smoothed by YZ using a band-passed filter (1Hz low cut-off; 1/28 Hz*)*, and data were z-scored (as observed in Benhamou et al., 2021). An exclusion criterion was set as 25% missing data within trials and within subjects. Applying this subsequently excluded 25.8% overall trials (280 out of 1083 trials) and 7 participants from the dataset. With a stricter exclusion criterion of 50% missing data within trials, 51.9% of trials (563 out of 1083 trials) and ten participants were excluded. This differs from the previous chapter findings owing to the finer-grain dataset used here (200Hz vs. 50Hz timebins). Baselining took place in line with the preceding chapters. The data were summarised at the trialwise level and exported for algorithm development and evaluation.

5.3.3 Neuropsychological data collection

Each patient completed a neuropsychological battery on the same day as taking part in the pupillometry paradigm. The battery included background tests of general cognitive function, verbal recognition memory, word comprehension, naming from verbal description, cognitive estimates, calculation, spelling, gesture production, and auditory-verbal short-term memory. The secondary analyses capitalised on the relatively-underused data collection within the paradigm, which was initially collected to verify expected respective dissociations of performance in visual and mnemonic domains within the PCA and tAD groups. The range of neuropsychological tasks undertaken form the UCL Dementia Research Centre PCA testing battery. Of particular interest were the following tests reflecting different domains of visual processing (basic, visuospatial and visuoperceptual; Table 5.1).

Table 5.1 Selection of neuropsychological tests accompanying the Play It Again data collection

Test	Description
Basic visual processing	
Acuity	Visual acuity test from the Cortical Visual Screening Test (CORVIST) (James, Plant, Warrington, & Thames Valley Test Company, 2001), requiring discrimination of squares, circles and triangles at decreasing stimulus sizes corresponding to Snellen form acuity levels.
Shape detection	Visual Object and Shape Perception (VOSP) subtest examining figure-ground discrimination (Warrington, James, & Thames Valley Test Company, 1991). Stimuli (N = 20) were random black patterns, half with a degraded 'X' superimposed. Patients were requested to state whether an 'X' was present.
Shape discrimination	The stimuli (N = 20) for this boundary detection task, adapted from Efron (1968), where a square (50x50 mm) or an oblong of edge ratio 1:1.2 matched for total flux. The task was to discriminate whether each shape presented was a square or an oblong.
Hue discrimination	CORVIST subtest involving stimuli (N=4) comprising 9 colour patches, 8 of the same hue but varying luminance and one target colour patch of a different hue.

Crowding	Participants were requested to identify a central target letter flanked on each side by two Arabic numerals (N=10).
Visuoperceptual processing	
Fragmented letters	VOSP subtest in which participants were asked to identify letters (N=20) visually degraded by 70%.
Object decision	VOSP subtest with stimuli (N=20) comprising 4 silhouette images, one of a real object (target) plus 3 non-object distractors.
Unusual & usual views	Comprising photographs of real objects (N=20) pictured from non-canonical (unusual) and canonical (usual) perspectives (Warrington, 1996). Unusual views are presented first, with items not identified subsequently re-presented from the usual perspective.
Visuospatial processing	
Dot counting	VOSP subtest comprising arrays (N=10) of 5-9 black dots on white background.
Number location	VOSP subtest in which stimuli (N=10) consist of two squares, the upper square filled with Arabic numerals in different positions, and the lower square with a single black dot. Participants are requested to identify the Arabic numeral whose spatial position corresponds to that of the target dot.
Letter cancellation	Participants were requested to mark as quickly as possible with a pencil the location of 19 targets (letter As) presented among distractors (letters B-E) in a grid on an A4 sheet (Willison & Warrington, 1992).

5.3.4 Music paradigm feature selection

A multimodal dataset comprising pupillometry features and latency metrics was considered following evidence of a previous eye-tracking and speech metrics study revealing the highest discriminative performance was achieved between AD, MCI and healthy controls for the model incorporating both data types (Jang et al., 2021). In addition, previous eye-tracking paradigms explored in PCA have analysed latency metrics to verbal responses in a categorisation task, reporting slower response times

175

in individuals with PCA relative to age-matched controls (Shakespeare et al., 2013). The previous chapter focused only on one feature (trialwise maximum pupil size) as an outcome measure. Other features, such as acceleration and deceleration of response, have been shown to be informative in previous work investigating between-group pupil dynamics in individuals with tAD and controls (Fotiou et al., 2015). For processed pupillometry signal, seven sample point-wise features were measured and summarised at the trialwise level in terms of mean and standard deviation, including: pupil size, gaze fixation in X-axis, gaze fixation in Y-axis, pupil change velocity, pupil change percent, constriction velocity, and dilation velocity and 4 trial-wise features (maximum pupil size ($pupil_{max}$) minimum pupil size ($pupil_{min}$), maximum constriction (*max-con)*, and maximum dilation (*max-dil*). While self-report responses were deemed unlikely to influence model accuracy due to minimal observation of between-group differences in Chapter 4, latency metrics (e.g. reaction time (ms) for verbal responses to the subjective familiarity Likert scale [*RT2*] and memory evocation question [*RT3*]) were incorporated as a previously unexplored feature. In total, 18 pupillometry features and two latency metrics were used for classification.

5.3.5 Inferential statistical analyses

Demographic characteristics were compared between the two syndromic groups using chi-square (or Fisher's Exact Test where expected frequencies were lower than 5) for categorical data, and two-sample *t*-tests (or Wilcoxon rank sum where normality assumptions were violated) to establish any between-group differences in numeric scores.

Group comparisons on neuropsychological tasks were undertaken using linear regression models for each numeric outcome (using the *lm* function in R), and a logistic regression (using the *glmer* function in R) for the binary Ishihara plate task outcome (pass/fail). Both models included *diagnosis* as the main predictor variable, and incorporated *music experience* (which was unbalanced between groups) as a nuisance covariate. Receiver Operating Characteristic (ROC) analyses were conducted for each neuropsychological test using the *pROC* package in R (Robin et al., 2011) to generate areas under the curve (AUCs) metrics from which classification accuracy could be inferred.

5.3.6 Machine-learning approaches

The random forest classifier was run comprising 200 decision trees for final classification (see Box 5.1).

Box 5.1 *Justification for random forest classifier selection*

The random forest classifier was selected as a classical machine learning classification technique which shows robustness to outliers and can effectively model complex relationships between variables, collating a large number of decision trees for final classification (Breiman, 2001). Each decision tree generates a class prediction; the class with the most votes is subsequently the final prediction. The class probability is indicated by the class votes:total votes ratio. The input of random forest classifier is the feature vector and the output is all-class probability. An imputed importance is calculated to determine the contribution of each feature to final classification. Baseline scores were evaluated on a subset of data; a specific feature from the validation set is permuted and the score is evaluated again. The permutation importance is defined to be the difference between the baseline score and metric from permutating the feature column. A larger difference indicates a higher importance to the eventual classification.

5.3.7 Model evaluation

To evaluate the classification performance of the fusion and pupil-only models' abilities to discriminate tAD from PCA, Area Under the Receiver Operating Characteristic Curves (AUROC) were computed to generate sensitivity and specificity metrics. The PCA music trials were randomly sub-sampled, to match the number as music trials obtained from AD participants (which was slightly smaller) and create a balanced dataset. All music trials were split randomly into training (60% of the total dataset) and testing sets (40% of the total dataset); the random forest classifier was trained and evaluated on these respective subsets. For the pupillometry data, non-parametric bootstrap for 200 iterations with each iteration including 80% of test data was used to calculate the 95% confidence interval for the AUROC.

AUC discrimination performance thresholds for all analyses were set as follows, in line with Hosmer, Lemeshow, and Sturdivant (2013): AUC = 0.50 [chance]; $0.50 \leq$ AUC $<$

0.70 [poor]; 0.70 ≤ AUC < 0.80 [acceptable]; 0.80 ≤ AUC < 0.90 [excellent]; 0.90 ≤ AUC [outstanding].

5.4   RESULTS

Significant findings at the level of $p < 0.001$ are denoted by double asterisks; single asterisks denote p-values at the level of $0.001 < p < 0.05$.

5.4.1 Participant characteristics

All relative frequencies and positive responses to the demographics questions posed are reported in Table 5.2. Music experience was revealed as the only covariate in the demographic questions between the patient groups.

Table 5.2 Between-group (AD vs PCA) differences in demographic information

	Between-group p-values (PCA vs. tAD)
No. in group	*39*
Age (years):	0.225 [a]
Gender	0.276 [c]
Handedness	1.000 [b]
UK-born	0.678 [b]
Residing in a different country to birth for > 1 year	0.389 [c]
Educated to at least undergraduate degree	0.252 [c]
Native English speaker	0.349 [b]
English fluency	No test[d]
Visual impairment present	0.258 [b]
Hearing impairment present	0.672 [b]
In a 'good' mood	No test [d]
Reporting a close (blood) relative with a neurological condition	0.191 [c]
Dyslexia (self-reported)	0.672 [b]

	Between-group p-values (PCA vs. tAD)
Self-reporting a neurological condition	No test [d]
No. receiving acetylcholinesterase medication	0.402 [b]
Scored music experience: 0-4: $\tilde{x}$ [IQR]	0.043* [a]

[a] = Wilcoxon rank sum test; [b] = Fisher's Exact Test; [c] = Pearson's Chi-Squared test [d] = Both group categories reported the exact same result (100%)

5.4.2 Model classification performance in discriminating tAD and PCA using selected features

Figure 5.1 shows the classification results of random forest classifier when its inputs include 18 trialwise features and latency metrics. The classifier achieved an AUROC of 0.81, sensitivity of 0.8, and specificity of 0.65. When only the pupil features were used, the classifier achieve an AUROC of 0.78, sensitivity of 0.76, and specificity of 0.62, as shown in Figure 5.2. A sensitivity analysis was performed to verify that results were robust to changes in pre-processing methodological decisions. Figure 5.3 shows the classification results with a data threshold of 0.5, i.e., a stricter data selection criterion requiring more valid sample points in each music trial. The AUROC is the same as that with threshold of 0.75. When the classifier only uses the pupil features as input, the AUROC achieved 0.76 (see Figure 5.4).
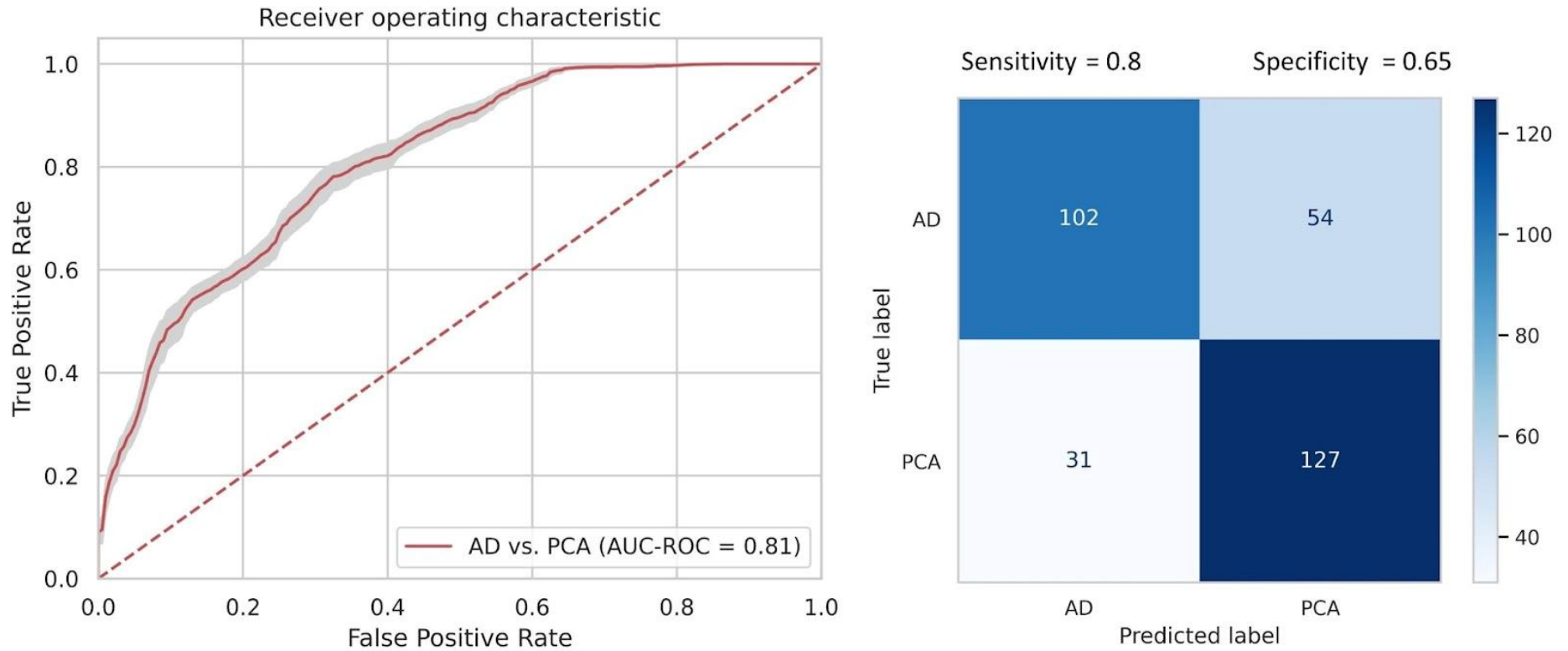
Figure 5.1 Receiver operating characteristics (ROC) curve and confusion matrix of classification between tAD and PCA using the fusion of all trialwise features and latency metrics. The classifier achieved an AUROC of 0.81. 95% confidence intervals were calculated via bootstraps on the test set and are represented by the grey shaded area. The diagonal of confusion matrix indicates the correct classification.
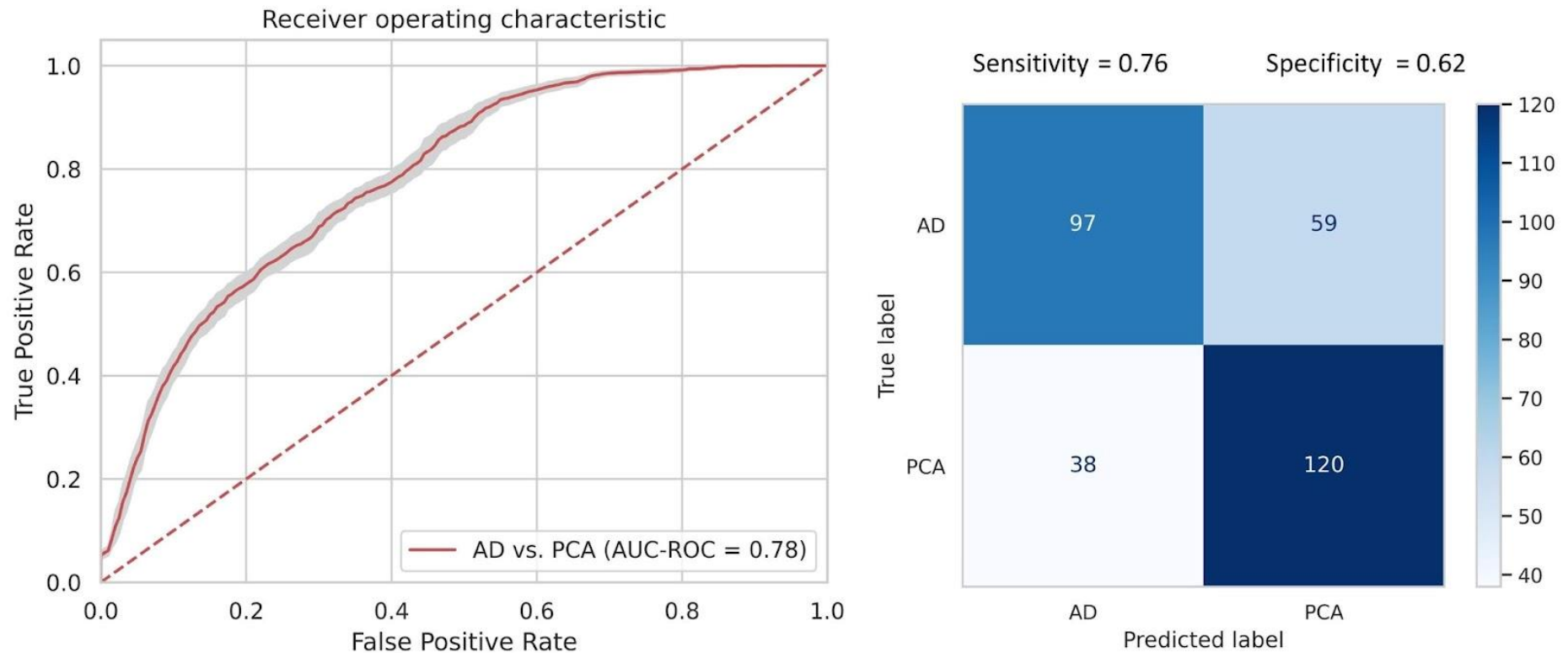
Figure 5.2 ROC curve and confusion matrix of classification results between PCA and tAD using a single modality model (pupil trialwise features only). The classifier achieved an AUROC of 0.78. 95% confidence intervals were calculated via bootstraps on the test set and are represented by the grey shaded area. The diagonal of confusion matrix indicates the correct classification.
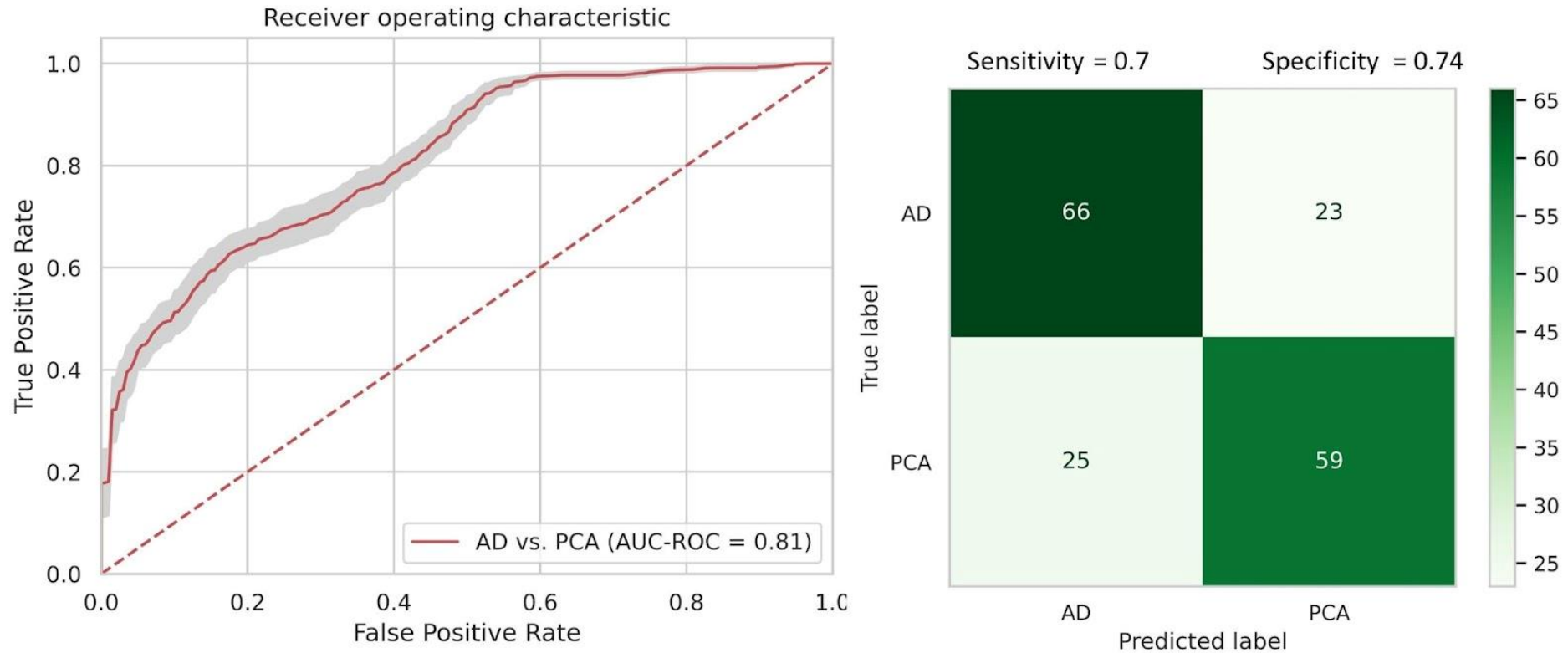
Figure 5.3 ROC curve and confusion matrix of classification between tAD and PCA using the fusion of all trialwise features and latency metrics, and reflecting stricter missing data exclusion thresholds. The classifier achieved AUROC of 0.81. 95% confidence intervals were calculated via bootstraps on the test set and are represented by the grey shaded area. The diagonal of confusion matrix indicates the correct classification.
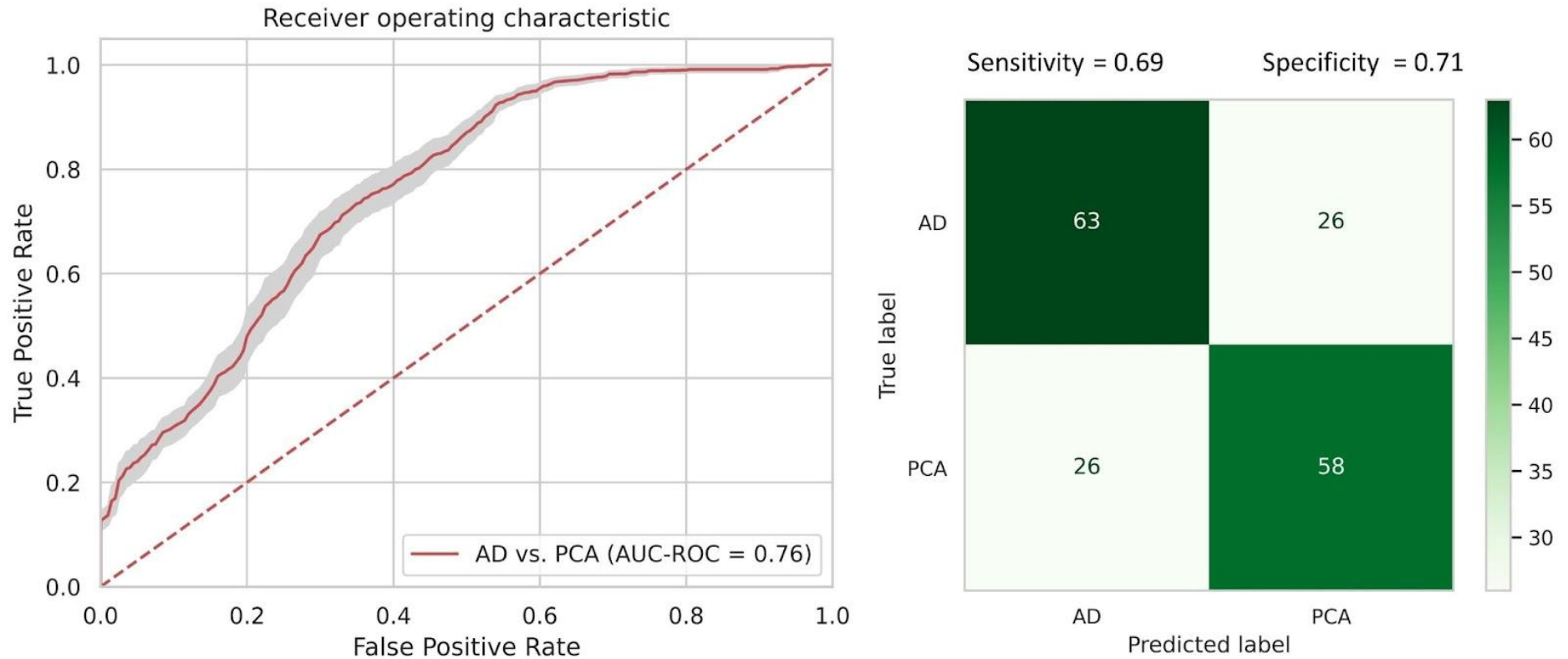
Figure 5.4 ROC curve and confusion matrix of classification between tAD and PCA using a single modality model (pupil trialwise features only), and reflecting stricter missing data exclusion thresholds. The classifier achieved AUROC of 0.76. 95% confidence intervals were calculated via bootstraps on the test set and are represented by the grey shaded area. The diagonal of confusion matrix indicates the correct classification.

5.4.3 Neuropsychological results

Neuropsychological test score means and standard deviations, split by PCA and AD participant groups, are shown alongside significance levels, coefficients and 95% confidence intervals for test differences in Table 5.3. As indicated in Chapter 4, *MMSE* scores were not significantly different between the two patient groups (p = 0.405). On *general background neuropsychological assessment*, the patient groups were broadly comparable (with the exception of memory-led AD participants being significantly more impaired on the Short Recognition Memory test (words) compared to PCA performance (p = 0.013). In line with expected cognitive profiling of the two syndromes, PCA participants were significantly more impaired than the AD group on all basic visual processing tasks (with the exception of the visual acuity task and Ishihara plate task where marginal differences were observed, but not at the threshold of statistical significance [p = 0.058 and p = 0.062 respectively]). The PCA group were also significantly more impaired on all *visuoperceptual* tasks compared to the AD group, and visuospatial tasks (with the exception of the *number location* task in which no differences were observed (p = 0.055).

Table 5.3 Neuropsychological raw scores (mean and standard deviation), *N* missing, regression coefficients and 95% confidence intervals of between-group differences for each test are presented alongside AUROC output. Significant between-group score differences are shaded (p-values, coefficients and confidence intervals in bold).

	Max. score	Mean raw scores ± sd		AD-PCA difference		
		PCA (N = 21)	AD (N = 19)	N missing (PCA:AD)	p-value, coefficient [95% CI]	Area Under the Curve (AUROC)
Background neuropsychology						
MMSE	30	20.9 ± 5.3	19.2 ± 5.2	-	0.405, -1.47 [-5.06, 2.09]	0.577
Short Recognition Memory test (words)	25	19.9 ± 3.8	15.1 ± 5.2	1:2	0.013, -4.25 [7.54, -0.97]*	0.772
Short Recognition Memory test (faces)	25	17.8 ± 4.7	19.4 ± 3.6	2:2	0.216, 1.97 [-1.21, 5.15]	0.600
Comprehension (concrete synonyms)	25	20.9 ± 2.3	19.4 ± 3.1	0:1	0.266, -0.95 [-2.65, 0.75]	0.646
Naming (verbal description)	20	14.8 ± 4.6	12.4 ± 5.7	-	0.277, -1.89 [-5.37, 1.58]	0.612
Cognitive estimates (error score)	30	8.6 ± 7.5	12.1 ± 7.8	3:1	0.196, 3.63 [-1.97, 9.23]	0.631
Calculation (GDA)	24	1.4 ± 2.0	2.6 ± 3.9	2:1	0.250, 1.25 [-0.92. 3.43]	0.616
Spelling (GDST - Set B, first 20 items)	20	9.6 ± 6.0	11.8 ± 5.4	1:0	0.079, 3.28 [-0.41, 6.97]	0.610

	Max. score	Mean raw scores ± sd		N missing (PCA:AD)	AD-PCA difference	Area Under the Curve (AUROC)
		PCA (N = 21)	AD (N = 19)		p-value, coefficient [95% CI]	
Gesture production test	15	14.7 ± 0.8	14.7 ± 0.8	-	0.887, -0.04 [-0.59, 0.52]	0.512
Digit span (max forwards)	8	5.5 ± 1.1	5.7 ± 1.0	-	0.347, 0.34 [-0.39, 1.07]	0.575
Digit span (max backwards)	7	3.1 ± 0.9	3.8 ± 1.3	1:1	0.098, 0.66 [-0.13, 1.46]	0.643
Single word reading (CORVIST)	16	14.7 ± 2.1	15.1 ± 2.7	1:0	0.650, 0.37 [-1.27, 2.02]	0.626
Basic visual processing						
Acuity (CORVIST) Snellen (6/9-6/36)	0.66	0.58 ± 0.2	0.66 ± 0.0	1:0	0.058, 0.08 [-0.01, 0.17]	0.600
Shape detection (figure-ground; VOSP)	20	15.4 ± 2.4	18.1 ± 2.2	3:0	0.006, 2.23 [0.68, 3.79]*	0.816
Shape discrimination	20	13.6 ± 4.7	17.7 ± 3.4	1:0	0.006, 4.42 [1.28, 7.02]*	0.735
Hue discrimination (CORVIST)	4	2.7 ± 1.3	3.7 ± 0.6	-	0.011, 0.95 [0.24, 1.66]*	0.713
Crowding	10	8.7 ± 2.1	9.9 ± 0.2	1:0	0.018, 1.30 [0.24, 2.35]*	0.778
Ishihara plate pass number (pass rate %)	-	12 (41%)	17 (59%)	3:0	0.062, 2.17 [0.21, 5.20]	0.639
Visuoperceptual processing						
Fragmented letters (VOSP)	20	4.8 ± 5.4	15.0 ± 5.5	1:1	<0.001, 9.98 [6.17, 13.79]**	0.891
Object decision (VOSP)	20	9.6 ± 6.2	16.3 ± 2.3	1:2	0.001, 6.39 [2.77, 10.02]*	0.825
Unusual views	20	3.8 ± 3.1	11.3 ± 5.1	4:2	<0.001, 7.45 [4.24, 10.66]**	0.879
Usual views	20	12.5 ± 5.4	18.9 ± 2.4	4:2	<0.001, 6.81 [3.61, 10.01]**	0.870
Visuospatial processing						
Dot counting (VOSP)	10	3.4 ± 3.1	9.2 ± 1.9	0:1	<0.001, 5.68 [3.84, 7.53]**	0.923
Number location (VOSP)	10	3.1 ± 3.3	5.6 ± 3.5	3:2	0.059, 2.411 [-0.10, 4.92]	0.686
A Cancellation: Completion time (s)	90	70.1 ± 24.1	49.0 ± 21.3	0:1	0.008, -22.06 [-37.84, -6.28]*	0.754
A Cancellation: N letters missed	19	5.5 ± 4.3	1.1 ± 2.2	0:1	0.001, -4.29 [-6.72, -1.85]*	0.819

5.4.4 Receiver operating characteristics: Neuropsychological test AUC metrics

AUC metrics for each neuropsychological test have been plotted in Figure 5.5 in order of magnitude, with shading corresponding to performance thresholds as outlined by Hosmer et al. (2013). Dot counting score [visuospatial processing domain]
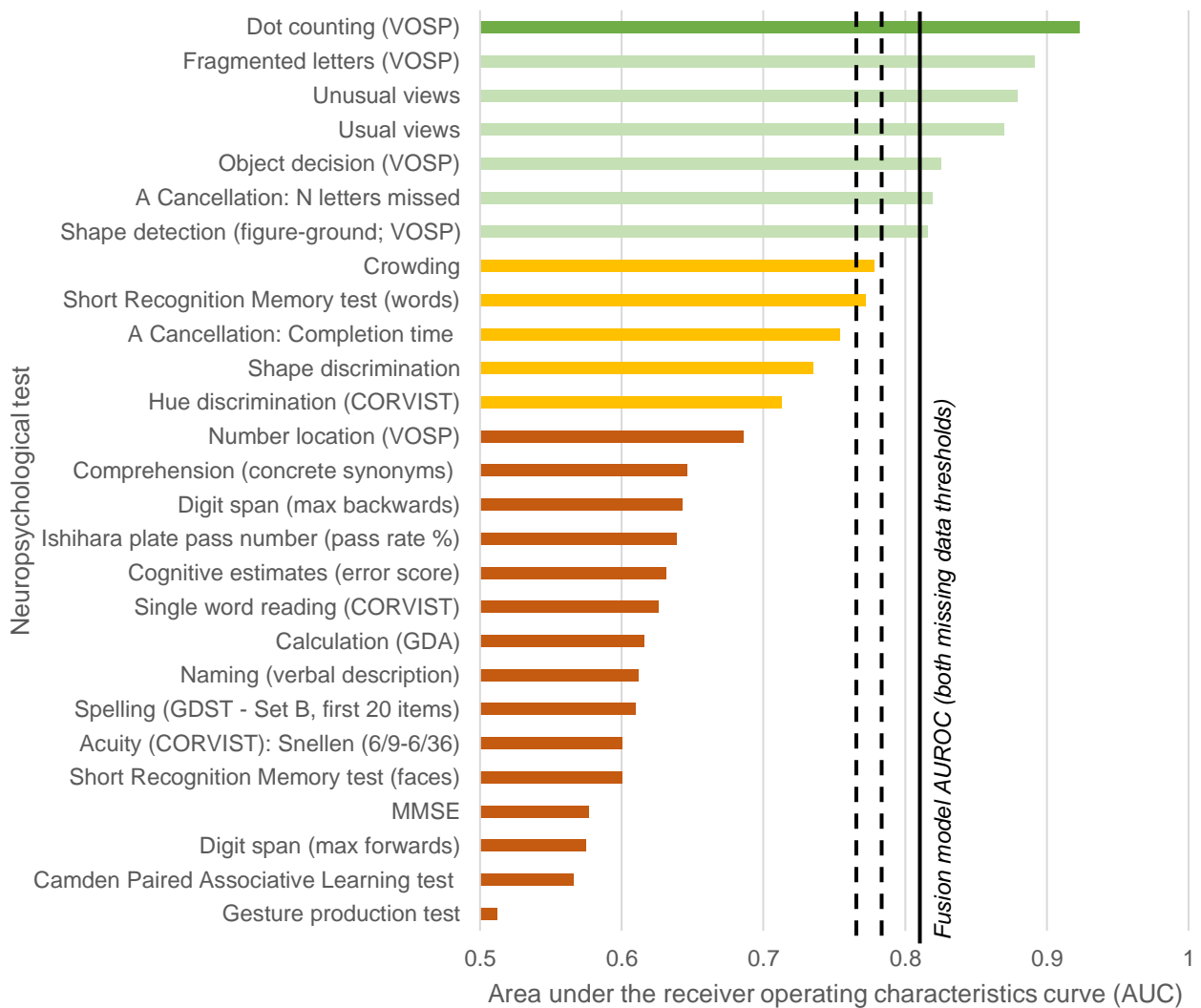
Figure 5.5 Horizontal bar chart depicting each area under the curve (AUC) metric for the neuropsychological battery undertaken by the Play It Again tAD and PCA participants. The minimum x-axis parameter reflects an AUC whereby discrimination is occurring at chance level (0.50). Bars shaded in green indicate tests with either excellent or outstanding discrimination performance [0.80 ≤ AUC]; yellow bars indicate tests with acceptable discrimination performance [0.70 ≤ AUC < 0.80]; orange bars indicate tests with poor discrimination performance [0.50 ≤ AUC < 0.70]. Solid black line indicates the AUROC value for the fusion model at both threshold cut-offs (AUC = 0.81). Dotted lines indicate the pupil-only model AUROC values (0.76 [stricter missing data threshold]; 0.78 [relaxed missing data threshold]).

demonstrated outstanding discriminatory ability [AUC = 0.92]. All tests in the visuoperceptual processing domain demonstrated excellent discriminatory ability (AUC range = 0.825-0.891), as well as one test from the visuospatial processing domain (A cancellation: N letters missed AUC = 0.819), and one basic visual processing test (Shape detection (figure-ground) AUC = 0.816). All other tests in the basic visual, visuospatial and visuoperceptual domains achieved acceptable discrimination performance (AUC range = 0.713-0.778) with the exception of Ishihara plate pass rate (AUC = 0.639; basic visual processing), visual acuity (AUC = 0.600; basic visual processing), and number location (AUC = 0.686; visuospatial processing) which achieved poor between-group discrimination performance. MMSE discrimination performance was just above chance (AUC = 0.577).

5.5 DISCUSSION

The current study findings consisted of the secondary exploration of the Play It Again paradigm. Analysis involved harnessing machine-learning methods to classify two syndromic groups (tAD and PCA) using pupillometry features and response latency metrics acquired within a music listening task. The findings demonstrated that the model performed acceptably in discriminating the two groups in the context of a pupil-only model (as indicated by the AUROC range of 0.76-0.78), and excellently in the context of a multimodal 'fusion' model whereby latency metrics were combined with pupil features in models using data with either a relaxed or strict missing data threshold cut-off (AUROC = both 0.81). To our knowledge, this is the first attempt to discriminate PCA from tAD using pupillometry and response latency metrics using machine-learning methods.

5.5.1 Interpretation of results

The discrimination performance of the multimodal model in classifying PCA and tAD was comparable to - and on occasions surpassed - that of neuropsychological tests which are routinely administered to probe visuospatial and visuoperceptual processing abilities in these two patient groups. The greater discrimination performance of the multimodal model relative to the single modality (pupil) model supports previous study findings whereby superior model discrimination performance was achieved in 'fusion model conditions' (Jang et al., 2021; Qiu et al., 2022). The findings indicate that PCA-

tAD classification is possible with a relatively low-burden paradigm which incorporates music listening, a relatively engaging activity, and a simple categorisation task requiring minimal instructions to complete (**Mengoudi et al., 2020**). The comparability of the fusion model to neuropsychological test performance in their ability to discriminate between the two syndromes highlights that this paradigm may serve as a low-burden efficient way of classifying an individual as having either PCA or tAD, either in complement with or in place of a long battery of neuropsychological testing.

5.5.2 Study strengths

The sample were well-characterised, and balanced in almost all aspects of background information collected which were relevant to the current study. Differences identified in music experience were marginal but were however taken into account in modelling between-group differences in neuropsychological test performance. Given various studies highlighting the cognitive enhancement afforded by music listening in PLWD, and the various benefits of musical experience on cognitive performance in later life (Baird & Samson, 2009) it was informative to incorporate music experience questions in the background data acquisition to inform the discrimination performance of each individual neuropsychological test against which to compare the pupil (single modality and fusion) model performances. In addition, it was encouraging that the PCA group's visuospatial and -perceptual differences relative to individuals with tAD could not be attributed to a global early visual deficit, indicated by retained performance on some basic visual processing tasks (Snellen chart visual acuity and Ishihara plate pass rate).

The consistency of results despite missing data threshold manipulation is reassuring. Both models - one which was more relaxed in its treatment of missing data and one which was more stringent - elicited similar AUROC values for the pupil-only models (0.78 and 0.76 respectively), and the same AUROC value (0.81) for the fusion models. This fact the model results are robust to these changes in methodological decisions indicates the strength of the signal. In addition, despite the pupil-only models eliciting an AUROC which met the criteria to be deemed 'acceptable' in either context (stringent or relatively relaxed missing data models), when examining the AUROC values themselves (fusion vs. single modality) they are relatively comparable (AUROC discrepancy of 0.03 [relatively relaxed context] and 0.05 [stringent] respectively). This indicates that while latency metrics do contribute to a more effective discrimination

188

performance, a large proportion of the contribution is driven by the pupil features. This suggests that further work incorporating more pupil features could in theory provide valuable insight into the classification of PCA and tAD, but at the very least that pupil metrics warrant further investigation due to their ease of acquisition relative to neuroimaging techniques.

Acceptable to excellent discrimination performance was achieved in the fusion model using only behavioural features which were automatically captured (e.g. latency metrics) rather than relying on the responses to the familiarity questions themselves. This was a pleasantly unexpected finding, however there are indications that adding familiarity responses to the model would have done little to improve discrimination performance, owing to the fact that Slattery et al. (2019) and the previous chapter did not observe between-group differences in familiarity ratings.

The findings indicate that this paradigm may serve a dual purpose in advancing our understanding of the clinical populations it seeks to explore. On the one hand, the findings in the previous chapters indicate the music listening task has the capacity to elicit within-group differences in emotional and cognitive appraisal of individual music clips, which may be informative for individualised music use in dementia care and support. On the other, this same task has demonstrated excellent discrimination performance in classifying tAD and PCA diagnoses by virtue of submitting the pupil data and familiarity latency metrics to machine-learning methods. This potential dual purpose seeks to uphold at least one of the 'three e's' medical research strives to achieve: efficiency in particular [the other two being effectiveness and equity (Hinrichs-Krapels & Grant, 2016)]. Exploring the dual-capacity for the Play It Again paradigm has optimised the time and energy participants dedicated to the task; both sets of findings have proven informative for application to two different stages of the disease (diagnosis/early-stage and severe/later-stage).

5.5.3 Study limitations

As discussed in the previous chapter, the current study was also limited by the small sample size (relative to other studies in this field). The higher resolution for the pupil features used to create the pupil features used in this chapter (four times the amount of data points per participant as in the previous chapters) sought to partly reconcile this, however missing data was an issue as demonstrated by the differing threshold

exclusion rates for individual trials and in some cases the exclusion of whole participant datasets based on these criteria. Whilst indeed it is promising that the differing thresholds resulted in similar AUROC values, and care was taken to select an interpolation method which is superior to handling missing data compared with other methods (Mathôt, 2018), the level of interpolation required for both approaches means that the pupil-only and fusion model discrimination performance should be interpreted with caution.

The neuropsychological findings were in keeping with expected results; individuals with PCA exhibited lower scores on tasks tapping visuospatial and -perceptual domains relative to individuals with tAD, whilst retaining normal performance on some basic visual processing. Conversely, individuals with tAD demonstrated impaired performance in the mnemonic task of word recall in the presence of a distractor. However, within-group heterogeneity of neuropsychological performance of visual tasks in PCA points to a "continuum of phenotypical variation" (Lehman et al., 2011, p. 2122). Indeed, examining the standard deviations of the neuropsychological results, some PCA individuals were performing above the threshold for normative results in some visual tasks. Whilst the ROC analyses to some extent overcomes this limitation by virtue of taking each individual into account rather than working off group means, this does point to some intra-group variation. Differences in performance could be associated with condition severity, or syndromic group heterogeneity which was not considered by the current study recruitment design. Crutch et al. (2017) highlight a widespread agreement that PCA presents on occasions with more than the core features of the syndrome as outlined by diagnostic criteria (Mendez, Ghajarania, & Perryman, 2002; Tang-Wai et al., 2004). Individuals with impairments matching only these diagnostic criteria are termed with having 'PCA pure'. However, some individuals exhibit additional symptoms which fulfil criteria for other clinical syndromes (e.g. LBD, lvPPA, or CBS); individuals with this experience are termed as having 'PCA plus' (Crutch et al., 2017).

5.5.4 Future direction

The capacity for music listening is one such example of a real-world clinically relevant complex cognitive process which invokes cortical systems higher than those considered the 'core' auditory regions (Benhamou & Warren, 2020). As observed in

previous chapters, the range of networks music listening invokes makes it a suitable candidate for detecting potential between-group differences. Whilst not entirely a naturalistic procedure, the Play It Again paradigm arguably holds higher ecological validity than the processes involved in undergoing a brain scan. However, with the finding that this engaging task has potentially excellent discrimination performance ability, there is a risk that future development of this work may inadvertently medicalise a pastime which many people with dementia engage with purely for enjoyment. If the capacity to use music in this way is to progress in the future, there would need to be in-depth ethical considerations around this principle of incorporating music listening in any future diagnostic 'toolkit'.

Noting these caveats, the findings described in the current work is suggestive of acceptable-to-excellent performance discrimination of PCA from tAD, by virtue of an algorithm developed to classify these two syndromic groups using several pupillometry features and automated latency metrics acquired during a low-burden music listening task. There is potential merit in expanding this work to address some of the limitations discussed within, notably in an attempt to explore further potential between-syndromic-group differences, and/or by exploring further candidates for features which may be added to future fusion models to increase performance.

5.6 CONTRIBUTIONS & ACKNOWLEDGEMENTS

6   OUR DEMENTIA CHOIR

6.1   ABSTRACT

This chapter describes the final empirical study in the thesis, relating to a real-world opportunity to explore physiological responses in PLWD as part of a British Broadcasting Corporation (BBC) television series, "Our Dementia Choir with Vicky McClure". The opportunity to participate in recreational choral singing for PLWD can give rise to a number of benefits, including improvements in emotional wellbeing. The current study sought to explore this via data collection of wellbeing and stress visual analogue scale (VAS) data pre- and post- six sessions of the 12-week choral programme. Physiological data (electrodermal activity, accelerometer data, skin temperature, and heart rate) were collected pre-, within- and post-sessions using wearable devices (Empatica® E4). Linear regressions indicated in-the-moment within-rehearsal changes in physiological responses; rehearsals were associated with reduced heart rate, elevated skin temperature and elevated electrodermal activity (indexing autonomic arousal) relative to baseline (pre-rehearsal). Autonomic responses declined in the 30 minutes following choir rehearsals, but remained elevated relative to baseline. Increased within-rehearsal EDA responses were found to be associated with positive change on the *wellness* VAS sub-domain score. Within-session EDA responses were not associated with previously-reported potential confounds such as changes in skin temperature and movement, nor disease severity. The findings are discussed in relation to the challenges of real world data collection, which precluded the inclusion of a number of important experimental controls. Nevertheless, the study adds to the evidence base that individuals with dementia respond differentially to participating in music activities such as recreational choir singing on multiple physiological levels. The suggestion of a relationship between physiological responses and wellbeing scale data adds support for the promotion of RCS activities to individuals living with dementia in community-based settings.

6.2   INTRODUCTION

6.2.1   Reported benefits of choirs to PLWD

Music-making, and in particular recreational choral singing (RCS), has been shown to have cognitive, emotional, wellbeing, quality of life and social benefits for people with dementias and their caregivers in a range of studies (Clements-Cortés, 2015; Petrovsky, Cacchione, & George, 2015; Särkämö et al., 2014). The Global Council on Brain Health report specifically recommends individuals "support or start a community-based music making group such as a choir" to improve mental wellbeing [*Music on Our Minds* (Global Council on Brain Health, 2020, p. 4)]. A recent large-scale (N = 318) randomised controlled trial investigating RCS with no RCS with standard care delivered over 6 months to PLWD revealed beneficial effects of RCS on depression rating scale scores at the halfway intervention point, primary endpoint, and 12 month follow-up (Baker et al., 2022). Interestingly, this effect was observed in the absence of any effect of group music therapy on depression scores (GMT) which had been contiguously compared as part of the MIDDEL (Music Interventions for Dementia and Depression in ELderly care) study (Baker et al., 2022; C. Gold et al., 2019). Whilst reflective of previous work in this area (Särkämö et al., 2016b), the extensive sample and statistical power owing to the experimental design (factorial cluster-randomised controlled trial) allows for increased confidence in asserting the potential power of participating in a choir to promote emotional wellbeing in individuals living with a dementia (Särkämö, 2022).

6.2.2   Limitations of choral research in dementia populations to date

The benefits in these studies have typically been captured via behavioural responses to RCS by administering visual analogue scales (VAS) and other measures before and after the sessions have taken place (Strohmaier et al., 2021). Including physiological measures in the study design, which could capture in real time the ever-changing demands, trials and delights when people with a dementia engage in an RCS programme, may enrich the evidence base for the use of these programmes across

community-based and residential care settings (Camic et al., 2018). Capturing in-the-moment responses during a choral session, compared with pre- and post- session parameters when a person is typically continuing to interact with others and receive social stimulation, may elucidate whether there is a 'specific ingredient' that the choral rehearsal itself offers in addition to the social functioning it provides (Clift, Hancox, Staricoff, & Whitmore, 2008). This may have implications for further funding provision, and address scepticism that it is possible to delineate the music activity from the social element of the session; recent attempts to investigate by Bowling et al. (2022) have opted for strict control conditions (speaking lyrics in a group setting/alone) which are arguable in terms of their ecological validity to social interaction opportunities.

6.2.3 Current in-the-moment choral research

Studies have attempted to capture physiological measures to objectively quantify responses of choir members while participating in organised choral programmes. Much of the data collection has been centred on salivary sampling, in an attempt to ascertain the potential immunobiological mechanisms underlying choral participation (Fancourt et al., 2014). The action of singing inevitably engages the lungs, resulting in research concentrating on physiological changes in respiration rate (Lewis et al., 2021; Philip et al., 2020). Choral singing in group and case studies has also been investigated in terms of cardiovascular autonomic responses. This has been indexed by increased within-session HRV responses in healthy populations, which have found to be coupled with respiratory and vocalization motor systems synchronised across choral membership, leading to the proposal of choirs acting as 'superorganisms' (Müller, Delius, & Lindenberger, 2018; Vickhoff et al., 2013). Studies collecting physiological data (HRV and saliva) alongside stress and wellbeing in an RCS group in people with dementia and their carers reported elevated HRV in PLWD within choral sessions relative to pre- and post-session responses (Bourne et al., 2019). Findings comparing EDA, HR, skin temperature and movement responses in PLWD singing the first song of a community singing programme repertoire relative to baseline have however produced mixed results (Walker et al., 2021).

6.2.4   Electrodermal responses (EDA) as a candidate biosignal for choral research

Historically referred to as galvanic skin response (GSR), EDA refers to an overall complex describing electrical phenomena measured on the skin (Boucsein, 2012). These electrical occurrences are usually observable by virtue of the skin becoming conductive to an electrical current, measured by exosomatic recordings of changing sweat secretion levels that occur in the absence of changes in temperature (Critchley, 2002). Changes in skin conductance levels reflect innervation of eccrine glands and diffuse hydration of the sweat duct wall below the surface of the skin, resulting in sweat secretion and re-absorption (Boucsein et al., 2012). Changes in EDA are an ideal candidate to measure psychophysiological arousal compared with other psychophysiological biomarker candidates, owing to the fact that sweat gland activity is mediated by sudomotor fibres linked to the sympathetic nervous system, and is therefore not contaminated by parasympathetic activity (Jabbari, Grimnes, & Martinsen, 2011; Regalia, Resnati, & Tognetti, 2023). Electrodermal activity comprises two recording states: skin conductance level (SCL, reflecting tonic responses) and skin conductance responses [SCR, reflecting phasic responses (Christopoulos, Uy, & Yap, 2016)]. SCLs are more languid in profile, reflecting longer-term differences in autonomic responses relative to task-evoked SCRs, but have nevertheless been explored in dementia-music studies as an indicator of sympathetic processing (Barradas et al., 2021).

In the current work, EDA data are recorded before, during, and after six choral rehearsals as part of a real-world study of dementia choir membership (Chapter 6). Whilst EDA activity remained the focus of the work owing to its relative capacity to indicate sympathetic activation, other physiological data strands were collected (HR, skin temperature and movement) which were explored both in their own right and in terms of their potential mediating effect on EDA.

6.2.5   Rationale for the current work

There are a number of nationally- and internationally-recognised RCS activities already deployed in community settings which are designed for PLWDs and their

caregivers (Alzheimer's Society, n.d.). These community-based choirs have demonstrated the capacity for individuals living with dementia to not only engage with previously-known music, but learn original songs (Bannan & Montgomery-Smith, 2008). A qualitative study exploring this specific set-up (whereby PLWD and carers are intermixed with the choral conductor not being aware of each individual's status) identified themes of increased social connectedness, impact of relationships and memory, lifting the spirits, created a shared experience and opportunity for diagnosis acceptance by virtue of participation (Osman, Tischler, & Schneider, 2016). To date, studies of RCS have focused on this type of choral group composition (I. N. Clark, Tamplin, & Baker, 2018). Mixed-group membership may be driving, or at least contributing to, an effect on increased wellbeing and reduced stress during the choral sessions (Thompson, Baker, Tamplin, & Clark, 2021). In addition, findings from Bourne et al. (2019) indicate there may be differential findings between effects of a choral rehearsal on physiological responses in PLWD and caregivers. HRV was reported to reach Bonferroni-corrected significance levels when comparing within-choral-session HRV with baseline and post-session responses in the PLWD group; this statistical threshold was not reached in the caregiver choir membership (Bourne et al., 2019). There is not, to the author's knowledge, a physiological evaluation undertaken on a choir whose membership exclusively comprises PLWD.

6.2.6 Current work

This study took place in collaboration with filming for a British Broadcasting Corporation documentary about a choir of people living with a diagnosis of a dementia ('Our Dementia Choir with Vicky McClure'), providing a complementary scientific perspective to the narrative of how PLWD respond to music activities (Ranscombe, 2021). While a common criticism of real-world studies is around the challenge of variable control, it is arguable that, without adopting a naturalistic setting, it is difficult to know the impact that being in an overtly 'experimental' environment has on participants' responses. It also addresses the clear lack of ecological validity within laboratory-based studies, although this type of critique has been queried by Holleman, Hooge, Kemner, and Hessels (2020). The study used passive physiological recording of HR, EDA, skin temperature (ST) and motion-based activity (ACC), alongside

participants' subjective self-report ratings of wellbeing and stress, to elucidate a richer understanding of the moment-to-moment changes in response to both the positive and challenging aspects of participating in an RCS programme where people with a dementia participated unaccompanied by their relatives. The physiological data collection included baseline and post-session conditions, both of which comprised social interaction opportunities (against which the within-session measurements were compared). As well as the differences in choral membership comprising only individuals living with a dementia, the current work takes up recommendations from Bourne et al. (2019) by virtue of analysing the multiple physiological data strands acquired via the Empatica® E4 device. In addition, the current study incorporated a longitudinal design, and used statistical modelling techniques which can account for repeated measures over a number of sessions as well as potential individual differences.

6.2.7 Hypotheses

RH1. Post-choral session VAS scores across all wellbeing measures will be higher and post-session stress reports lower compared with pre-session reports, in line with Bourne et al. (2019).

RH2. A difference will be observed in autonomic responses during the choir rehearsal sessions compared with the 30 minutes preceding and following the choir sessions, reflecting HRV findings in Bourne et al. (2019) and findings of PWLD-only differences in within-session autonomic responses which were not observed in the caregiver group.

6.3   METHODS

We employed a quasi-experimental repeated-measures exploratory approach for this investigation, in a naturalistic setting.

6.3.1 Choir members/participants

20 people (11 male, 9 female) with a diagnosis of a dementia volunteered to join the choir. All were subsequently invited by university researchers to take part in the scientific aspect of the documentary filming. Participants were presented with a written information sheet which described the project. All choir members consented to participating in writing. Participants' ages were calculated on the last data collection point over 12 weeks. Owing to the requirements set out within the choral programme filming, background information gathered was selective. Self-reported diagnosis, date of birth, and handedness was established. A Mini Mental State Examination (Telephone Version; MMSE-T) was administered at follow-up 4 months after the last testing session (Newkirk et al., 2004). MMSE-T scores were converted to predicted in-person scores following guidelines in Newkirk et al. (2004), and reflecting findings of comparable performance in patient and healthy older adults on a range of clinical and neuropsychological testing regardless of whether administered in-person or in a remote setting whereby the researcher is not in the same room (Requena-Komuro et al., 2022). According to the converted MMSE scores at follow-up, all participants who were assessed were in the mild-moderate stages of dementia (see Table 6.2). We were unable to administer the MMSE-T to nine participants at follow-up (n=1 due to illness, n=2 had moved to a residential care home and n=6 were lost to follow-up). Neither previous choral singing nor musical experience were pre-requisites to choir membership.

6.3.2 Ethical approval

The study was approved by a Canterbury Christ Church University Research Ethics Committee (V:075\Ethics\2015-18) and adhered to British Psychological Society ethical standards. The study was supported by funding from the Hub Award, Wellcome (200783/Z/16/Z).

6.3.3 Measures

Wellbeing and stress scales

Canterbury Wellbeing Scales (CWS) were used to ascertain subjective wellbeing. These scales were developed specifically for PLWD (early to middle stages) and dementia caregiver population (Camic, 2020). The CWS is a Visual Analogue Scale (VAS)-style questionnaire that measures subjective 'in-the-moment' wellbeing across five subscales using a scale of 0-100 (happy/sad, well/unwell, interested/bored, confident/not confident and optimistic/not optimistic) in addition to a composite overall wellbeing score (0-500). It is easily completeable in an arts/health context (Strohmaier et al., 2021). An additional VAS to measure stress was administered alongside the CWS (relaxed/stressed) where higher scores indicate lower levels of stress. This was incorporated to ascertain potential coupling of scale score with autonomic responses, and scored inversely.

Skin conductance measurements

A change in a cognitive or emotional response to music is likely to elicit a host of reactions (e.g. reduction in a stress response will impact BP, HR, along with changes in plasma stress hormone levels) (Buss, Jaffee, Wadsworth, & Kliewer, 2018). A multistrand approach collected within one device will increase the likelihood of improving our understanding of the physiological and subjective responses of PLWD as they engage with music. The literature review (Chapter 2) highlighted potential ambulatory wearable devices that could be incorporated into the study design. Heart rate collection methods in earlier studies were unspecified, in contrast to subsequent HR data collection which tended to favour ambulatory ECG systems (Kurita et al., 2006; Okada et al., 2009) or finger-pulse oximeters (Corrêa et al., 2020; Maseda et al., 2018). Whilst Okada et al. (2009) specifically commented that their utilised ambulatory ECG system was tolerable in severe clinical populations, other studies have highlighted difficulties in measuring physiological responses in PLWD. In a feasibility study, Hsu et al. (2015) used separate monitors to measure HR/HRV(attached to the chest and wrist) and a 'Q sensor' measuring EDA, TEMP, and ACC data (attached to the wrist). Whilst the physiological findings were not

199

published (warranting the paper being excluded from this review), the authors remarked on the variable uptake from participants in agreeing to have these monitors fitted. Consequentially, the authors recommended that devices measuring physiological responses in research should be - ideally - minimally disruptive to fit. Recent studies seem to be addressing this concern by capitalising on the use of discreet wearables, such as the Empatica® E4 device, within dementia-music research (Bourne et al., 2019; Walker et al., 2021; Sun et al., 2021). The Empatica® E4 is an unobtrusive wristband, which measures not only HR, but BVP (from which HRV can be derived), ACC, EDA, and TEMP data, thus serving a dual-purpose of a). being tolerated in clinical populations, an b). creating opportunities for multistrand physiological data collection, as cited in Bourne et al. (2019), Walker et al., (2021) and Sun et al. (2021). These devices have been used in a number of studies investigating EDA in PLWD with success (Section 2.3). EDA as measured from the wrist (rather than the palmar surface) has been proposed to reflect changes more in thermoregulation than in sympathetic nervous system activation (Boucsein et al., 2012). This potential confounding variable is able to be explored by virtue of the multistrand data collection acquired via the Empatica® E4 device.

6.3.4 Procedure

The choir members were aware of the intention to undertake an empirical study as part of the dementia choir programme. All participants read the participant information sheet, and had the opportunity to ask researchers questions. Consent was sought and obtained from all choir participants in Week 0. Up to 16 participants were selected from the group at random each week to complete the CWS before the RCS sessions (due to the number of devices available). There were minor fluctuations in attendance across the 12-week programme owing to availability and minor illness. This was documented (see Table 6.2); however attendance indicated a good consistency of singers (minimum of 16 present at each rehearsal out of a total of 20 singers involved enabling 16 observations at the physiological level at each data collection timepoint). The Empatica® E4 devices were synchronised to Unix time, and participants were subsequently fitted with one device on their dominant hand on arrival (or post-consent in Week 0) to establish a pre-rehearsal baseline period before the choral session

began. During the baseline period, participants were seated around tables and interacted with each other and relatives. The Empatica® E4s passively recorded HR, TEMP, ACC, and EDA outputs before, during and after the RCS sessions. The research team documented the Unix time of the choir session beginning [*Choir start Unix time*] and finishing [*Choir end Unix time*] to allow for standardised baselining before and after the rehearsal to reconcile differences in arrival and Empatica® fitting times.

No specific instructions were given by the researchers during the RCS activities. The choir was conducted primarily by a highly-experienced choral director with over 20 years of working with singers, and assisted by the Nottinghamshire Singing for the Brain group leader who has extensive experience of working with PLWD in choral settings (O'Neill, 2019). The sessions broadly consisted of a welcome, verbal and physical warm-up, and singing a mixture of familiar melodies and one broadly-unfamiliar song (see Table 6.1). Following each session, the Empatica® devices continued to record physiological responses for at least a further 30 minutes while participants interacted with relatives. The devices were then removed. Participants subsequently completed the CWS for a second time. This method was repeated throughout the RCS programme (a total of six data collection sessions on Weeks 0, 2, 4, 6, 8, and 10).

Table 6.1 Description of choral rehearsal content on the weeks whereby data acquisition occurred

Data collection timepoint [Programme week number]	Song list
Timepoint (T)0	Amazing Grace [giving lyrical prompts]
[*Week 0*]	White Cliffs of Dover;
	Pack Up Your Troubles In Your Old Kit Bag;
	No Matter What;
	Build Me Up Buttercup
	All Things Bright and Beautiful
	Bring Me Sunshine [performed sitting, then standing]
	Close To You
	Oh Danny Boy
	Daisy Daisy
	Lean On Me
	Get Me To The Church On Time
	I'm getting to know you

	Somewhere Over The Rainbow
	There's A Hole In My Pocket
	Pack Up Your Troubles In Your Old Kit Bag
	It's A Long Long Way To Tipperary
Timepoint (T)2 [*Week 2*]	Welcome song (Banuwa)
	Stand By Me
	Lean On Me
	Don't Be So Hard on Yourself [new song]
Timepoint (T)3 [*Week 4*]	Lean On Me
	Stand By Me
Timepoint (T)4 [*Week 6*]	Stand By Me
	In My Life
Timepoint (T)5 [*Week 8*]	In My Life
Timepoint (T)6 [*Week 10*]	Stand By Me
	In My Life

6.3.5 Physiological data pre-processing

Data were split into three timebins (pre-, within- and post-session time periods), determined using the choir start and end Unix time using a bespoke script in R. Baseline (pre-session) data was determined by the mean output of each physiological response, calculated as *Choir start Unix time - 30 minutes*. Post-session was calculated as *Choir end Unix time + 30 minutes.* One participant's data were removed in Session 2 owing to insufficient baseline data collection due to late arrival and fitting time (<5 minutes data collected at baseline). Individuals' within- and post-session data for each rehearsal were baselined to demonstrate proportional change from each individual's pre-session responses, split by rehearsal. All physiological data strands were processed in their raw format (EDA [microsiemens (µS)]; TEMP [Celsius degrees (°C)] HR [average heart rate extracted from blood volume pulse (BVP) signals]); with the exception of accelerometer data, owing to the three values ascertained for x, y, and z-axis movements. For this data strand, the three values (units: 1/64g) were summed to create a measure of absolute activity.

6.3.6 Statistical analyses

Data were analysed using STATA (wellbeing score analyses) and R (physiological data modelling and coupling with wellbeing scores). Data were interrogated for each week and time of CWS (pre-/post-session) to ascertain normality. Reflecting the inconsistency of the scale data in meeting parametric assumptions, Wilcoxon signed-rank tests were undertaken on CWS composite scores and subdomain pre- and post-session scores; p-values are reported alongside Z-scores. The same approach was applied to the Stress scale, which was scored inversely (higher scores related to lower stress). To establish potential coupling with within-rehearsal physiological responses, change in wellbeing and stress scores were calculated by subtracting the pre-session scale scores from post-session scale scores to create seven further variables (*d*[*happy, well, relaxed, interested, optimistic*]; *dCompositeCWS; dStress*). Each of these differences were recoded into seven factor variables '*sign-diff*' comprising three levels: *positive change, no change, negative change.* To avoid prevalence of Type 1 errors owing to multiple statistical testing, post-hoc Bonferroni-corrected significance levels were calculated for testing pre- and post- scale score differences in individual weeks by dividing the threshold alpha of 0.05 by the number of tests conducted (CWS corrected α [0.05/42] = 0.001).

Mixed-effect linear regressions with robust standard errors (using the HC3 method to account for a small sample) were run firstly for each physiological measure (ACC, EDA, TEMP and HR) overall. Time (baseline[pre-session], within- and post-session) was included in each model as a fixed effect; baseline (pre-session) responses were set as the reference level against which initial associations between physiological responses and within-/post-session time periods were interrogated. Where post-session differences relative to baseline were observed, models were re-run with *within-session* as the reference level to establish within- and post-session differences. Models were then run separately for each week using the same approach.

As electrodermal activity was of particular interest, models using within-session EDA responses were run to determine if this was influenced by factors which have been cited in the literature as potentially contaminating EDA signal, such as skin temperature and movement (particularly when using wrist-positioned data acquisition devices, as raised in Boucsein et al. (2012)). Two separate models were run to

determine the potential influence of these two physiological measures on within-session EDA in the current study. Within-session EDA signal was also modelled to investigate potential associations with change in wellbeing scores (domain and composite; *d*[*happy, well, relaxed, interested, optimistic*]; *dCompositeCWS; dStress*) and each *sign-diff*[*happy, well, relaxed, interested, optimistic*]; *sign-diffCompositeCWS; sign-dffStress*) variable.

All models investigating overall physiological responses (i.e. considering data collected across the weeks) included both *session* and *participant* as random effects to account for repeated measurements from the choir members. The separate models run for each week incorporated only *participant* as a single random effect.

Potential associations between physiological responses and disease severity were modelled separately for HR, EDA, temperature and movement using *MMSE score* (where available) as an indicator of disease severity, and random effects of session and participant.

6.4   RESULTS

Significant findings at the level of p < 0.001 are denoted by double asterisks; single asterisks denote p-values at the level of 0.001 < p < 0.05.

6.4.1 Choir membership/participant characteristics

In contrast to the Play It Again paradigm, there was little experimental control over the diagnostic breakdown of the choral membership, owing to the opportunistic sampling approach which was led by the production company. Individuals were recruited by virtue of an advertisement for the choral programme. Individuals who expressed an interest were made explicitly aware from the initial invitation to join that there would be a scientific element to participating in the programme. As perhaps expected, the composition of the choral membership comprised mainly individuals with more prevalent dementias, although some individuals did not have a diagnosis of a particular dementia subtype [Alzheimer's disease (AD; n = 7), frontotemporal dementia (FTD; n = 2), vascular dementia (VD; n=1), mixed-type [usually vascular/Alzheimer's]

dementia (MTD; n=6), Parkinson's disease dementia (PDD; n=1), or dementia of an unspecified type (UD; n=3)].

Basic background and clinical information about the choir members are reported in Table 6.2. The choir membership was relatively balanced in terms of gender; within-group left-handed prevalence (~10%) was representative of general population left-handedness incidence. According to the converted MMSE scores at follow-up, all participants who were assessed were in the mild-moderate stages of dementia (see Table 6.2).

Table 6.2 Background and clinical information of the choir membership

Participant Number	Handedness	Gender	Age at last session (years)	Diagnosis	MMSET score (/26)	MMSE in-person conversion score (/30)	Sessions attended
1	R	M	80.00	UD	NA	-	5
2	R	F	78.20	VD	NA	-	4
3	L	M	71.75	AD	10	14	5
4	R	M	67.33	FTD	23	24	5
5	R	F	70.30	AD	NA	-	5
6	R	M	82.94	AD	11	15	5
7	R	M	52.00	AD	15	18	6
8	R	F	53.00	UD	NA	-	6
9	R	F	81.05	PDD	NA	-	5
10	R	F	82.58	AD	14	17	5
11	R	M	84.08	MDT	17	20	6
12	R	F	75.15	MDT [VD/AD]	8	13	6
13	R	F	70.03	AD	24	25	5
14	R	F	83.19	Mixed vascular/AD	NA	-	4
15	R	M	75.78	Mixed vascular/AD	14	17	5
16	R	M	62.00	Unknown	NA	-	5
17	L	M	84.79	Mixed vascular/AD	NA	-	4
18	L	F	50.00	FTD	NA	-	5
19	R	M	31.46	YOAD	10	14	4
20	R	M	87.60	AD/PDD	13	16	2
Total	L:R [3:17]	M:F [11:9]	$\tilde{x}$ = 75.47 [IQR = 16.67]	-	$\tilde{x}$ = 14 [IQR = 5.5] 7	$\tilde{x}$ = 17 [IQR = 4.5]	$\tilde{x}$ = 5 [IQR = 0.25]

NA = Not Administered; MMSE(T) = Mini Mental State Examination (Telephone).

6.4.2 Subjective wellbeing and stress scores

P-values and derived Z-scores for each CWS domain, stress scale and CWS composite scores are outlined in Table 6.3. Post-session composite wellbeing scores

had a general tendency to be higher than pre-session scores, with these values reaching Bonferroni-corrected significance ($α<0.008$) at Weeks 0 ($p=0.001$), 4 ($p<0.001$), and 8 ($p=0.001$) (see Figure 6.1). Whilst none of the wellbeing scale domain pre- and post- session differences reached Bonferroni-corrected significance, significant post-session increases in *happiness* scores were observed in all but two weeks (Week 2 ($p=0.347$), and Week 10 ($p=0.636$), in the *wellness* domain in Weeks 0, 2, 4, and 6, in post-rehearsal interest scores (Weeks 4 and 6). Confidence was rated significantly higher on the CWS post-rehearsal in half the rehearsals (Weeks 0, 4, and 8), which was a similarly-observed profile for Optimistic scores (with the additional post-rehearsal increase at Week 6). Stress was significantly reduced at Weeks 0, 4, 6, and 8. None of the CWS domains or stress scales revealed statistically significant differences in Week 10.

Table 6.3 Within-week difference in pre- and post-session CWS and stress scale scores. Z-score and p values demonstrated for each measure. Single asterisks indicate where $α<0.05$; double asterisks indicates where findings reached Bonferroni-corrected $α=0.001$. Borderlines indicate comparisons where post-rehearsal mean score was lower than pre-rehearsal mean score.

Pre-post (within subjects, within week) comparisons

	Week 0	Week 2	Week 4	Week 6	Week 8	Week 10
Happy	2.7	0.94	2.81	2.31	3.17	0.47
	0.007*	0.347	0.005*	0.021*	0.002*	0.636
Well	2.51	2.06	2.91	2.27	1.38	0.86
	0.012*	0.04*	0.004*	0.023*	0.166	0.392
Interested	1.1	0.67	2.16	2.34	0.57	0.12
	0.273	0.504	0.031*	0.019*	0.566	0.908
Confident	3.06	0.8	2.49	1.87	2.71	1.37
	0.002*	0.425	0.013*	0.061	0.007*	0.172
Optimistic	2.51	0.12	2.52	2.26	2.2	0.19
	0.012*	0.904	0.012*	0.024*	0.028*	0.851
Composite	3.27	0.09	3.66	2.72	3.26	0.91
	0.001**	0.932	<0.001**	0.007	0.001*	0.363
Relaxed [stress scale]	2.88	0.026	1.99	2.41	3.18	1.74
	0.004*	0.979	0.047*	0.016*	0.002*	0.082
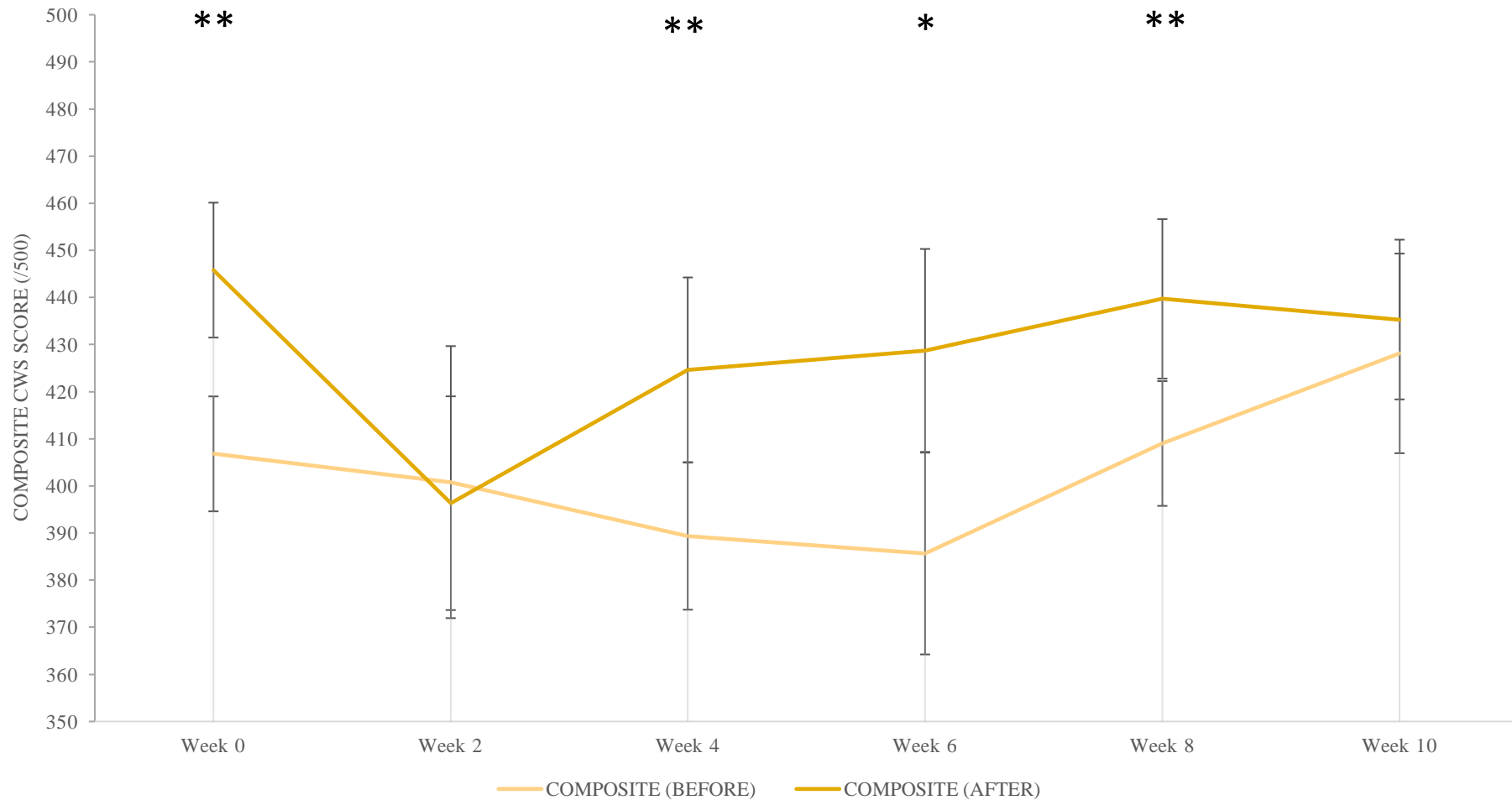
Figure 6.1 Overall composite Canterbury Wellbeing Scale (CWS) scores week-by-week. Double asterisks indicate significant increases in wellbeing following RCS rehearsals at Weeks 0, 4 and 8. Week 6 reached significance at levels of less than α=0.05 (p = 0.011, indicated by a single asterisk). Lowest post-rehearsal wellbeing scores were observed in Week 2, with participants reporting lower levels of interest, confidence, and optimism compared with pre-session scores, however these differences did not reach statistical significance.

Noting the different profile of responses in Week 2, ranksum unmatched sample comparisons were generated between Week 0 vs Week 2, 4, 6, 8, 10, and Week 2 versus Week 4, 6, 8, 10 to establish any differences in wellbeing. These comparisons revealed no significant differences, however Week 0 vs 2 post-rehearsal confidence showed a trend towards a decline (p=0.076).

6.4.3   Physiological responses

Coefficients (with 95% confidence intervals) for each physiological response, split by week, are reported in Table 6.4.

No associations were demonstrated between MMSE scores and EDA responses (-0.046 [-0.177, 0.086], p = 0.491), skin temperature (0.0008 [-0.001, 0.003], p = 0.478) or movement (0.027 [-0.045, 0.100], p = 0.458); associations were however revealed between MMSE score (increases indicating lower disease severity) and lower heart rate levels (-0.004 [-0.008, -0.001], p = 0.011). Heart rate models were therefore run with MMSE score as an additional covariate.

Movement (accelerometer) responses

Looking at responses over the six data collection timepoints, no associations between movement and rehearsal timepoint were demonstrated (relative to baseline: [within-session; -0.031 [-0.847, 0.785] p = 0.940); post-rehearsal -0.329 [-0.749, 0.091], p = 0.124; see Figure 6.2]. Splitting this week-by-week, there was no association between movement and rehearsal timepoint in any week (see Table 6.4 for relevant coefficients).
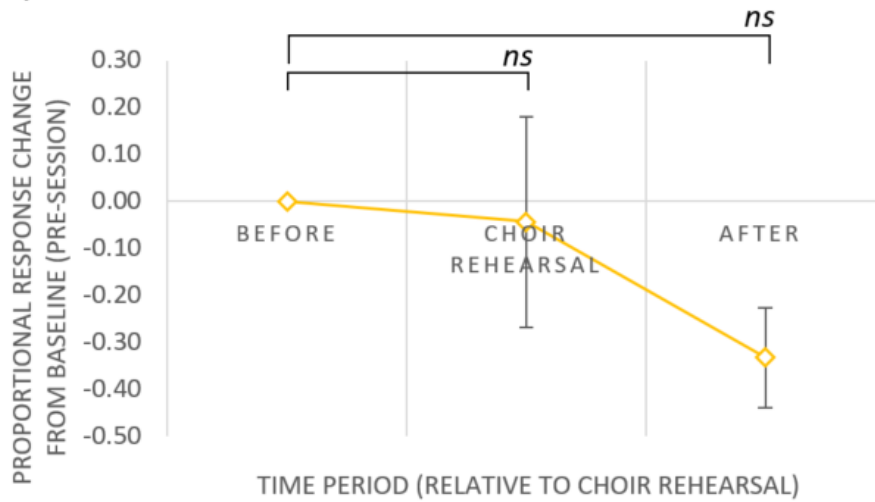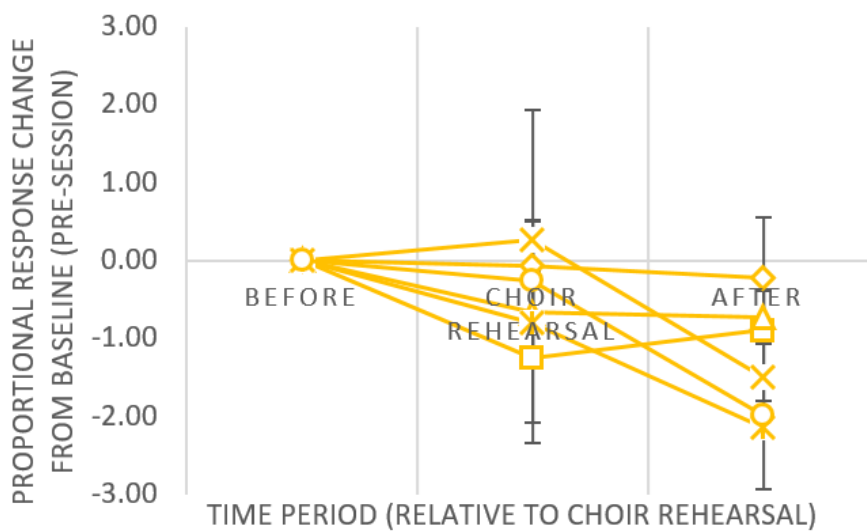
Figure 6.2 Line charts depicting mean proportional change from baseline in movement data before (baseline period), during and after choir rehearsals; (A) overall responses; (B) split by week, with error bars displayed. *ns* in A indicates the comparable within- and post-session movement levels relative to baseline (pre-session) movement levels.

Skin temperature (ST) responses

Looking at responses over the six data collection timepoints, the within-session period was significantly associated with higher skin temperature (0.038 [0.023, 0.053], $p < 0.001$) relative to baseline. Skin temperature remained elevated relative to baseline once the choral rehearsals ended (0.040 [0.024, 0.055], $p < 0.001$; see Figure 6.3). Re-running the model with within-session as the reference level revealed no significant differences between post-session temperature increase from baseline and rehearsal skin temperature increase from baseline (0.002 [-0.020, 0.023] 0.870). Modelling skin temperature week-by-week revealed this profile was reflected in four of the six weeks of data collection points (Weeks 0, 4, 6, and 8); the remaining weeks (2 and 10) demonstrated no associations between rehearsal timepoint and skin temperature (see Table 6.4).

Heart rate (HR) responses

Looking at responses over the six data collection timepoints (and additionally controlling for MMSE), the within-session period was significantly associated with lower HR (-0.058 [-0.086, -0.031], $p < 0.001$) relative to baseline, which returned to a rate comparable to baseline in 30 minutes following rehearsal (0.013 [-0.015, 0.042], $p = 0.357$) (see Figure 6.4). Splitting this week-by-week, all within-session HR was reduced relative to baseline with the exception of Week 10; this reduction in HR reached statistical significance in week 8 only, with marginal findings in Week 0 ($p = 0.051$), Week 4 ($p = 0.056$) (see Table 6.4 for relevant coefficients). HR responses in the 30 minutes following rehearsals were significantly elevated from baseline in Week 10 only; in all other weeks, post-session HR was comparable to pre-session levels.

TEMP

A) AVERAGE CHANGE ACROSS ALL SESSIONS
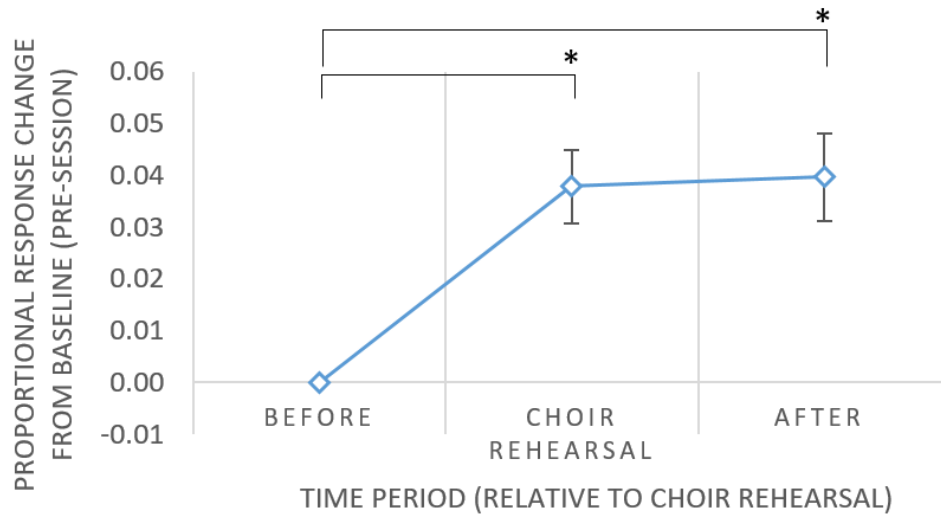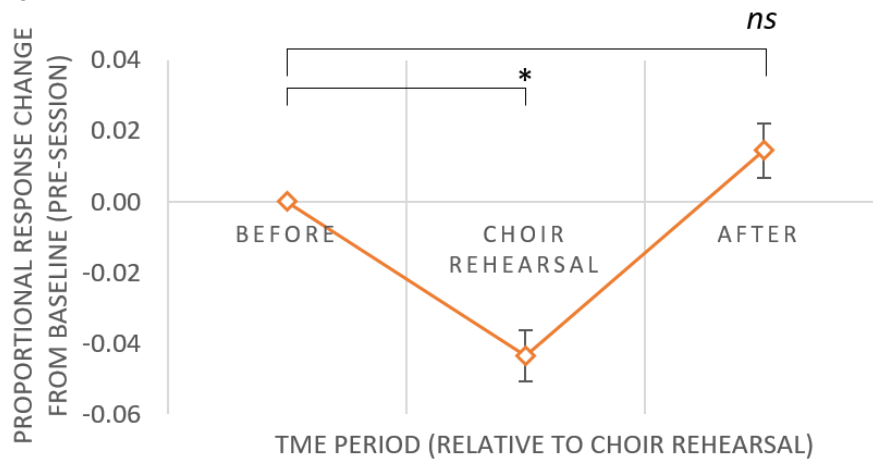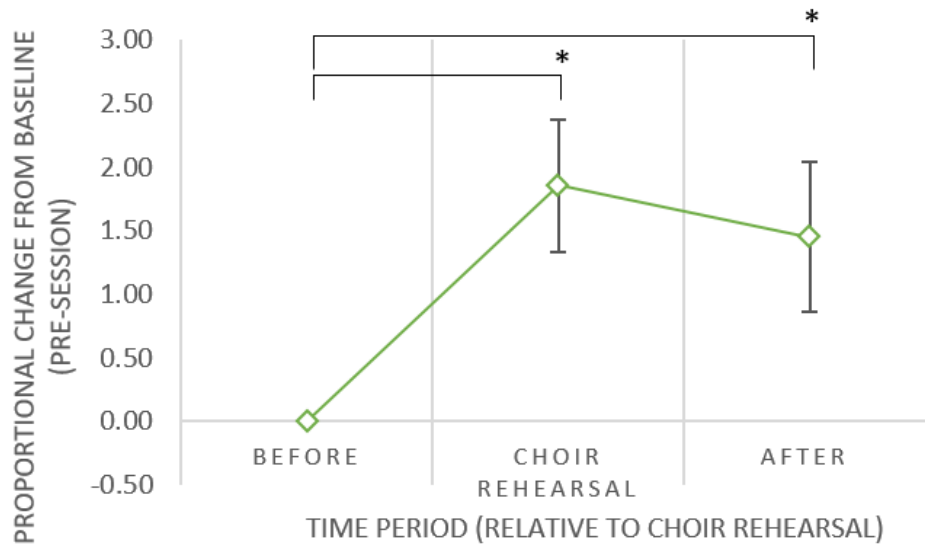
B) SPLIT WEEK-BY-WEEK
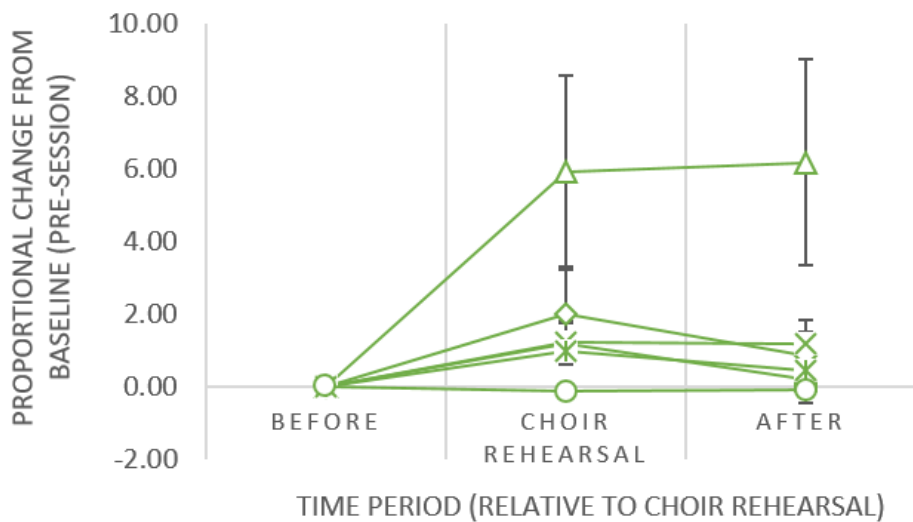
—◇— WEEK 0 —□— WEEK 2 —△— WEEK 4 —✕— WEEK 6 —✳— WEEK 8 —○— WEEK 10

Figure 6.3 Line charts depicting mean proportional change from baseline in skin temperature before (baseline period), during and after choir rehearsals; (A) overall responses; (B) split by week. Asterisks in A represent significantly larger within- or post-session mean skin temperature responses relative to baseline (pre-session) responses.

Figure 6.4 Line charts depicting mean proportional change from baseline in heart rate before (baseline period), during and after choir rehearsals; (A) overall responses; (B) split by week. The asterisk represents the significantly higher within- session mean HR response relative to baseline (pre-session); *ns* in A reflects comparable post-session HR responses relative to baseline.

Electrodermal (EDA) responses

Looking at responses over the six data collection timepoints, the within-session period was significantly associated with higher levels of electrodermal activity (1.858 [0.793, 2.922], p <0.001) relative to baseline (see Figure 6.5). Electrodermal activity continued to be significantly higher post-session than pre-session (baseline) levels (1.463 [0.398, 2.528], p < 0.007); re-running the model with *within-session* as the reference variable

212

level revealed that this post-session EDA response was lower relative to within-session EDA response levels, however this was not statistically significant (-0.395 [-1.901, 1.110], p = 0.606).


Figure 6.5 Line charts depicting mean proportional change from baseline electrodermal activity before (baseline period), during and after choir rehearsals; (A) overall responses; (B) split by week. Asterisks in A represent significantly larger within- or post-session mean EDA responses relative to baseline (pre-session) responses.

Modelling within-session electrodermal activity to interrogate potential confounds of movement and skin temperature revealed no association between EDA and movement (0.078 [-0.052, 0.207], p = 0.238) and a marginal association between EDA and skin temperature, although this did not reach statistical significance (17.854 [-2.484, 38.192], p = 0.085).

Modelling associations between within-session EDA and wellbeing score differences revealed no significant associations, although weak associations were demonstrated between within-session EDA responses and increased differences in *wellness scores* (0.128 [-0.013, 0.269], p = 0.075) and in *CWS composite scores* (0.031 [-0.005, 0.066], p = 0.088). Modelling within-session EDA and *signdiff* wellbeing scores did reveal a significant association between increased within-session EDA responses and positive change in *wellness* CWS subdomain scores (2.788 [0.806, 4.770], p = 0.006).

Modelling electrodermal activity week by week revealed a statistically significant association between increased electrodermal responses and within-session time period (relative to baseline) in all weeks except Weeks 2 and 10. Post-session electrodermal responses were elevated relative to baseline in all these weeks except weeks 6 and 8, whereby electrodermal responses returned to a level comparative to baseline responses (Week 6: 1.188 [-0.131, 2.508], p = 0.076; Week 8: 0.426 [-0.224, 1.075], p = 0.193).

Splitting the physiological responses by week revealed no significant associations between physiological responses and time period relative to baseline in Weeks 2 and 10 in any of the data strands acquired, with the exception of heart rate response in Week 10 which was significantly elevated post-rehearsal 0.063 [0.014, 0.112] p < 0.014; see Table 6.4).

Table 6.4 Within- and post-session coefficients [95% CI] and p-values relative to baseline physiological responses, split by physiological response (ACC, TEMP, HR, EDA) and week.

	Session time	Week 0 Coefficient [95% CI]	p-value	Week 2 Coefficient [95% CI]	p-value	Week 4 Coefficient [95% CI]	p-value	Week 6 Coefficient [95% CI]	p-value	Week 8 Coefficient [95% CI]	p-value	Week 10 Coefficient [95% CI]	p-value
ACC	Within	-0.076 [-0.841, 0.690]	0.843	-1.173 [-3.537, 1.190]	0.322	0.577 [-0.351, 1.505]	0.217	0.941 [-2.517, 4.398]	0.586	-1.066 [-3.745, 1.614]	0.427	0.538 [-0.353, 1.430]	0.230
	Post	-0.222 [-1.206, 0.763]	0.652	-0.675 [-1.759, 0.409]	0.216	0.167 [-0.560, 0.895]	0.646	-0.772 [-2.037, 0.493]	0.225	-0.642 [-2.290, 1.005]	0.436	0.150 [-0.247, 0.546]	0.451
TEMP	Within	0.047 [0.024, 0.070]	<0.001**	0.003 [-0.008, 0.015]	0.571	0.076 [0.014, 0.138]	0.018*	0.045 [0.005, 0.085]	0.027*	0.050 [0.009, 0.091]	0.019*	0.006 [-0.001, 0.012]	0.105
	Post	0.042 [0.016, 0.067]	0.002*	0.005 [-0.012, 0.022]	0.532	0.088 [0.025, 0.150]	0.007*	0.042 [0.003, 0.081]	0.037*	0.062 [0.019, 0.104]	0.005*	-0.0005 [-0.014, 0.013]	0.937
HR	Within	-0.086 [-0.173, <0.001]	0.051	-0.075 [-0.163, 0.014]	0.094	-0.055 [-0.112, 0.002]	0.056	-0.047 [-0.113, 0.018]	0.148	-0.107 [-0.165, -0.050]	<0.001**	0.009 [-0.066, 0.083]	0.811
	Post	-0.018 [-0.070, 0.034]	0.493	0.022 [-0.094, 0.139]	0.685	-0.002 [-0.068, 0.063]	0.947	-0.0006 [-0.090, 0.089]	0.989	-0.016 [-0.069, 0.037]	0.532	0.063 [0.014, 0.112]	0.014*
EDA	Within	1.984 [0.309, 3.658]	0.021*	1.181 [-1.089, 3.451]	0.300	5.890 [0.323, 11.457]	0.039*	1.220 [0.084, 2.356]	0.036*	0.968 [0.166, 1.771]	0.019	-0.139 [-0.378, 0.101]	0.249
	Post	0.850 [0.033, 1.667]	0.042*	0.174 [-0.830, 1.177]	0.729	6.157 [0.256, 12.057]	0.041*	1.188 [-0.131, 2.508]	0.076	0.426 [-0.224, 1.075]	0.193	-0.098 [-0.482, 0.285]	0.608

ACC = movement; TEMP = skin temperature; HR = heart rate; EDA = electrodermal activity

6.5   DISCUSSION

The current study measured in-the-moment physiological responses elicited whilst individuals with a diagnosis of dementia participated in a choir. The findings suggest that overall within-session autonomic responses can be differentiated from physiological baseline (pre-session responses) and in some cases demonstrate a return to baseline-comparable physiological responses within 30 minutes of the rehearsal ending. An increase in composite wellbeing scores post-rehearsal relative to pre-rehearsal scores was observed for the majority of data collection timepoints. A coupling between electrodermal activity and positive change in *wellness* scores was demonstrated. To the author's knowledge, this is the first description of physiological responses of PLWD in an RCS programme without relatives accompanying them throughout the songs.

6.5.1   Interpreting the results

RH1. Post-choral session VAS scores across all wellbeing measures will be higher and post-session stress reports lower compared with pre-session reports, in line with Bourne et al. (2019).

RH1 can be partially accepted. Significant increases in self-reported wellbeing and a reduction in stress following four out of the six rehearsals (half of the overall weeks reaching stricter significance thresholds to account for multiple testing) indicated a partial acceptance of RH1; a positive impact of RCS sessions on wellbeing and stress reduction. These positive findings are in line with initial predictions and replicate results from previous studies (Bourne et al., 2019) and reports of group singing specifically benefitting PLWD through its effects on individuals' mood and enjoyment (Dingle et al., 2021; Moss, Lynch, & O'Donoghue, 2018).

Understanding the difference in context across the Our Dementia Choir programme sessions may explain the unexpected findings in wellbeing scales at Weeks 2 and 10. For example, in Week 2, a rhythmically complex and unfamiliar song was introduced to the repertoire, which may account for the non-significant differences in pre- and

post-rehearsal scores of relaxation, interest, confidence, optimism and composite wellbeing scores. Previous studies have demonstrated that learning new music whilst living with Alzheimer's disease is possible (see case studies; [*Patient GW* (W. W. Beatty et al., 1988); *Patient ML* to an extent (W. W. Beatty et al., 1999)]; also Baird, Umbach, and Thompson (2017)). When challenging music (e.g. singing in rounds) is introduced to groups sensitively and slowly, potential negative effects are eliminated (Camic, Williams, & Meeten, 2013). This indicates that offering challenging music too quickly or novel pieces too close together may have a transient negative effect on wellbeing. Protocols for designing RCS programmes for PLWD outline helpful guidance on navigating these potentially challenging aspects (McDermott, 2018). The final week of the Our Dementia Choir data collection did not elicit post-rehearsal differences in wellbeing scores, and only minimal physiological changes within-session relative to the other rehearsal data collection weeks. This finding may reflect a degree of apprehension amongst the choir as they approached a significant end-of-programme performance to an audience of more than 2000 people at a major arts venue, or to the coming to a close of the regular weekly rehearsals. These possibilities highlight a broader issue of music activity provision in dementia care in terms of the need for it to be a sustainable offer, unlikely to be revoked. (It was subsequently confirmed that the choir would continue, and it has been maintained ever since, to the great credit of all involved: https://www.ourdementiachoir.com/).

RH2. A difference will be observed in autonomic arousal responses during the choir rehearsal sessions compared with the 30 minutes preceding and following the choir sessions.

RH2 can be partially accepted; autonomic arousal responses (as indexed by EDA) were elevated within choral rehearsals relative to pre-session baseline. While some studies did not find any effects of EDA (Vickhoff et al., 2013), these current findings reflect earlier-discovered patterns of increased EDA response in healthy adults participating in RCS (Valentine & Evans, 2001). It is not possible for any single measure to accurately capture the multiple complexities of participatory RCS involving people with a dementia. The advantage of the current study is the use of multiple measures, which improves confidence in interpreting individual responses (e.g. a

relationship between EDA and subjective wellbeing). Multistrand physiological data analyses revealed different response profiles within each modality acquired via the Empatica® E4 device. For example, skin temperature demonstrated a similar within-session profile to EDA. Within-session heart rate showed an inverse pattern relative to baseline responses, indicated by a lower heart rate during choir rehearsals. This differs from earlier work outlined in the literature review (Chapter 2) which reported increases in five out of nine individuals living with dementia recorded whilst engaging in the first song of a community music session relative to baseline (Walker et al., 2021). The contrasting finding here may reflect the longer time period investigated in the current study. No significant differences in movement were observed over the course of the rehearsal relative to baseline.

6.5.2   Study strengths

This is, to the author's knowledge, the first study to report EDA responses in PLWD as they participated in a longitudinal choral programme. The presence of findings differs from previous studies in healthy populations in a choral setting whereby no EDA changes were observed despite reported effects on HRV (Vickhoff et al., 2013). This may relate to the inter-group differences in physiological responses observed between caregiver choir members and PLWD in Bourne et al. (2019); individuals with dementia may get more enjoyment or engage more intently with a choral programme than individuals without the condition, thus leading to observable changes. However, the absence of a control singing group renders this interpretation as speculation only.

Boucsein (2012) cautions against electrodermal activity being recorded from sites such as the wrist, citing that responses acquired there are more likely to reflect thermoregulatory processes rather than psychophysiologically relevant electrodermal phenomena. Indeed, in the current study EDA and skin temperature elicited similar profiles of within- and post-rehearsal overall responses relative to baseline. However, the fact that within-session EDA were demonstrated as not being associated with either skin temperature or movement (plus associations with subjective wellness) is reassuring, as it indicates a likelihood of processes indexing the intended changes of autonomic arousal (i.e. reflecting a cognitive and/or emotional change in state), rather

218

than thermoregulation. Whilst overall EDA responses indicated EDA levels remained elevated relative to baseline even in the 30 minutes following the choir rehearsal, the coefficients indicate that there was a decline in EDA post-session relative to within-session. Although there may be a slightly longer-term effect of a choral rehearsal on EDA that extends beyond the choir rehearsal time itself, it appears that this starts to decline within as short a time as 30 minutes after the experience finishes, thus supporting the evidence for choral participation having relatively transient effects at the physiological level.

Subjective wellness was a significant positive predictor of the increase in electrodermal activity - an objective measure of autonomic arousal, over which the participants have no control. These significant changes in EDA held true even when accounting for individual differences and choir session number. To our knowledge, this is also the first time a link between PLWDs' subjectively rated wellness and changes in EDA response has been demonstrated. The context of the wellbeing scale also provides reassurance that the increase in EDA response is likely to be reflecting an enjoyable experience, rather than indexing the 'fight or flight' initiation sequence when reacting to alarming stimuli which can also lead to increased EDA (Christopoulos et al., 2016). Despite the absence of further statistically significant coupling of autonomic and CWS scale changes, it is striking that the week-by-week observations revealed that across all measures of physiological change, only one (HR) occured within-session during Weeks 2 and 10 relative to baseline; rehearsals in these weeks did not elicit changes in CWS scores post-rehearsal relative to baseline wellbeing and stress ratings. Nonetheless, we must remain cautious when using a handful of physiological measures to infer changes in complex psychological constructs within naturalistic settings (Harding, Sullivan, & Crutch, 2017).

The sample was well-balanced in terms of gender, although in terms of generalisability we need to consider the possibility of bias introduced by self-selection of individuals who volunteer to take part in RCS-related activities (Daykin et al., 2018). The overall sample size was similar as recruited to Bourne et al. (2019) but by virtue of the choir being exclusively PLWD, the current study had a larger sample of dementia participants (N = 20 cf. Bourne N = 10). The modelling approaches selected allowed

for consideration of intersessional and interpersonal variation in each analyses, which are often overlooked but remain important to consider, particularly in older or clinical populations that demonstrate higher levels of physiological response variability relative to younger control participants (Zhao et al., 2019).

While the variations in wellbeing and stress across the current programme were only partially in line with our initial predictions of a consistent post-rehearsal elevation in scores, the differential responses in Weeks 2 and 10 increase our confidence that the CWS and stress scale reports are an accurate reflection of the context and members' experience of each session. Although the primary purpose of the study was to examine associations between subjective and objective indices of wellbeing within singers with dementia, future studies involving separate or co-singing healthy participants can examine the relative magnitude of the impact of rehearsals upon these psychological and physiological metrics. Furthermore, the lack of significant differences in wellbeing ratings across weeks, taken together with the coupling of within-session EDA and wellness domain positive change, seems to support the primacy of in-the-moment over longer term benefits upon wellbeing for this particular choir group.

Innes et al. (2021) highlights the importance of a care comparator in studies investigating the impact of music-based activities. One common criticism of assessing the impact of participatory creative activities in PLWD is that increases in wellbeing may in fact be driven by the nature of participatory activities (e.g. increased opportunity for social interaction and leaving the home environment) rather than the creative aspect itself. While we recognise our experimental design did not include a canonical control condition to explore this explicitly, our baseline and post-session intervals recorded physiological responses while participants engaged in the social, non-musical aspects of their community choir, interacting with friends, family and choir staff. The fact that levels of arousal were higher during the choral rehearsals compared with these meaningful, non-neutral baseline periods, adds to confidence that the observed within-session differences were being driven by the music and the choir.

Acknowledging that MMSE score was unattainable for nearly half the choir membership, it is encouraging that available MMSE scores were not associated with

three of the four physiological responses acquired in the study; furthermore, when controlling for MMSE score, heart rate effects (which had been associated with disease severity) were retained reflecting a decrease in HR within-session relative to the baseline. We therefore recognise the potential of physiological responses acquired using discreet wearable devices to demonstrate how people living with a more advanced stage of dementia who are nonverbal may be responding to an RCS programme. Further replication studies are however needed in populations with early-to-mid stages of dementia before strong evidence-based claims can be made for the overall value of RCS for wellbeing in PLWD. Future studies (e.g. the Pilot Randomised Evaluation of Singing in Dementia (PRESIDE) study) which have currently been protocolised to include individuals who have capacity to consent (e.g. mild-moderate stages) will no doubt further our understanding of the role choral programmes play for PLWD (Dowson et al., 2021). Exploration is however needed in individuals with severe dementia in order to establish how or if music-making activities may be provided considering varying factors relating to that disease stage (Eisenmann, Golla, Schmidt, Voltz, & Perrar, 2020).

6.5.3 Study limitations

The main evident study limitation relates to the lowered capacity for standardising choral rehearsals by way of testing an 'intervention', owing to the real-world setting of the overall investigation. Given the understandable priority given to filming, it was not possible to gather the level of information typically collected as part of empirical investigations of this nature, nor to avoid within-session changes, which may have unduly influenced physiological responses (e.g. dividing the choir into groups for specific harmonising rehearsals). Furthermore, whilst descriptions of the choral rehearsal content were provided to address previous criticisms in the literature of vague procedural outlines, reviewing the heterogeneity of between- and within-session content indicates that it is highly unlikely that precise replicability is possible (spontaneous interactions and events being a natural and essential part of such authentic, non-formulaic musical experiences). The latter limitation is not specific to this study *per se*, but speaks more generally to limitations regarding any evaluation involving multicomponent musical activities which incorporates a real-world element.

Specific efforts were made in previous studies to recruit from a well-established choral group, motivated by potential confounding variables such as uncertain expectations, unfamiliarity with other choir members and facilitators (Bourne et al., 2019). Owing to the real-word opportunistic nature of the research aligning with a broadcast depicting the creation of a 'dementia choir', this was impossible to avoid in the current study. In spite of these limitations, differential findings were reported within-session relative to baseline periods in three of the four physiological parameters measured, and post-session differences were observed in the wellbeing scales - these findings were observed to an extent even in Week 0, which was the first time the choir members had met and sung together. Additional confounds may also have been unintentionally introduced by virtue of the broadcasting element which are not typically features in a choral singing experience (e.g. camera equipment and crew, filming the bulk of the footage within the choir rehearsals). This within-session addition may have contributed towards a high intensity within-session environment, which may have influenced physiological responses rather than the choral activity itself. The current findings must therefore be considered in light of this particularly unusual context. Replication of the current paradigm in a well-established choir unaccompanied by filming would strengthen the current findings and address any concerns around baseline results varying by virtue of a novel experience, or within-session results arising due to filming.

Despite arguable ecological validity of the 'control' conditions in this real-world investigation revealing within-rehearsal elevation of autonomic arousal, teasing apart the underlying mechanisms of what is driving this change is a challenge. In fact, recent publications have stated that due to the link between choral singing and interpersonal action being so great, choral singing is proposed as being a biomarker candidate for social interaction (Delius & Müller, 2022). Warran, Burton, and Fancourt (2022) have recently proposed a framework (INgredients iN ArTs in hEalth [INNATE]) to determine the active ingredients which may be responsible for arts activities innervating health and wellbeing benefits. However, it may be that establishing these active ingredients within real-world multicomponent activities, such as a choir, may be beyond reach when conducting autonomic response investigations, which can be triggered in reality by a multitude of the elements involved in this type of activity. The challenge of

establishing whether music or social interaction (or both) leads to an elevated psychological and/or autonomic response is arguable from a neuroanatomical point of view. This is by virtue of an overlap in neurobiological circuits underlying music perception and social cognition; a recent meta-analysis has additionally revealed both are compromised in dementias such as FTD (van't Hooft et al., 2021).

Given the proposed link between EDA and autonomic arousal (whether that indexes engagement or exasperation; see Harding, Sullivan & Crutch, 2017) it was reasonable to expect that perhaps an autonomic-wellbeing scale coupling may have occurred for either the 'interested-bored' CWS subdomain, or the 'relaxed-stressed' subscale, as a proxy for self-reported autonomic arousal. The fact it was only the 'wellness' subdomain positive change which demonstrated a coupling with EDA responses was unexpected, perhaps reflecting the difficulty of pinpointing the exact feeling that a heightened EDA response elicits. One limitation was that the HR data acquired was utilised in the current context as a complementary physiological measurement to EDA, rather than investigated in its own capacity to be transformed into HRV data, which can index sympathetic activity. This physiological measure also has the advantage of being able to be assessed non-invasively in people living even with advanced stages of dementia with relative confidence of testing normal autonomic function (Femminella et al., 2014), with findings of relative comparability to health controls when assessing HRV reactions to music in individuals with TBI and/or in a vegetative state (Riganello, Candelieri, Quintieri, Conforti, & Dolce, 2010). The extension of HR data acquisition to investigate HRV therefore warrants further investigation in future analyses which may reveal tighter coupling with CWS/stress responses relative to the EDA-wellbeing scale links.

6.5.4 Future direction

Despite the current findings, there are queries around the capability of wearables such as Empatica® E4 to acquire autonomic data that correlates sufficiently with gold-standard autonomic data acquisition techniques (Menghini et al., 2019). This, coupled with the challenges of the current real-world study, indicates that an experimental paradigm with tighter controls and data acquisition techniques may be needed. One

such avenue which may enable this type of research to be conducted in a group singing setting is investigating potential collaborations with research groups who already lead laboratories which may satisfy, to an extent, both the real-world elements with sensitive data acquisition techniques (e.g. the Person Environment Activity Research Laboratory [PEARL], University College London). Future research collaborating with engineers and computer scientists based at PEARL or other similar research environments may enable further investigations of choral programmes on PLWD, enabling both the necessary data acquisition to form more confident conclusions about physiological response while these rehearsals take place, plus enabling additional 'sham' conditions to ascertain any extraneous effects of wearing sensors on self-report scales whilst participating in such a group, and/or participating in music activities within a laboratory setting.

In addition, the current bespoke script developed as part of the pre-processing pipeline could in principle be used to differentiate between different parts of choral rehearsals. There is evidence for example that choir singing and solo singing can elicit differential responses (Schladt et al., 2017). Being able to separate solo singers from the rest of the group through documenting different timings within sessions, and relating those timings to data segmentation, could open up the potential to explore particular components within music-making interventions, thus avoiding the oft-cited limitation of uncertainty around what components of the choral rehearsal may be driving observed changes. This is currently being tested in an existing audiovisual dataset acquired from a Royal Academy of Music (RAM) online programme for people living with a range of rare dementias. The script enables automated segmentation of different parts of the programme (e.g., singing, discussion, listening to professional RAM musicians perform) to enable exploratory analyses of continuous automated facial emotional data captured from the video of participants over six sessions.

The findings suggest that measuring skin conductance may help us understand how PLWD are responding to RCS programmes. These within-rehearsal findings were not accounted for by change in either movement or skin temperature, thus increasing the likelihood of within-session changes reflecting autonomic arousal changes elicited by the choir rehearsals as they were taking place. This interpretation was extended

further still following the association between increased EDA and positive change on a wellness scale. The lack of association between MMSE score and EDA, skin temperature and movement data (and observed effects when controlled for in HR responses) indicates that these devices may provide insight into changes in cognitive and/or emotional state in response to other music activities whereby participants are living with more severe disease stages. The extension of the dementia choir format to other countries by virtue of the success of the BBC broadcast (e.g. in Germany, the Netherlands and Norway) may serve as a potential collaboration for future work in this area ("Dementia Choir format seals international deals," n.d.).

6.6   CONTRIBUTIONS AND ACKNOWLEDGEMENTS

7   GENERAL DISCUSSION

7.1   General summary

Overall, the thesis rationale related to the coupling of physiological responses related to autonomic processing (pupillometry and electrodermal activity) with self-report in the context of music and dementia, to establish whether there is mileage in the capacity for biosignals to be called upon as way of approximating a self-reported 'felt' musical experience for PLWD. Exploring data capture of physiological responses to different music formats (music listening and recreational choral singing) broadens our understanding of whether biosignals have the potential to act as a proxy for subjective musical experiences irrespective of level of engagement (e.g. passive versus active). The motivation for undertaking the work centred on establishing evidence for physiological-phenomenological links to music activities that continue to be observed in PLWD. This evidence could inform the development of future music interventions that incorporate biosignal/feedback systems to enable music provision that can be attuned in real-time to either match or modulate an autonomic response in PLWD who may not be able to communicate their musical experience via verbal self-report.

More specifically, the work described demonstrated a coupling of a variety of physiological responses with (a) emotional and cognitive appraisals of song excerpts in older adults and individuals with typical Alzheimer's disease (pupillometry; Play It Again) and (b) scales of wellbeing (EDA, HR and skin temperature; Our Dementia Choir). Play It Again findings highlighted the importance of including different syndromic groups in a music paradigm. Firstly, this enabled investigation of syndromic differences in psychophysiological profiles and self-report behaviours to create disease-specific considerations for incorporating music in dementia care and support. Secondly, this design enabled secondary analyses to establish the extent to which low-burden paradigms can serve a dual purpose as a syndromic group classification tool. Our Dementia Choir provided the opportunity to explore physiological data capture in a real-world community music setting, with findings that add to the evidence base for the importance of musical-based activity for PLWD, reflected not only in wellbeing scales but additionally by changes on multiple within-session physiological levels.

7.2   Further consideration of study results

7.2.1 Confounding variables and potential impact on psychophysiological responses

A limitation plaguing many dementia-music studies is the limited sample size; the work described in the thesis is unfortunately prone to the same critique. The sample size of the PLWD participant groups reflects in this case either: the challenges involved in recruiting individuals with a relatively rare form of dementia (Play It Again); or availability of devices to undertake simultaneous data collection at group-level (Our Dementia Choir). Nevertheless, both studies indicated promising results, which justify further investigation at scale. Larger sampling may also elucidate the potential contributions of other participant-led factors to physiological and behavioural responses observed in the current study (Särkämö et al., 2016a). Researchers have cited that, for example, interindividual differences in musicianship may be a factor in determining the prevalence of spared musical memory in PLWD (Baird & Samson, 2009). A certain level of selection (either self- or researcher-led) in terms of dementia-music studies has the potential to propagate different results in documented retained musical abilities (Baird & Samson, 2009). Musicianship was initially reported in Agustus et al. (2018) as a contributing factor to various activation patterns observed during music familiarity processing, however when analysis was re-run without the two most-experienced musicians' data included, this effect was attenuated. It would be of interest in further investigations with a larger sample to establish the role of musicianship as a potentially mediating factor in music-linked pupillometry and self-rating responses.

The potential role of medication in mediating autonomic responses also warrants further consideration. Acetylcholinesterase inhibitors (AcHE; generic names: donepezil[Aricept], rivastigmine and galantamine) specifically seek to extend Ach presence in neurosynaptic clefts to maintain cognitive functioning in people living with Alzheimer's disease. These medications may also influence pupillometry responses and increase the likelihood of preservation effects relative to controls. For example, Patient EN (Cuddy & Duffin, 2005) – reported to have preserved semantic musical memory in severe dementia stages - was also taking prescribed symptomatic medications at the time. Furthermore, donepezil has been demonstrated to enhance

degraded speech comprehension in individuals with tAD (Hardy, Hwang, et al., 2017), indicating that AcHE medication may serve as an aid in ameliorating processing of complex acoustic stimuli. The current sample size of individuals in the PLWD participant groups not taking a symptomatic medication was too small to investigate group differences with confidence. On the one hand, the fact that the PCA group (which reported an overall higher rate of individuals taking medication) was the group which demonstrated only marginal effects, indicates that this concern may not be warranted. However, a double-blind placebo-controlled cross-over clinical trial revealed that some AcHE medication (donepezil) had no significant treatment effect on the primary outcome measure (MMSE) when administered to 18 people living with PCA (Ridha et al., 2018). This implies that whilst AcHE medication may not be sufficient to mask symptoms or retain autonomic responding in the PCA group, its capacity to mask any cholinergic deficits (and potentially preserve autonomic responses) may still be occurring for other clinical cohorts whereby AcH inhibition is successful, for example in tAD . Therefore, whilst the inclusion of the medication questionnaire in the current study was to some extent informative, a larger sample of participants not receiving medication, plus future iterations requesting a list of brand-name medication may be warranted. Owing to findings whereby pupil response reduced by 7.4% in participants with tAD four weeks post-5mg daily donepezil dosage compared with baseline measurements (Estermann et al., 2006), ascertaining the length of time a person has received AcHE medication may also prove informative.

7.2.2 *Methodological challenges (and lessons) for future pupillometry studies in PCA*

The challenges encountered in acquiring pupillometric data in individuals with PCA became a galvanising point to seek out methods to reconcile the issue regarding on-screen fixation position. Traditionally, corrective methods in pupillometry studies are either not attempted (with the issue being cited as a study limitation) or methods are invoked which result in extreme cut-off thresholds for data inclusion (e.g. any data points with a gaze position exceeding greater than one degree of visual angle from central fixation). The latter approach is often a tenable approach in younger neurotypical populations whereby relatively little data are lost (e.g. 6% as observed in Jagiello et al., 2019), however as noted in the current study this was not a viable approach. Reconciling this issue is particularly pertinent to the current work, in that

future application would involve testing individuals with more severe stages of dementia, who are likely to incur oculomotor challenges either by virtue of normal physiological aging, as a marker of progression in the disease, or a combination of the two (Bowen et al., 2016).

Hayes and Petrov (2016) undertook a series of experiments which established the possibility of applying a correction to acquired data whereby pupil rotation occurs by virtue of the experimental paradigm (e.g. in reading studies). This correction can be applied post hoc, provided the experimenter has a precise measurement of particular X-Y-Z distances reflecting their bespoke experimental setup. Unfortunately, Y-axis measurements (e.g. distance between the participants' pupil(s) and table position) were not acquired in the current study, and therefore this correction could not be applied to the dataset. However, equipped with the knowledge of what is required to calculate and implement the correction, this will be a relatively simple addition to incorporate into future experimental designs pertaining to pupillometry research. Correcting for this nuisance artefact means that it will be possible to obtain more accurate representations of pupil size in clinical populations regardless of the confounds of eye position, potentially uncovering spared psychosensory responses previously thought unattainable. Pupillometry therefore would remain a candidate biomarker for measuring psychophysiological responses to music in PCA participants, and in future for those in more severe disease stages relative to the current sample.

Whilst important to establish pupil profiles of different syndromic groups, it is arguably also informative to study individuals with PCA as a proxy for people living with later stages of dementia. The Prevalence of Visual Impairment in people with Dementia (PrOVIDe) study indicates that 50% of people with dementia in care homes have uncorrected visual impairments (owing predominantly to either cataracts, macular degeneration, or both) (Bowen et al., 2016). While these visual impairments may not be the same as the corticovisual problems in PCA, they are likely to give rise to difficulties in performing fixation tasks. This work has been a helpful exercise in considering and learning potential ways to overcome this from a methodological point of view, to enable pupillometry research to continue to be accessible for people living with more advanced stages of dementia.

In addition, despite the fact that effects were observed in the current study using standard pre-processing techniques, it is sensible to consider experimental designs that have yielded results in other psychosensory pupillometry studies who have recruited older populations, and some further checks which future iterations of Play It Again could include. For example, Piquado et al. (2010) baselined pupillometry responses relative to individuals' pupil change to light intensity as opposite to pre-task baseline, reconciling any potential confounds of reduced physiological reactivity in the physiological aging process, with or without a diagnosis of a dementia. Incorporating these alternative approaches should be considered in future experimental design to enable testing between pre-trial and light-intensity response baselining when establishing psychosensory pupillometry responses to music in individuals with a progressive condition. In light of the trial number-pupil response relationship findings, another experimental design of future interest would be to implement a paradigm requiring even less cognitive effort than Play It Again. One example is the structure employed by Jagiello et al. (2019) in splicing one personally-familiar song with an analogous unfamiliar song (provided adjustment was made for likely more latent pupillary velocities by extending the excerpt snippet lengths). This approach - while questionable in terms of application to dementia care settings where music is often other-selected for PLWD rather than self-selected - would further reduce the cognitive effort involved for participants, and enable us to test whether or not the current pupillometry effects pervade in the absence of subsequent trial-by-trial active categorisation prompts. Finally, the inclusion of a brief audiometry test, which was regrettably not part of the current design, should be incorporated in any future iteration of Play It Again (Hardy et al., 2016).

7.2.3 Complexities of autonomic responses: what do they mean?

One broader limitation of investigating canonical autonomic responses in general is the complexity in relating the rather global 'increase/decrease' profile to clinically-relevant outcomes, and/or interpretations of an individual's complex internal experience. The curvilinear profile of the pupil-affect relationship in Play It Again demonstrated that increases in autonomic responses were observed for both canonically 'good' (self-rated pleasant) and 'bad' (self-rated unpleasant) excerpts. Chapter 3 discussed the circumplex model of emotion in relation to a number of

previous studies grappling with the same predicament when using physiological markers such as pupillometry, which is accepted as a marker of increased arousal, but "more controversial" in its capacity to distinguish valence (Salimpoor et al., 2009; p. 2). It therefore appears questionable as to the clinical use this may have if one were to only have physiological response data available. The choice of experimental design and statistical analyses in future work could, to an extent, overcome this challenge. It is however important to query that: if increases are observed for both positive and low valenced music appraisals, how might we infer from these increases in a real-world setting whether a patient is enjoying the music, or hating every second (Harding et al., 2017)?

Multistrand physiological data collection collected simultaneously may provide a more holistic understanding of the physiological processes underpinning a musical experience, leading to reduced ambiguity in interpretation of results. For example, concurrent data collection in Our Dementia Choir via the Empatica E4 device comprised four types of physiological data collection, which provided further context to interpretation of the main physiological response of interest (i.e., EDA). Increases in autonomic arousal, at first glance, may appear to be reflecting a stress 'fight-or-flight' response. However, when observed with a concurrent decrease in heart rate (e.g., as observed in Chapter 6), the interpretation of the autonomic arousal findings point to EDA responses reflecting a more positive experience (e.g. engagement). Leveraging multistrand physiological data collection alongside sophisticated statistical approaches may therefore further elucidate alignment or interdependencies between autonomic indices, whilst increasing confidence in the generalisability of the results of a single physiological index where alignment is found between autonomic indices (e.g. cross-validation between pupillometry and EDA; see Bradley et al., (2008)). More broadly, this question also causes us to reflect on the complexity of our relationship with music, which makes this stimulus far more challenging to work with on an 'intervention' level relative to other types of stimulation (Särkämö & Sihvonen, 2018). Positive emotions have been linked with sad music (eliciting feelings of 'comforting' and 'sweet sorrow'; Eerola and Peltola (2016). In addition, Granot et al. (2021) reported that, throughout the Covid-19 pandemic, music listening was used as a tool to process associated negative emotions. If music is used for PLWD as a silver bullet to only serve an 'up', we may be overlooking opportunities for 'sad' music to have

some positive, affirmatory effects. Individuals without dementia interact with sad music to match or externalise emotions (Eerola, Peltola, & Vuoskoski, 2015) reflecting that canonically 'sad' music may in fact act as a catharsis of sorts. The opportunity to engage with 'sad' music may be arguably more pertinent for individuals with progressive health conditions as they journey through the unchartered territory of ill-health. Efforts have been made, including in the current work, to tease apart different physiological profiles for subjective appraisals of music; however, there remains a lot of ground to cover before being able to confidently incorporate this nuanced use of music in healthcare settings.

7.2.4 Coupling physiological responses with alternative measures

The considerations above have ramifications in terms of selecting future candidate scales and self- and/or proxy-report measures in dementia-music studies. Exploration of other self- and/or proxy- report scales relating more to a global measure of 'engagement' may be a more feasible starting point for any potential system juggling a biosignal-based dementia-music interface, rather than encountering limitations around attempting to profile particular emotions on a physiological level. A range of tools are available to measure engagement in PLWD (Matcham, Thoman, Sobol, Sanchez, & Gaspar, 2022), some of which are specifically designed as a measure of the holistic experience of engaging with music (Music in Dementia Assessment Scale; MiDAS) (McDermott et al., 2015). Another consideration is whether or not future experimental design in this field should consider other outcomes with which to establish associations between physiological responses and other in-the-moment data capture. This could involve a mixed-methods design whereby associations between physiological responses are explored in relation to coded qualitative observations of in-the-moment music experiences (Dowlen et al., 2021). Alternatively, the recent development of an 'in-the-moment' Video Analysis Scale of Engagement (VASE) offers the opportunity for within-session observable engagement ratings of PLWD behaviours using a dynamic and continuous measurement tool (D. L. L. Lai et al., 2020). Combining the VASE with pre- and post-session established measurements of engagement (e.g. MiDAS) may improve further the ability to couple physiological increases/decreases specifically to the types of behavioural outcomes which are prioritised in care settings.

7.3   A comment on the differing profiles of 'the moment' explored in the current work

Whilst both Play It Again and Our Dementia Choir explored within-session data acquisition using passive techniques recording continuously in the moment, the length of 'the moment' differed between the experimental studies. This even differed between chapters despite invoking the use of the same paradigm. For the purposes of Chapters 3 and 4, Play It Again focused firstly on the measurement of much more transient responses to music, isolated by 12-second excerpts. For Chapter 5, the moment was extended to explore the entire within-session recordings across the music listening activity to elucidate tAD-PCA classification performance. Whilst the experimental design differed in many other ways, the moment as measured in Our Dementia Choir similarly retained a focus on within-session measurements of the choral rehearsal as a whole. Elongating the moment in this context was intentional in order to address the recognised moment-to-moment heterogeneity within each rehearsal. It may nevertheless be beneficial to further segment whole-session music activities to determine specific contributors to physiological responses observed. It is clear that despite different timeframe parameters, within-session continuous recording has added insight into some of the physiological processes occurring at the intra-session level, either due to emotional and cognitive appraisal of music (Play It Again; Chapters 3 and 4); syndromic group differences (Play It Again; Chapter 5) and engagement in a music activity relative to social interaction (Our Dementia Choir, Chapter 6). The current work adds to the evidence base of differential autonomic responses occurring in PLWD in both short timeframes (e.g. within 12 seconds) and in response to whole music activities as they occur.

7.4   A proposal for in-the-moment RCTs

It is encouraging to observe that references to 'the moment' are becoming more prevalent within the wider dementia and music literature (Dowlen et al., 2021; Keady et al., 2020; Zeilig, West, & van der Byl Williams, 2018) and policy publications (All-Party Parliamentary Group on Arts Health and Wellbeing, 2017). Hackett, Sabat, and Giovannetti (2022) have even asserted that these clinically-relevant shorter-term effects should be prioritised over long-term effects when evaluating the efficacy of

music interventions in this group. However, work remains to be done to disseminate the nuanced understanding necessary to successfully incorporate music in dementia care and support, and achieve buy-in that social prescribing provision continues to be a valuable tenet of healthcare even if only short-term effects are likely. Despite a publication date nearly a decade ago, DeNora and Ansdell's (2014) explanation as to why music may not be more readily available in mainstream healthcare provision remains pertinent, despite the drive to increase social prescribing via NHS 'link workers' and an increasing evidence base in the power of music in the moment for PLWD (Greengross, 2020).

DeNora and Ansdell (2014) venture that RCT experimental designs arguably set music up to 'fail' in proving itself as a useful tool in dementia care and support. For example, as observed in Chapter 2, the relationship between in-the-moment responses and longitudinal pre-post effects (i.e., over the entire course of an intervention) has been shown to be relatively weak (e.g., no observed long-term changes in heart rate at 8 or 12 weeks in final sessions versus pre-intervention levels) (Gok Ugur et al., 2019; da Rocha et al., 2022). It therefore appears as though acquiring in-the-moment response data at the same timepoints as typical pre-and post-whole intervention data collection is unlikely to yield useful information about a music intervention's impact on PLWD. However, recent findings implementing a previous feasibility, efficacy and acceptability trial protocol (Weise et al., 2018) demonstrated a reduction in observed problematic behaviours (BPSD) that were coded within the music listening sessions relative to BPSD frequency coded at pre- and post-session timepoints (Hillebrand, Weise, & Wilz, 2023). Had data collection taken place at pre- and post-session timepoints only, there would have been no differences observed, with ramifications on interpretation of the music's capacity to reduce BPSD. Arguably, the inclusion of in-the-moment data collection during music interventions that are being trialled enable researchers to investigate whether the music intervention can be associated with momentary benefits to PLWD. Paradigms such as RCT which additionally incorporate in-the-moment physiological response investigation on a larger scale and in PLWD provide promise, in that they may satisfy policymakers in accepting the power of the music in its short-term context in dementia care and support (Hackett et al., 2022).

Another proposal around the inclusion of physiological measures being incorporated into RCTs is the development of biosignal-based feedback music interventions, whereby the physiological measures become part of the music intervention that needs to be trialled. The attraction of RCT designs in arts in healthcare research is so great owing to the fact that RCTs are most readily accepted in medical research (and therefore in policy) as the gold standard method to investigate a capacity of an intervention to effect some sort of change - extending this to a hopeful improvement - in a clinical population. A compromise of sorts, which may in fact result in powerful findings for the field, could emerge from a combination of in-the-moment physiological research with large-scale RCT design. In line with this proposal is a biosignal-based feedback music provision system currently being developed as part of the Music Attuned Technology – Care via eHealth (MATCH) study led by Felicity Baker (https://www.musicattunedcare.com/), due for feasibility testing in 2024-2025. The results in the thesis could be utilised to inform the development of similar applications which could be subjected to future RCT testing.

7.5 Disseminating findings of transient effects in the context of 'music as panacea'

Differential responses observed at both the physiological self-report levels in both studies indicate that some musical experiences did not have as much of an impact as others. Whilst this underscores the argument for conducting in-the-moment research in this field, it is important to note the sensitivity around gently challenging the ubiquitous impression that, for PLWD, all music 'works'. Whilst a well-intended mantra, the current work supports the proposal that this statement is digestible only if invoking the right music, on an individualised basis, at that specific moment for them. It should also be noted that not all studies investigating the effect of music on PLWD report universally positive responses (Garland, Beer, Eppingstall, & O'Connor, 2007; Nair et al., 2013). Although numbers of negative published findings in this research area are small, this may simply reflect publication bias rather than demonstrating a true representation of music's impact on this clinical population. Where negative effects are observed, these must be considered carefully alongside positive results. The findings in the current work are particularly helpful in that they concurrently emphasise

the importance of finetuning our understanding of responses to individualised music in PLWD. The messaging around music in dementia care and support, whilst well-intended, can often give the impression that music can 'do no harm'. Differing responses and engagement with music pieces, alongside protocols in place to avoid 'red flag' songs (Playlist for Life, 2019), indicates that this is unlikely to be the case. However, disseminating this narrative whilst charitable and healthcare organisations struggle to find funding for arts in health provision for PLWD remains a challenge for researchers in the field. Conveying the interpretation of one's findings sensitively and couching this as a desire to optimise and retain personhood in PLWD is paramount to being perceived as helpful for future music practice in dementia care, rather than dismissive of the entire endeavour (Camic et al., 2018).

7.6   Shorter-term research and practice implications: Findings of note

The consistent null finding across healthy older adults and PLWD for any associations between reminiscence bump-released music and subjective appraisals of music is an important finding which shapes recommendations for music practitioners in establishing potential music selection for PLWD whereby self-selection is not a possibility. A recent survey (American Association of Retired Persons' *Music and Brain Health* survey*)* revealed that only 9% of adults exclusively enjoy music released within their generation; 49% enjoy music released after their generation and over 80% of adults enjoy music released before, with some evidence for intergenerational influences of parents' and grandparents' own reminiscence bumps on individual preferences (Jakubowski, Eerola, Tillmann, Perrin, & Heine, 2020; Mehegan & Rainville, 2020). The current findings therefore reflect that music familiarity is not sufficient to elicit pleasant musical experiences, therefore requiring more of a music practitioner's undoubted skill in exploring other avenues to increase the chances of music eliciting meaningful moments for PLWD. The increased recognition of algorithm-generated new music (Raglio, Maestri, et al., 2022) or improvisatory music-making by, with and for PLWD (Zeilig et al., 2018) may also help to steer away from the long-held view that music familiarity is a guarantee of 'success' in dementia care and support. Improvisatory music also acts as a potential leveller in terms of removing the constrictions around a 'shared musical lexicon' which can be exclusionary for individuals across a range of cultures accessing music activities. Music improvisation

also provides the opportunity for music therapists to explore further heightened physiological responses and the extent to which this may reflect joy or distress (e.g. as outlined in the *Music and Dementia podcast,* O'Hare, 2020).

Whilst some music-physiology paradigms have utilised participant-selected stimuli in neurotypical research (e.g. Jagiello et al., 2019), there are undoubted advantages to researcher-led stimulus selection from a scientific perspective as the field of dementia-music-physiology research starts to grow. For one, researcher-selected stimuli enables standardisation (i.e., physiological and behavioural responses are acquired from each participant in relation to the same pieces of music). Researcher stimulus selection also facilitates the systematic control over certain variables, not only in terms of fulfilling the main study objectives where condition manipulation is key (e.g., identifying known/unknown songs) but also creating a stimulus set which balances potential extraneous variables such as tempo and lyrical content. A researcher-selected stimulus set also enables future work to take place to determine the replicability of results in larger samples and different syndromic groups for direct comparison to the current findings. Whilst the application of the work (to support individualised music use in PWLD) therefore appears on first glance at odds with experimental paradigm designs in dementia-music research, the motivation for testing researcher-selected music at this early stage of this research in PWLD is to increase our understanding of physiological responses to a set of test musical stimuli which fulfil the scientific objectives for the work (as discussed in Salimpoor et al., 2009). By creating a researcher-led stimulus set that was hypothesised to have particular influences physiological responses (e.g. the pupil old/new effect and pupil-affect relationship), the current stimulus set has the potential to inform future work that may incorporate participant-led stimulus sets designed to elicit similar underlying processing.

In addition, there are real-world examples whereby music selection is tailored to the individual (hence 'individualised') by the PLWD being consulted about their music preferences and supported in the provision of that music. However, with the reality of later-stage dementia creating difficulties in communicating verbally, music is often selected by others. This can either be achieved using either informally through discussion with the PLWD's loved ones who are aware of their preferences, or via questionnaires (e.g. using the Assessment of Personal Music Preference (Family

Version)) (Gerdner, Hartsock, & Buckwalter, 2000). Alternatively, observations of the PLWD whilst listening to music can also provide visual cues about a person's musical experience in a person with later-stage dementia (D. L. L. Lai et al., 2020). For example, interpretations as to a PLWD's experience of music may be inferred by observations of the person's facial expressions (e.g. smiling, frowning, downward turn of the mouth), emotional outbursts (e.g. crying) or movement (e.g. clapping, tapping feet, reaching out to hold the hand of a peer or loved one). Observations of this kind may be especially informative to formal caregivers who may not be aware of a person's musical preference, or to music therapists who may be working with a person with later-stage dementia in order to achieve a particular functional outcome (e.g. playing upbeat music to encourage dressing). The role of music for people with late stage dementia is eloquently described by Brancatisano et al. (2020) in the TMCM framework, in that music has the capacity to: engage, increase alertness, evoke emotion, encourage physical movement including verbal fluency, facilitate synchronous experience with others and promote social interaction, and promote a sense of identity through an individual's musical likes and dislikes.

7.7 Practical challenges associated with pupillometry and ANS measurements in persons with dementia

The practical challenges associated with autonomic data collection in PLWD in many cases relates to the autonomic index a researcher has selected to measure the response, and the technical requirements which influence continuous data recording from the particular devices involved. For example, practical challenges associated with the collection of pupillometry data in PLWD may reflect syndromic group symptom profiles. The current work highlighted that individuals with PCA may struggle to maintain fixation on a single point of a screen for pupillometry recording to take place, owing to degradation of cortical areas engaged in motor tasks such as sustained and stable eye positioning. Results from the PrOVIDe study also indicated that this difficulty may be encountered by individuals with later-stage dementias encountering visuospatial and perceptual difficulties (Bowen et al., 2016). In addition, for individuals with moderate to later-stage dementia, the capacity to follow task instructions, upon which obtaining continuous pupillometry data collection are reliant (e.g. when to blink), may be diminished. Attentional levels may also be compromised in this population, as dementia can lead to wandering attention or a decrease in alertness levels. Therefore,

whilst music may act as an engaging stimulus in the short-term, to avoid attention (and therefore eye movements) wandering, any pupillometry data collection paradigms are required to be relatively short, which can compromise the number of trials included and call into question the reliability of the results with relatively few opportunities to test different stimulus properties. A shorter paradigm aligns with previous recommendations with music intervention delivery and optimum effects (e.g. 30 minutes for a single musical event) (Kurita et al., 2006; Sakamoto et al., 2013). However, with increased task demands owing to pupillometry data collection, it is likely that any musical paradigm collecting this type of data would need to be shorter than this. Indeed, the Play It Again paradigm duration was approximately 10-15 minutes, although individual variation was introduced by not restricting the timeframe for subjective self-report responses. Challenges which relate to general data collection of autonomic responses relate to discomfort, stress and anxiety influencing results. Although feasibility was not directly assessed as part of this work, all participants were happy to wear the Empatica devices and engage with the computer screen, however as these individuals were all research-keen and had capacity to weigh up and understand the invitation to don wearables, this high tolerance may not be reflective of the take-up of wearables in people living with later-stage dementia. To address these challenges, researchers should adapt autonomic data collection protocols to accommodate specific needs and capabilities of individuals with dementia, taking factors into account such as: experience in taking part in research and familiarity with the research process, capacity to consent, paradigms with simple instructions which can accommodate flexibility that may need to be exercised when collecting data from PLWD. Subject to permission from the PLWD where possible, researchers should also collaborate with family members, loved ones and professional caregivers which will ensure PLWDs' wellbeing when taking part in this type of research and facilitate data collection.

7.8 Longer-term/future research and practice implications: Incorporating physiological measurements into music provision for PLWD

The future outlook for the use of physiological responses to music in PLWD centres around the potential for wearables and sensors to be incorporated into music interventions delivered to PLWD. Integrating technology could include incorporating biofeedback capacities into future iterations of music streaming platforms via

accompanying wearables, and/or to stream physiological responses in real-time to a dashboard readable by music practitioners, healthcare professionals and caregivers (Melander et al., 2018). The current findings contribute to the knowledge base required to develop such platforms, with particular reference to the extent physiological biosignals may complement music practitioners' work in dementia care and support, particularly to verify the impact of music in later-stage post-verbal individuals (Au-Yeung et al., 2022). One venture, *MediMusic,* has sought to integrate physiological readings into a feedback system which updates in relative real time (trialled initially in the Lancashire Teaching Hospitals NHS Trust) (Tasker, 2021). The findings from Play It Again and Our Dementia Choir add to the potential physiological candidate biosignals that could be explored in this way for PLWD at different disease stages, confirming moment-to-moment differences in physiological signals during two different formats of music delivery. The success of any technological integration in dementia care is however contingent on the acceptability of the wearable devices or recording sensors in the clinical populations from which researchers seek to acquire physiological data. A recent review by MacRitchie, Floridou, Christensen, Timmers, and de Witte (2023) cites that, for the most part, integrated designs remain in the protoyping stages, however there is promise for this field in particular to flourish in the coming years as our knowledge of physiological responses to music in PLWD of different etiologies, syndromic groups, and/or disease stages increases. Whilst a longer-term direction, systems which could accurately incorporate biosignal responses to music in a feedback format may be of particular value in care home settings, in which some activities incorporating music (e.g. reminiscence therapy) have been proposed as being more effective relative to other settings (Woods, O'Philbin, Farrell, Spector, & Orrell, 2018). Music generation systems based on an individual's physiological response in a care home setting may also facilitate the professional carer-PLWD relationship, by virtue of revealing musical preferences which may surprise, or forge previously-unknown communalities between formal caregivers and residents (Waters, Sousa, Orrell, & McDermott, 2022). It is important to note however that this area of research should not be interpreted as a means to replace in-person music provision or the enjoyment of music with loved ones present. The motivation for this research and its future direction reflects a proposed solution to the funding challenges and sustainability of delivering individualised music therapy to PLWD, which is incredibly valuable but sadly not an accessible option for all. Coaching

opportunities for caregiver delivery (extending previous work in this area) would be an important accompaniment to any biosignal-based system which may be developed in the future (Särkämö, 2020; Särkämö et al., 2016a, 2016b; Särkämö et al., 2014).

Considering the challenges mentioned in Section 7.7, there are a number of practical limitations that hinder the extent to which the current findings can be taken up immediately for the development of neurofeedback-type music interventions in severe dementia. Further knowledge of how tightly specific ingredients of a phenomenological musical experience can be linked with physiological responses (derived from research in both clinical and neurotypical populations) is required to advance the possible clinical applications of this work. Encouragingly, studies have shown it is feasible to collect physiological data in individuals with severe dementia, in certain cases using the same devices as adopted in Our Dementia Choir (e.g., Sun et al., 2021). Acquiring and analysing further autonomic data (pupillometric and electrodermal data in particular) alongside other physiological and behavioural data strands during music experiences in individuals with later-stage dementia will be one avenue for future research.

Psychometric properties of outcome measures used to assess the benefits of music for PLWD broadly refer to the extent to which a measurement (a) accurately assesses the construct it is designed to measure (validity); (b) provides stable and consistent results over time (reliability); and (c) is sensitive to detecting meaningful change in the intended construct, either in response to an intervention/treatment or longitudinally (responsiveness). As observed in Chapter 2, music-dementia-physiology studies are relatively scarce, and there has been little direct investigation, comment or discussion thus far in the field as to the validity, reliability and responsiveness of the selected measures in a dementia-music context. Establishing the psychometric properties of physiological measures in this context will be important for ascertaining the utility and credibility of these measures for assessing responses to music in PLWD. Future research based on this initial work, outlined below, aims to further our understanding of the psychometric properties selected to measures responses to music in PLWD.

7.9   Future direction

7.9.1   Further analyses of acquired Play It Again datasets

Other data acquired as part of Play It Again have yet to be analysed, as this was deemed beyond the scope of this thesis. This includes timecourse information within the pupillometry dataset (extending beyond the velocity measures explored within Chapter 5), using more sophisticated statistical modelling approaches such as Generalised Additive Mixed Modelling (GAMMs). These modelling techniques are recommended for pupillometry data due to their ability to take into account non-linear trajectories over the time course of pupil data acquisition within a trial or experimental session, whilst continuing to take into account interindividual/stimulus differences by virtue of including random effects (van Rij et al., 2019).

Given the limitations discussed in relation to relatively global parameters in physiological responses such as pupil dilation (as raised in Section 7.2.3), multistrand data collection could elucidate the landscape of a person's experience beyond what has been described in the work. Further in-depth analysis of the Empatica® E4 data, particularly acquired during Play It Again, could address some of the challenges in acquiring pupillometry data in individuals with PCA, and provide a multistrand narrative of whether old/new and/or curvilinear biosignal-affect profiles could be observed in other autonomic responses (e.g EDA (Bradley et al., 2008), and/or HRV) in this clinical population.

The link between EDA and pupillometry is particular pertinent to the thesis owing to the different approaches adopted for the Play It Again study and Our Dementia Choir. Table 1.1. demonstrates an overlap of the preganglionic neurons that communicate signals to both pupil and skin, indicating a potential convergence in pupillometry and EDA profiles. Findings of similar overall profiles of EDA and pupillometry responses have been reported in previous studies, increasing the specificity of pupil dilation reflecting sympathetic processing (as pupil dilation may also be mediated by parasympathetic processing, see Table 1.1.). Bradley et al. (2008) tested this specifically by collecting pupillometry, EDA and heart rate data while viewing pleasant, neutral and negative affective images, whereby pleasant and unpleasant images were controlled for arousal. The authors hypothesised that, should pupil dilation demonstrate a similar profile to HR, this would indicate that this process was governed

by parasympathetic innervation; in contrast, profiling of the pupil dilation response similar to the EDA observations would indicate the likelihood of pupil dilation reflecting sympathetic processes. Hierarchical multiple regression analyses indicated that skin conductance changes (but not heart rate changes) was a "highly reliable predictor" of pupil dilation changes (Bradley et al., 2008, p. 5). Furthermore, principal component analyses conducted on pupil response, skin conductance and cardiovascular (RR interval) data revealed grouping of maximum response latencies for pupil and EDA responses in a study exploring multistrand autonomic responses to affective touch in children with autism (Bufo et al., 2022).

Extending the Play It Again analyses which incorporated both timeseries exploration and other physiological data strands may therefore increase our understanding of the complex processing that underlies a musical experience in healthy older adults, and people living with typical and atypical Alzheimer's disease. This approach will also enable associations between physiological measures to be investigated, extending our understanding of the measures' convergent and discriminant validity when applied to this clinical population as they engage with music.

7.9.2 *Extending Play It Again: Further data collection and extension*

The investigation of real-world music activities (e.g. Our Dementia Choir) holds certain ecological validity advantages over laboratory-based work. However given the current technological and dementia-music knowledge landscape and the repeated citing of within-session heterogeneity hindering advancement of how real-life music activities 'work' in PLWD, focusing instead on music-listening activities (with the potential to extrapolate to naturalistic settings) is an optimal starting point for any subsequent investigations.

There is clinical rationale to implement the Play It Again paradigm in other clinical cohorts. A third dementia subtype, termed logopenic variant Primary Progressive Aphasia (lvPPA), is primarily accounted for by underlying pathology due to Alzheimer's disease (Marshall et al., 2018). lvPPA is characterised by a different trajectory of atrophy compared with tAD and PCA, predominantly involving left tempero-parietal junctions (comprising inferior parietal and posterior-superior temporal cortices) and resulting in initial presenting symptoms of difficulty with sentence repetition and word finding difficulty, coupled with 'tip-of-the-tongue' hesitations (Marshall et al., 2018;

Ruksenaite et al., 2021). While the current work sought to explore potential syndromic differences between the visual variant form of Alzheimer's disease and its typical memory-led counterpart, a natural extension may be to revive the ethical approvals for this work and explore within- and between-group differences in this syndromic group. Whilst noting findings of core auditory processing deficits in people with PPA, it appears that individuals with the logopenic variant do not encounter these to the same extent as observed in other variants (noting small sample sizes) (Grube et al., 2016). Furthermore, whilst Hardy, Agustus, et al. (2017) reported reduced activation of posterior superior temporal cortices in people with lvPPA in relation to processing phonemic spectral speech signals, processing of music familiarity appears to be preserved in lvPPA relative to controls (Golden et al., 2017). Whilst this result was acquired from a small sample of lvPPA participants (N = 5; Golden et al., 2017), it indicates that a preservation of autonomic processing may also be demonstrable in this population. This work is due to take place as part of an upcoming PhD project within the Dementia Research Centre, UCL.

7.9.3 Lessons learned from Our Dementia Choir: Applying the analysis pipeline to future work

Work led by Professor Jason Warren (UCL Dementia Research Centre, Co-I Crutch) is underway to establish ecologically-valid individualised music interventions for people living with typical Alzheimer's disease and FTD within the MANDDOLIN grant (Music And Neuroscience against Dementia: from Designs to Outcomes through Listening INclusively INformed for INdividuals). This interdisciplinary project is incorporating physiological indices measured by the Empatica® E4. Physiological responses and video recordings will be collected before, during and after music listening in different experimental conditions. The video recordings will be used to generate automated facial electromyography (fEMG) data using FaceReader. All study procedures have been designed to take place virtually to explore the possibility of remote biosignal data collection, whereby participants can engage in research in their own homes (Owens, 2020). The music conditions of interest will initially focus on music familiarity (e.g. responses to a personal resonance playlist vs. general familiarity playlist, experienced simultaneously by a PLWD and their care partner acting as a control) and 'musical groove'. The extent to which the within-playlist Empatica® and fEMG data can be coupled with a variety of proxy- (MiDAS; caregiver rated) and in-

the-moment (VASE; independent academic and music therapist-rated) scales will be established (D. L. L. Lai et al., 2020; McDermott et al., 2015). My contribution to this work to-date was co-establishing the Standard Operating Procedures for data collection to enable subsequent time synchronisation with video and Empatica® data. This involved adapting the bespoke R script I created for the Our Dementia Choir analysis to respond to the requirements of this project (e.g. automatically selecting data windows of interest in accordance with relevant data label ('baseline') and time-frequency labelling, removing irrelevant data). This work is currently underway, facilitated by the knowledge gained from the work described in this thesis to further benefit the dementia-music field ("MANDDOLIN4 - Music And Neuroscience against Dementia," 2022).

7.9.4 Other physiological candidates for future research

Physiological biomarker candidates may be more or less suited to different dementia subtypes based on leading symptoms (e.g. pupillometry in PCA populations). There are, however, other burgeoning physiological markers that were not considered at the time of the experimental design, but may enable automated distinction between typical and atypical Alzheimer's disease. Recorded vocal responses to the Play It Again paradigm in the healthy control, PCA and tAD groups could - in principle - be analysed to detect differences in vocal properties according to disease presence or absence, type, and/or add insight into disease severity. The current knowledge base of vocal analysis has been derived mostly from progressive neurodegenerative clinical populations, including Parkinson's disease, AD and MCI (Ahmed, Haigh, de Jager, & Garrard, 2013; Hlavnička et al., 2020; Tóth et al., 2018). The main tenets of vocal analysis involve audio signal decomposition (voice from background), voice feature (analysing components of the voice signal), vocal biomarker (relating one or more vocal features to a clinical outcome) and vocal assistance (analysing the use of devices such as Alexa) (Fagherazzi, Fischer, Ismael, & Despotovic, 2021). Models analysing x-vector vocal features have demonstrated high levels of specificity and sensitivity in line with standardised scales of depression (PHQ-8) and anxiety (GAD-7) (Kwon, 2022). As this work develops, it may be possible for brief verbal responses to the Play It Again music ratings to provide an additional insight into how physiological responses to music in PLWD may also be being mediated by these psychological

conditions, noting mood difference findings in Garrido et al. (2018). One evidently major limitation of the vocal biomarker approach is that, by nature, it limits itself to fixed parameters of disease severity (e.g. earlier-mid stages of dementia prior to individuals losing the ability to communicate via language in later stages).

7.9.5 Future prospects for the line of research presented

While our understanding of the dementia-music-physiological relationship is in the early stages relative to other dementia-music enquiry, the current findings contribute to evidence that we may in future be able to use wearables and technology to 'read' in-the-moment response to music in PLWD. The need for technological solutions to tailor music provision for PLWD increases as dementia projections continue to rise. A number of factors (including socioeconomic, geographical and other demographic factors) can preclude PLWDs' access to consistent and sustainable qualified music therapy. Music therapists are able to conduct in-the-moment musical recalibrations to attune music provision to PLWD, to achieve desired therapeutic outcomes and support their emotional wellbeing (Raglio et al., 2018). The line of research outlined in the thesis recognises that there is a potential to provide a technological solution to this. Digital platforms may in the future be able to interpret biosignals in a similar way to a music therapist noticing subtle expression and/or movement changes in a person's engagement with a particular song, and use that data to automate the continuation or changing of songs to support that person's therapeutic outcomes and wellbeing. The development of such platforms would enable more individuals with dementia the opportunity to access music which can be attuned in real time to their own biosignal feedback, over the course of the disease, using resources likely to already be in their possession (e.g. a smartwatch and smartphone).

7.10 Closing remarks

Ultimately, evidence for preserved processing in PLWD which may serve future wellbeing is of course approached with excitement and interest from healthcare professionals and researchers - and even more so from the individuals living with or alongside these conditions. The current work has focused on the available literature conducted with PLWD, contrasting this to neurotypical findings for the most part. Music-physiology research however is gaining traction in many different healthcare settings and contexts (Dingle et al., 2021; Fancourt & Finn, 2019) which may combine

with current and future findings to justify and inform the development of larger scale studies, and/or potential technological platforms that may be developed. Parallel technical and technological enhancements across different disciplines only serve to complement this research endeavour in the search for new and novel ways to keep lines of communication open for PLWD to engage with music, even when experiencing later disease stages. The recency of a large proportion of studies and large-scale study protocols cited within this thesis is indicative of a watershed moment for this field. Understanding how to create and offer truly individualised, biosignal-based musical experiences for people living with a dementia may well be in our reach.

8   REFERENCES

Agrawal, Y., Carlson, E., Toiviainen, P., & Alluri, V. (2022). Decoding Individual differences and musical preference via music-induced movement. *Sci Rep, 12*(1), 2672. doi:10.1038/s41598-022-06466-3

Agresti, A. (2010). *An introduction to categorical data analysis* (2 ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Agustus, J. L., Golden, H. L., Callaghan, M. F., Bond, R. L., Benhamou, E., Hailstone, J. C., . . . Warren, J. D. (2018). Melody Processing Characterizes Functional Neuroanatomy in the Aging Brain. *Front Neurosci, 12*, 815. doi:10.3389/fnins.2018.00815

Ahmed, S., Haigh, A. M., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain, 136*(Pt 12), 3727-3737. doi:10.1093/brain/awt269

Alam, M., Roy, N., Holmes, S., Gangopadhyay, A., & Galik, E. (2016). *Automated Functional and Behavioral Health Assessment of Older Adults with Dementia*. Paper presented at the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE).

All-Party Parliamentary Group on Arts Health and Wellbeing. (2017). *Creative Health: The Arts for Health and Wellbeing*. Retrieved from https://www.culturehealthandwellbeing.org.uk/appg-inquiry/

Allan, L. M., Ballard, C. G., Allen, J., Murray, A., Davidson, A. W., McKeith, I. G., & Kenny, R. A. (2007). Autonomic dysfunction in dementia. *J Neurol Neurosurg Psychiatry, 78*(7), 671-677. doi:10.1136/jnnp.2006.102343

Almkvist, O. (1996). Neuropsychological features of early Alzheimer's disease: preclinical and clinical stages. *Acta Neurol Scand*, 63-71.

Alzheimer's Research UK. (2022). Can music help someone living with dementia? Retrieved from https://www.facebook.com/AlzheimersResearchUK/videos

Alzheimer's Society. (n.d.). Singing for the Brain. Retrieved from https://www.alzheimers.org.uk/get-support/

Amendment 114: Health and Care Bill - Report (3rd Day) – in the House of Lords at 4:23 pm on 7th March. (2022). Retrieved from https://www.theyworkforyou.com

Au-Yeung, W. M., Miller, L., Beattie, Z., May, R., Cray, H. V., Kabelac, Z., . . . Vahia, I. V. (2022). Monitoring Behaviors of Patients With Late-Stage Dementia Using Passive Environmental Sensing Approaches: A Case Series. *Am J Geriatr Psychiatry, 30*(1), 1-11. doi:10.1016/j.jagp.2021.04.008

Baird, A., & Samson, S. (2009). Memory for music in Alzheimer's disease: unforgettable? *Neuropsychol Rev, 19*(1), 85-101. doi:10.1007/s11065-009-9085-2

Baird, A., & Samson, S. (2015). Music and dementia. *Prog Brain Res, 217*, 207-235. doi:10.1016/bs.pbr.2014.11.028

Baird, A., Umbach, H., & Thompson, W. F. (2017). A nonmusician with severe Alzheimer's dementia learns a new song. *Neurocase, 23*(1), 36-40. doi:10.1080/13554794.2017.1287278

Baker, F. A., Bloska, J., Braat, S., Bukowska, A., Clark, I., Hsu, M. H., . . . Odell-Miller, H. (2019). HOMESIDE: home-based family caregiver-delivered music and reading interventions for people living with dementia: protocol of a randomised controlled trial. *BMJ Open, 9*(11), e031332. doi:10.1136/bmjopen-2019-031332

Baker, F. A., Lee, Y. C., Sousa, T. V., Stretton-Smith, P. A., Tamplin, J., Sveinsdottir, V., . . . Gold, C. (2022). Clinical effectiveness of music interventions for dementia and depression in elderly care (MIDDEL): Australian cohort of an international pragmatic

cluster-randomised controlled trial. *Lancet Healthy Longev, 3*(3), e153-e165. doi:10.1016/S2666-7568(22)00027-7

Balconi, M., Cotelli, M., Brambilla, M., Manenti, R., Cosseddu, M., Premi, E., . . . Borroni, B. (2015). Understanding Emotions in Frontotemporal Dementia: The Explicit and Implicit Emotional Cue Mismatch. *Journal of Alzheimer's Disease, 46*, 211-225. doi:10.3233/JAD-142826

Bannan, N., & Montgomery-Smith, C. (2008). 'Singing for the brain': reflections on the human capacity for music arising from a pilot study of group singing with Alzheimer's patients. *J R Soc Promot Health, 128*(2), 73-78. doi:10.1177/1466424007087807

Barradas, G. T., Juslin, P. N., & Bermúdez i Badia, S. (2021). Emotional Reactions to Music in Dementia Patients and Healthy Controls: Differential Responding Depends on the Mechanism. *Music & Science, 4*. doi:10.1177/20592043211010152

Barrett, L. F. (2014). The Conceptual Act Theory: A Précis. *Emotion Review, 6*(4), 292-297. doi:10.1177/1754073914534479

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of psychophysiology, 2nd ed.* (pp. 142-162). New York, NY, US: Cambridge University Press.

Beatty, W. W., Rogers, C. L., Rogers, R. L., English, S., Testa, J. A., Orbelo, D. M., . . . Ross, E. D. (1999). Piano playing in Alzheimer's disease: Longitudinal study of a single case. *Neurocase, 5*(5), 459-469. doi:10.1080/13554799908402740

Beatty, W. W., Zavadil, K. D., Bailly, R. C., Rixen, G. J., Zavadil, L. E., Farnham, N., & Fisher, L. (1988). Preserved musical skill in a severely demented patient. *International Journal of Clinical Neuropsychology, 10*, 158-164.

Beissner, F., Meissner, K., Bar, K. J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. *J Neurosci, 33*(25), 10503-10511. doi:10.1523/JNEUROSCI.1103-13.2013

Benhamou, E., & Warren, J. D. (2020). Disorders of music processing in dementia. In *Music and the Aging Brain* (pp. 107-149).

Benhamou, E., Zhao, S., Sivasathiaseelan, H., Johnson, J. C. S., Requena-Komuro, M. C., Bond, R. L., . . . Warren, J. D. (2021). Decoding expectation and surprise in dementia: the paradigm of music. *Brain Commun, 3*(3), fcab173. doi:10.1093/braincomms/fcab173

Bianco, R., Gold, B. P., Johnson, A. P., & Penhune, V. B. (2019). Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians. *Sci Rep, 9*(1), 17060. doi:10.1038/s41598-019-53510-w

Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain Cogn, 138*, 103621. doi:10.1016/j.bandc.2019.103621

Bittner, D. M., Wieseler, I., Wilhelm, H., Riepe, M. W., & Muller, N. G. (2014). Repetitive pupil light reflex: potential marker in Alzheimer's disease? *J Alzheimers Dis, 42*(4), 1469-1477. doi:10.3233/JAD-140969

Bolton, L. M., Jiang, J., & Warren, J. D. (2022). Music as a person centred intervention for dementia. *BMJ, 376*, o518. doi:10.1136/bmj.o518

Bonifacio, G., & Zamboni, G. (2016). Brain imaging in dementia. *Postgrad Med J, 92*(1088), 333-340. doi:10.1136/postgradmedj-2015-133759

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *J Pers Soc Psychol, 105*(4), 703-717. doi:10.1037/a0033770

Boucsein, W. (2012). *Electrodermal Activity* (2 ed.): Springer New York, NY.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., roth, W. T., Dawson, M. E., . . . Society for Psychophysiological Research Ad Hoc Committee on Electrodermal, M. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*(8), 1017-1034. doi:10.1111/j.1469-8986.2012.01384.x

Bourne, P., Camic, P. M., & Crutch, S. J. (2021). Psychosocial outcomes of dyadic arts interventions for people with a dementia and their informal caregivers: A systematic review. *Health Soc Care Community, 29*(6), 1632-1649. doi:10.1111/hsc.13267

Bourne, P., Camic, P. M., Crutch, S. J., Hulbert, S., Firth, N. C., & Harding, E. (2019). Using Psychological and Physiological Measures in Arts-Based Activities in a Community Sample of People with a Dementia and Their Caregivers: A Feasibility and Pilot Study. *Journal of Aging Studies and Therapies, 1*(1), 1-11. doi:10.16966/jast.102

Bowen, M., Edgar, D. F., Hancock, B., Haque, S., Shah, R., Buchanan, S., . . . O'Leary, N. (2016). The Prevalence of Visual Impairment in People with Dementia (the PrOVIDe study): a cross-sectional study of people aged 60-89 years with dementia and qualitative exploration of individual, carer and professional perspectives. *Health Services and Delivery Research, 4*(21). doi:10.3310/hsdr04210

Bowling, D. L., Gahr, J., Ancochea, P. G., Hoeschele, M., Canoine, V., Fusani, L., & Fitch, W. T. (2022). Endogenous oxytocin, cortisol, and testosterone in response to group singing. *Horm Behav, 139*, 105105. doi:10.1016/j.yhbeh.2021.105105

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602-607. doi:10.1111/j.1469-8986.2008.00654.x

Brancatisano, O., Baird, A., & Thompson, W. F. (2020). Why is music therapeutic for neurological disorders? The Therapeutic Music Capacities Model. *Neurosci Biobehav Rev, 112*, 600-615. doi:10.1016/j.neubiorev.2020.02.008

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32. doi:https://doi.org/10.1023/A:1010933404324

Brotherhood, E., Ball, P., Camic, P. M., Evans, C., Fox, N., Murphy, C., . . . Crutch, S. (2017). Preparatory planning framework for Created Out of Mind: Shaping perceptions of dementia through art and science. *Wellcome Open Res, 2*, 108. doi:10.12688/wellcomeopenres.12773.1

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x

Bufo, M. R., Guidotti, M., Mofid, Y., Malvy, J., Bonnet-Brilhault, F., Aguillon-Hernandez, N., & Wardak, C. (2022). Atypical Response to Affective Touch in Children with Autism: Multi-Parametric Exploration of the Autonomic System. *J Clin Med, 11*(23). doi:10.3390/jcm11237146

Burns, A., & Morris, S. (2018). Music and dementia: a powerful connector. Retrieved from https://www.england.nhs.uk/blog/music-and-dementia-a-powerful-connector/

Buss, K. A., Jaffee, S., Wadsworth, M. E., & Kliewer, W. (2018). Impact of psychophysiological stress-response systems on psychological development: Moving beyond the single biomarker approach. *Dev Psychol, 54*(9), 1601-1605. doi:10.1037/dev0000596

Camic, P. M. (2020). Canterbury Wellbeing Scales: directions and scales. *Wellcome Research Open*. doi: https://doi.org/10.5281/zenodo.4063768

Camic, P. M., Crutch, S. J., Murphy, C., Firth, N. C., Harding, E., Harrison, C. R., . . . Zeilig, H. (2018). Conceptualising and Understanding Artistic Creativity in the Dementias: Interdisciplinary Approaches to Research and Practise. *Front Psychol, 9*, 1842. doi:10.3389/fpsyg.2018.01842

Camic, P. M., Williams, C. M., & Meeten, F. (2013). Does a 'Singing Together Group' improve the quality of life of people with a dementia and their carers? A pilot evaluation study. *Dementia (London), 12*(2), 157-176. doi:10.1177/1471301211422761

Carrasco, R., Baker, F. A., Bukowska, A. A., Clark, I. N., Flynn, L. M., McMahon, K., . . . Wosch, T. (2020). *Empowering Caregivers of People Living with Dementia to Use Music Therapeutically at Home*. Paper presented at the 32nd Australian Conference on Human-Computer Interaction.

Chatterjee, H. J., Camic, P. M., Lockyer, B., & Thomson, L. J. M. (2018). Non-clinical community interventions: a systematised review of social prescribing schemes. *Arts & Health, 10*(2), 97-123. doi:10.1080/17533015.2017.1334002

Chiari, A., Vinceti, G., Adani, G., Tondelli, M., Galli, C., Fiondella, L., . . . Vinceti, M. (2021). Epidemiology of early onset dementia and its clinical presentations in the province of Modena, Italy. *Alzheimers Dement, 17*(1), 81-88. doi:10.1002/alz.12177

Christensen, R. H. B. (2022). ordinal—Regression Models for Ordinal Data; R package version 2022.11-16. Retrieved from https://CRAN.R-project.org/package=ordinal

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2016). The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience. *Organizational Research Methods, 22*(1), 394-420. doi:10.1177/1094428116681073

Chu, H., Yang, C. Y., Lin, Y., Ou, K. L., Lee, T. Y., O'Brien, A. P., & Chou, K. R. (2014). The impact of group music therapy on depression and cognition in elderly persons with dementia: a randomized controlled study. *Biol Res Nurs, 16*(2), 209-217. doi:10.1177/1099800413485410

Clare, A., & Camic, P. M. (2020). Live and recorded group music interventions with active participation for people with dementias: a systematic review. *Arts Health, 12*(3), 197-220. doi:10.1080/17533015.2019.1675732

Clare, A., Camic, P. M., Crutch, S. J., West, J., Harding, E., & Brotherhood, E. (2020). Using Music to Develop a Multisensory Communicative Environment for People with Late-Stage Dementia. *Gerontologist, 60*(6), 1115-1125. doi:10.1093/geront/gnz169

Clark, C. N., Golden, H. L., McCallion, O., Nicholas, J. M., Cohen, M. H., Slattery, C. F., . . . Warren, J. D. (2018). Music models aberrant rule decoding and reward valuation in dementia. *Soc Cogn Affect Neurosci, 13*(2), 192-202. doi:10.1093/scan/nsx140

Clark, C. N., & Warren, J. D. (2015). Music, memory and mechanisms in Alzheimer's disease. *Brain, 138*(Pt 8), 2120-2122. doi:10.1093/brain/awv164

Clark, I. N., Tamplin, J. D., & Baker, F. A. (2018). Community-Dwelling People Living With Dementia and Their Family Caregivers Experience Enhanced Relationships and Feelings of Well-Being Following Therapeutic Group Singing: A Qualitative Thematic Analysis. *Front Psychol, 9*, 1332. doi:10.3389/fpsyg.2018.01332

Claron, J., Royo, J., Arcizet, F., Deffieux, T., Tanter, M., & Pouget, P. (2022). Covariations between pupil diameter and supplementary eye field activity suggest a role in cognitive effort implementation. *PLoS Biol, 20*(5), e3001654. doi:10.1371/journal.pbio.3001654

Clements-Cortés, A. (2015). Singing for Health, Connection and Care. *Music and Medicine, 7*(4), 13-23. doi:10.47513/mmd.v7i4.428

Clift, S., Hancox, G., Staricoff, R., & Whitmore, C. (2008). *Singing and Health: A systematic mapping and review of non-clinical research*. Retrieved from Canterbury Canterbury Christ Church University:

Cohen, D., & Rossato-Benett, M. (2014). Alive Inside: A Story of Music and Memory.

Corrêa, L., Caparrol, A. J. d. S., Martins, G., Pavarini, S. C. I., & Gratão, A. C. M. (2020). Efeitos da música nas expressões corporais e faciais e nos sintomas psicológicos e comportamentais de idosos. *Cadernos Brasileiros de Terapia Ocupacional, 28*(2), 539-553. doi:10.4322/2526-8910.ctoAO1889

Cousins, E., Tischler, V., Garabedian, C., & Dening, T. (2020). A Taxonomy of Arts Interventions for People With Dementia. *Gerontologist, 60*(1), 124-134. doi:10.1093/geront/gnz024

Critchley, H. (2002). Electrodermal responses: What happens in the brain. *The Neuroscientist, 8*(2), 84-185.

Crous-Bou, M., Minguillon, C., Gramunt, N., & Molinuevo, J. L. (2017). Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimers Res Ther, 9*(1), 71. doi:10.1186/s13195-017-0297-z

Crutch, S. J., Lehmann, M., Schott, J. M., Rabinovici, G. D., Rossor, M. N., & Fox, N. C. (2012). Posterior cortical atrophy. *Lancet Neurol, 11*(2), 170-178. doi:10.1016/S1474-4422(11)70289-7

Crutch, S. J., Schott, J. M., Rabinovici, G. D., Murray, M., Snowden, J. S., van der Flier, W. M., . . . Associated Syndromes Professional Interest, A. (2017). Consensus classification of posterior cortical atrophy. *Alzheimers Dement, 13*(8), 870-884. doi:10.1016/j.jalz.2017.01.014

Cuddy, L. L., & Duffin, J. M. (2005). Music, memory, and Alzheimer's disease: is music recognition spared in dementia, and how can it be assessed? *Med Hypotheses, 64*(2), 229-235. doi:10.1016/j.mehy.2004.09.005

Cuddy, L. L., Duffin, J. M., Gill, S. S., Brown, C. L., Sikka, R., & Vanstone, A. D. (2012). Memory for Melodies and Lyrics in Alzheimer's Disease. *Music Perception, 29*(5), 479-491. doi:10.1525/mp.2012.29.5.479

Cuddy, L. L., Sikka, R., & Vanstone, A. (2015). Preservation of musical memory and engagement in healthy aging and Alzheimer's disease. *Ann N Y Acad Sci, 1337*, 223-231. doi:10.1111/nyas.12617

Cummings, A. (2015). A brief evaluation of Vocal Flourish, a York choir for people with dementia.

da Rocha, L. A., Siqueira, B. F., Grella, C. E., & Gratão, A. C. M. (2022). Effects of concert music on cognitive, physiological, and psychological parameters in the elderly with dementia: a quasi-experimental study. *Dement Neuropsychol, 16*(2), 194-201. doi:10.1590/1980-5764-dn-2021-0088

David, M., & Malhotra, P. A. (2022). New approaches for the quantification and targeting of noradrenergic dysfunction in Alzheimer's disease. *Ann Clin Transl Neurol, 9*(4), 582-596. doi:10.1002/acn3.51539

Daykin, N., Mansfield, L., Meads, C., Julier, G., Tomlinson, A., Payne, A., . . . Victor, C. (2018). What works for wellbeing? A systematic review of wellbeing outcomes for music and singing in adults. *Perspect Public Health, 138*(1), 39-46. doi:10.1177/1757913917740391

de Fleurian, R., & Pearce, M. T. (2021). Chills in music: A systematic review. *Psychological Bulletin, 147*(9), 890-920. doi:10.1037/bul0000341

Delius, J. A. M., & Müller, V. (2022). Interpersonal synchrony when singing in a choir. *Front Psychol, 13*, 1087517. doi:10.3389/fpsyg.2022.1087517

Dementia Choir format seals international deals. (n.d.). Retrieved from https://curvemedia.com/dementia-choir-format-seals-international-deals/

DeNora, T., & Ansdell, G. (2014). What can't music do? *Psychological Well-Being, 4*(23). doi:https://doi.org/10.1186/s13612-014-0023-6

Dingle, G. A., Sharman, L. S., Bauer, Z., Beckman, E., Broughton, M., Bunzli, E., . . . Wright, O. R. L. (2021). How Do Music Activities Affect Health and Well-Being? A Scoping Review of Studies Examining Psychosocial Mechanisms. *Front Psychol, 12*, 713818. doi:10.3389/fpsyg.2021.713818

DiNuzzo, M., Mascali, D., Moraschi, M., Bussu, G., Maugeri, L., Mangini, F., . . . Giove, F. (2019). Brain Networks Underlying Eye's Pupil Dynamics. *Front Neurosci, 13*, 965. doi:10.3389/fnins.2019.00965

Dougherty, P. (2020). Central Control of the Autonomic Nervous System and Thermoregulation. Retrieved from https://nba.uth.tmc.edu/neuroscience

Dowlen, R., Keady, J., Milligan, C., Swarbrick, C., Ponsillo, N., Geddes, L., & Riley, B. (2021). In the moment with music: an exploration of the embodied and sensory experiences of people living with dementia during improvised music-making. *Ageing and Society, 42*(11), 2642-2664. doi:10.1017/s0144686x21000210

Dowson, B., McDermott, O., & Schneider, J. (2019). What indicators have been used to evaluate the impact of music on the health and wellbeing of people with dementia? A review using meta-narrative methods. *Maturitas, 127*, 26-34. doi:10.1016/j.maturitas.2019.06.001

Dowson, B., Schneider, J., Guo, B., Bath, P. M., McDermott, O., Haywood, L. J., & Orrell, M. (2021). Pilot Randomised Evaluation of Singing in Dementia (PRESIDE): protocol for a two-arm, parallel-group randomised controlled feasibility study with waiting-list control. *Pilot Feasibility Stud, 7*(1), 15. doi:10.1186/s40814-020-00759-y

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., . . . Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol, 13*(6), 614-629. doi:10.1016/S1474-4422(14)70090-0

Edwards, R. D., & Hodges, D. A. (2007). Neuromusical research: An overview of the literature. In W. G. a. F. Rauscher (Ed.), *Neurosciences in Music Pedagogy* (pp. 1-25): Nova Science Publishers, Inc.

Eerola, T. (2011). Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical, Film, Pop and Mixed Genres. *Journal of New Music Research, 40*(4), 349-366. doi:10.1080/09298215.2011.602195

Eerola, T., & Peltola, H. R. (2016). Memorable Experiences with Sad Music-Reasons, Reactions and Mechanisms of Three Types of Experiences. *PLoS One, 11*(6), e0157444. doi:10.1371/journal.pone.0157444

Eerola, T., Peltola, H. R., & Vuoskoski, J. K. (2015). Attitudes toward sad music are related to both preferential and contextual strategies. *Psychomusicology: Music, Mind, and Brain, 25*(2), 116-123. doi:10.1037/pmu0000096

Efron, R. (1968). What is perception? Boston studies in the philosophy of science. *Humanities: New York*.

Eisenmann, Y., Golla, H., Schmidt, H., Voltz, R., & Perrar, K. M. (2020). Palliative Care in Advanced Dementia. *Front Psychiatry, 11*, 699. doi:10.3389/fpsyt.2020.00699

Ekman, P. (1999). Basic emotions. In T. D. M. Power (Ed.), *Handbook of Cognition and Emotion*: John Wiley & Sons Ltd. .

El Haj, M., Chapelet, G., Moustafa, A. A., & Boutoleau-Bretonniere, C. (2022). Pupil size as an indicator of cognitive activity in mild Alzheimer's disease. *EXCLI J, 21*, 307-316. doi:10.17179/excli2021-4568

Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digit Biomark, 5*(1), 78-88. doi:10.1159/000515346

Fancourt, D., & Finn, S. (2019). *What is the evidence on the role of the arts in improving health and well-being? A scoping review*. Retrieved from Health Evidence Network synthesis report: https://www.ncbi.nlm.nih.gov/books/NBK553773/

Fancourt, D., Ockelford, A., & Belai, A. (2014). The psychoneuroimmunological effects of music: a systematic review and a new model. *Brain Behav Immun, 36*, 15-26. doi:10.1016/j.bbi.2013.10.014

Felisberti, F. M. (2021). Hedonic Preferences to Audio and Visual Stimulation in Seniors with Cognitive Impairments. *J Alzheimers Dis, 83*(3), 1353-1366. doi:10.3233/JAD-210520

Femminella, G. D., Rengo, G., Komici, K., Iacotucci, P., Petraglia, L., Pagano, G., . . . Ferrara, N. (2014). Autonomic dysfunction in Alzheimer's disease: tools for assessment and review of the literature. *J Alzheimers Dis, 42*(2), 369-377. doi:10.3233/JAD-140513

Fernández, G., & Parra, M. A. (2021). Oculomotor behaviors and integrative memory functions in the Alzheimer clinical syndrome. *Journal of Alzheimer's Disease, 82*(3), 1033-1044. doi:10.3233/JAD-201189

Fink, L. K., Hurley, B. K., Geng, J. J., & Janata, P. (2018). A linear oscillator model predicts dynamic temporal attention and pupillary entrainment to rhythmic patterns. *J Eye Mov Res, 11*(2). doi:10.16910/jemr.11.2.12

Fletcher, P. D., Downey, L. E., Golden, H. L., Clark, C. N., Slattery, C. F., Paterson, R. W., . . . Warren, J. D. (2015). Auditory hedonic phenotypes in dementia: A behavioural and neuroanatomical analysis. *Cortex, 67*, 95-105. doi:10.1016/j.cortex.2015.03.021

Fletcher, P. D., Nicholas, J. M., Downey, L. E., Golden, H. L., Clark, C. N., Pires, C., . . . Warren, J. D. (2016). A physiological signature of sound meaning in dementia. *Cortex, 77*, 13-23. doi:10.1016/j.cortex.2016.01.007

Fletcher, P. D., Nicholas, J. M., Shakespeare, T. J., Downey, L. E., Golden, H. L., Agustus, J. L., . . . Warren, J. D. (2015a). Dementias show differential physiological responses to salient sounds. *Front Behav Neurosci, 9*, 73. doi:10.3389/fnbeh.2015.00073

Fletcher, P. D., Nicholas, J. M., Shakespeare, T. J., Downey, L. E., Golden, H. L., Agustus, J. L., . . . Warren, J. D. (2015b). Physiological phenotyping of dementias using emotional sounds. *Alzheimers Dement (Amst), 1*(2), 170-178. doi:10.1016/j.dadm.2015.02.003

Fong, J. (2012). Eye-Opener: Why Do Pupils Dilate in Response to Emotional States? Retrieved from https://www.scientificamerican.com

Fotiou, D., Kaltsatou, A., Tsiptsios, D., & Nakou, M. (2015). Evaluation of the cholinergic hypothesis in Alzheimer's disease with neuropsychological methods. *Aging Clin Exp Res, 27*(5), 727-733. doi:10.1007/s40520-015-0321-8

Franzen, L., Cabugao, A., Grohmann, B., Elalouf, K., & Johnson, A. P. (2022). Individual pupil size changes as a robust indicator of cognitive familiarity differences. *PLoS One, 17*(1), e0262753. doi:10.1371/journal.pone.0262753

Freitas, C., Manzato, E., Burini, A., Taylor, M. J., Lerch, J. P., & Anagnostou, E. (2018). Neural Correlates of Familiarity in Music Listening: A Systematic Review and a Neuroimaging Meta-Analysis. *Front Neurosci, 12*, 686. doi:10.3389/fnins.2018.00686

Frost, S., Robinson, L., Rowe, C. C., Ames, D., Masters, C. L., Taddei, K., . . . Kanagasingam, Y. (2017). Evaluation of Cholinergic Deficiency in Preclinical Alzheimer's Disease Using Pupillometry. *J Ophthalmol, 2017*, 7935406. doi:10.1155/2017/7935406

Gardner, H. (1985). *Frames of Mind; The theory of multiple intelligences*. New York: Basic.

Gardner, R. M., Mo, S. S., & Borrego, R. (1974). Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory. *Bulletin of the Psychonomic Society, 3*(3), 237-238. doi:10.3758/BF03333458

Gardner, R. M., Mo, S. S., & Krinsky, R. (1974). Inhibition of pupillary orienting reflex by heteromodal novelty. *Bulletin of the Psychonomic Society, 4*, 510-512. doi:10.3758/BF03334274

Garland, K., Beer, E., Eppingstall, B., & O'Connor, D. W. (2007). A comparison of two treatments of agitated behavior in nursing home residents with dementia: simulated

family presence and preferred music. *Am J Geriatr Psychiatry, 15*(6), 514-521. doi:10.1097/01.JGP.0000249388.37080.b4

Garrido, S., Dunne, L., Stevens, C., Chang, E., & Perz, J. (2019). *Music playlists for people with dementia: A guide for carers, health workers and family*. Retrieved from https://musicfordementia.github.io/

Garrido, S., Stevens, C. J., Chang, E., Dunne, L., & Perz, J. (2018). Music and Dementia: Individual Differences in Response to Personalized Playlists. *J Alzheimers Dis, 64*(3), 933-941. doi:10.3233/jad-180084

Garrido, S., Stevens, C. J., Chang, E., Dunne, L., & Perz, J. (2019). Musical Features and Affective Responses to Personalized Playlists in People With Probable Dementia. *Am J Alzheimers Dis Other Demen, 34*(4), 247-253. doi:10.1177/1533317518808011

Gaudreau, D., & Peretz, I. (1999). Implicit and explicit memory for music in old and young adults. Netherlands: Elsevier Science.

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behav Res Methods, 52*(5), 2232-2255. doi:10.3758/s13428-020-01374-8

Gerdner, L. A. (2012). Individualized music for dementia: Evolution and application of evidence-based protocol. *World J Psychiatry, 2*(2), 26-32. doi:10.5498/wjp.v2.i2.26

Gerdner, L. A. (2021). Evidence-Based Guideline: Individualized Music for Persons with Dementia (7th Edition, 2021). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10798453

Gerdner, L. A., Hartsock, J., & Buckwalter, K. C. (2000). Assessment of personal music preference (Family version)

Gingras, B., Marin, M. M., Puig-Waldmuller, E., & Fitch, W. T. (2015). The Eye is Listening: Music-Induced Arousal and Individual Differences Predict Pupillary Responses. *Front Hum Neurosci, 9*, 619. doi:10.3389/fnhum.2015.00619

Global Council on Brain Health. (2020). Music on Our Minds: The Rich Potential of Music to Promote Brain Health and Mental Well-Being. doi:https://doi.org/10.26419/pia.00103.001

Gök Ugur, H., Orak, O. S., Yaman Aktas, Y., Enginyurt, O., & Saglambilen, O. (2019). Effects of Music Therapy on the Care Burden of In-Home Caregivers and Physiological Parameters of Their In-Home Dementia Patients: A Randomized Controlled Trial. *Complement Med Res, 26*(1), 22-30. doi:10.1159/000490348

Gold, C., Eickholt, J., Assmus, J., Stige, B., Wake, J. D., Baker, F. A., . . . Geretsegger, M. (2019). Music Interventions for Dementia and Depression in ELderly care (MIDDEL): protocol and statistical analysis plan for a multinational cluster-randomised trial. *BMJ Open, 9*(3), e023436. doi:10.1136/bmjopen-2018-023436

Gold, C. A., & Budson, A. E. (2008). Memory loss in Alzheimer's disease: implications for development of therapeutics. *Expert Rev Neurother, 8*(12), 1879-1891. doi:10.1586/14737175.8.12.1879

Golden, H. L., Clark, C. N., Nicholas, J. M., Cohen, M. H., Slattery, C. F., Paterson, R. W., . . . Warren, J. D. (2017). Music Perception in Dementia. *J Alzheimers Dis, 55*(3), 933-949. doi:10.3233/JAD-160359

Goll, J. C., Kim, L. G., Ridgway, G. R., Hailstone, J. C., Lehmann, M., Buckley, A. H., . . . Warren, J. D. (2012). Impairments of auditory scene analysis in Alzheimer's disease. *Brain, 135*(Pt 1), 190-200. doi:10.1093/brain/awr260

Goodwin, J. (2007). Cranial Nerves III, IV, and VI: The Oculomotor System.

Granholm, E. L., Panizzon, M. S., Elman, J. A., Jak, A. J., Hauger, R. L., Bondi, M. W., . . . Kremen, W. S. (2017). Pupillary Responses as a Biomarker of Early Risk for Alzheimer's Disease. *J Alzheimers Dis, 56*(4), 1419-1428. doi:10.3233/JAD-161078

Granot, R., Spitz, D. H., Cherki, B. R., Loui, P., Timmers, R., Schaefer, R. S., . . . Israel, S. (2021). "Help! I Need Somebody": Music as a Global Resource for Obtaining Wellbeing Goals in Times of Crisis. *Front Psychol, 12*, 648013. doi:10.3389/fpsyg.2021.648013

Greengross, S. (2020). Making the case for music becoming mainstream in dementia care. Retrieved from https://musicfordementia.org.uk

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25* (7). doi:https://doi.org/10.1016/S0166-2236(02)02191-4

Groarke, J. M., MacCormac, N., McKenna-Plumley, P. E., & Graham-Wisener, L. (2022). Music Listening Was an Emotional Resource and Social Surrogate for Older Adults During the COVID-19 Pandemic: A Qualitative Study. *Behaviour Change, 39*(3), 168-179. doi:10.1017/bec.2022.10

Groussard, M., Tyler, G., Copalle, R., & Platel, H. (2019). Preservation of musical memory throughout the progression of Alzheimer's Disease? Toward a reconciliation of theoretical, clinical and neuroimaging evidence. *Journal of Alzheimer's Disease, 68*(3), 857-883. doi:10.3233/JAD-180474

Grube, M., Bruffaerts, R., Schaeverbeke, J., Neyens, V., De Weer, A. S., Seghers, A., . . . Vandenberghe, R. (2016). Core auditory processing deficits in primary progressive aphasia. *Brain, 139*(Pt 6), 1817-1829. doi:10.1093/brain/aww067

Gupta, P., & Srivastava, S. (2017). Use of Pupillometry in Dementia: A Review of Published Literature.

Hackett, K., Sabat, S. R., & Giovannetti, T. (2022). A person-centered framework for designing music-based therapeutic studies in dementia: current barriers and a path forward. *Aging Ment Health, 26*(5), 940-949. doi:10.1080/13607863.2021.1931029

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews, 18*(2), e1230. doi:https://doi.org/10.1002/cl2.1230

Halpern, A. R., & O'Connor, M. G. (2000). Implicit memory for music in Alzheimer's disease. *Neuropsychology, 14*(3), 391-397. doi:10.1037//0894-4105.14.3.391

Harding, E., Sullivan, M. P., & Crutch, S. (2017). *"Engaged or exasperated? Interpreting physiological data in dementia research,".* Paper presented at the Proceedings of the Powerful Partners: Advancing Dementia Care through the Arts and Sciences Conference London: Royal Society for Public Health.

Hardy, C. J. D., Agustus, J. L., Marshall, C. R., Clark, C. N., Russell, L. L., Brotherhood, E. V., . . . Warren, J. D. (2017). Functional neuroanatomy of speech signal decoding in primary progressive aphasias. *Neurobiol Aging, 56*, 190-201. doi:10.1016/j.neurobiolaging.2017.04.026

Hardy, C. J. D., Hwang, Y. T., Bond, R. L., Marshall, C. R., Ridha, B. H., Crutch, S. J., . . . Warren, J. D. (2017). Donepezil enhances understanding of degraded speech in Alzheimer's disease. *Ann Clin Transl Neurol, 4*(11), 835-840. doi:10.1002/acn3.471

Hardy, C. J. D., Marshall, C. R., Golden, H. L., Clark, C. N., Mummery, C. J., Griffiths, T. D., . . . Warren, J. D. (2016). Hearing and dementia. *J Neurol, 263*(11), 2339-2354. doi:10.1007/s00415-016-8208-y

Hardy, C. J. D., Yong, K. X. X., Goll, J. C., Crutch, S. J., & Warren, J. D. (2020). Impairments of auditory scene analysis in posterior cortical atrophy. *Brain, 143*(9), 2689-2695. doi:10.1093/brain/awaa221

Harmer, B. J., & Orrell, M. (2008). What is meaningful activity for people with dementia living in care homes? A comparison of the views of older people with dementia, staff and family carers. *Aging Ment Health, 12*(5), 548-558. doi:10.1080/13607860802343019

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behav Res Methods, 48*(2), 510-527. doi:10.3758/s13428-015-0588-x

Hillebrand, M. C., Weise, L., & Wilz, G. (2023). Immediate Effects of Individualized Music Listening on Behavioral and Psychological Symptoms of Dementia: A Randomized Controlled Trial. *Int J Geriatr Psychiatry*, e5893. doi:10.1002/gps.5893

Hinrichs-Krapels, S., & Grant, J. (2016). Exploring the effectiveness, efficiency and equity (3e's) of research and research impact assessment. *Palgrave Communications, 2*(1). doi:10.1057/palcomms.2016.90

Hlavnička, J., Tykalová, T., Ulmanová, O., Dušek, P., Horáková, D., Ruzicka, E., . . . Rusz, J. (2020). Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses. *Clin Neurophysiol, 131*(5), 1155-1165. doi:10.1016/j.clinph.2020.02.005

Hodges, D. A. (2010). Psychophysiological measures. In P. N. Juslin (Ed.), *Handbook of Music and Emotion: Theory, Research, Applications*.

Hodges, D. A., & Wilkins, R. W. (2015). How and Why Does Music Move Us? *Music Educators Journal, 101*(4), 41-47. doi:10.1177/0027432115575755

Holcombe, A. O. (2019). Contributorship, Not Authorship: Use CRediT to Indicate Who Did What. *Publications, 7*(3). doi:10.3390/publications7030048

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. *Front Psychol, 11*, 721. doi:10.3389/fpsyg.2020.00721

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Assessing the Fit of the Model. In S. L. a. R. X. S. D.W. Hosmer (Ed.), *Applied Logistic Regression*.

Howland, R. H. (2016). Hey mister tambourine man, play a drug for me: Music as medication. *Journal of Psychosocial Nursing & Mental Health Services, 54*(12), 23-27. doi:https://doi.org/10.3928/02793695-20161208-05

Hsieh, S., Hornberger, M., Piguet, O., & Hodges, J. R. (2011). Neural basis of music knowledge: evidence from the dementias. *Brain, 134*(Pt 9), 2523-2534. doi:10.1093/brain/awr190

Hsu, M. H., Flowerdew, R., Parker, M., Fachner, J., & Odell-Miller, H. (2015). Individual music therapy for managing neuropsychiatric symptoms for people with dementia and their carers: a cluster randomised controlled feasibility study. *BMC Geriatr, 15*, 84. doi:10.1186/s12877-015-0082-4

Imtiaz, D., Anwar, Y., & Khan, A. (2020). Wearable sensors and a multisensory music and reminiscence therapies application: To help reduce behavioral and psychological symptoms in person with dementia. *Smart Health, 18*. doi:10.1016/j.smhl.2020.100140

Innes, K. E., Montgomery, C., Selfe, T. K., Wen, S., Khalsa, D. S., & Flick, M. (2021). Incorporating a Usual Care Comparator into a Study of Meditation and Music Listening for Older Adults with Subjective Cognitive Decline: A Randomized Feasibility Trial. *J Alzheimers Dis Rep, 5*(1), 187-206. doi:10.3233/ADR-200249

International Federation of the Phonographic Industry. (2022). Engaging with music report Retrieved from https://www.ifpi.org/resources/

Irish, M., Cunningham, C. J., Walsh, J. B., Coakley, D., Lawlor, B. A., Robertson, I. H., & Coen, R. F. (2006). Investigating the Enhancing Effect of Music on Autobiographical Memory in Mild Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders, 22*(1), 108-120. doi:10.1159/000093487

Jabbari, A., Grimnes, S., & Martinsen, O. G. (2011). Electrodermal activity by DC potential and AC conductance measured simultaneously at the same skin site. *Skin Res Technol, 17*(1), 26-34. doi:10.1111/j.1600-0846.2010.00459.x

Jacobsen, J. H., Stelzer, J., Fritz, T. H., Chetelat, G., La Joie, R., & Turner, R. (2015). Why musical memory can be preserved in advanced Alzheimer's disease. *Brain, 138*(Pt 8), 2438-2450. doi:10.1093/brain/awv135

Jagiello, R., Pomper, U., Yoneya, M., Zhao, S., & Chait, M. (2019). Rapid Brain Responses to Familiar vs. Unfamiliar Music - an EEG and Pupillometry study. *Sci Rep, 9*(1), 15570. doi:10.1038/s41598-019-51759-9

Jakubowski, K., Eerola, T., Tillmann, B., Perrin, F., & Heine, L. (2020). A Cross-Sectional Study of Reminiscence Bumps for Music-Related Memories in Adulthood. *Music & Science, 3*. doi:10.1177/2059204320965058

James, M., Plant, G. T., Warrington, E. K., & Thames Valley Test Company. (2001). Corvist: Cortical vision screening test

Janata, P. (2005). Brain networks that track musical structure. *Ann N Y Acad Sci, 1060*, 111-124. doi:10.1196/annals.1360.008

Jang, H., Soroski, T., Rizzo, M., Barral, O., Harisinghani, A., Newton-Mason, S., . . . Field, T. S. (2021). Classification of Alzheimer's Disease Leveraging Multi-task Machine Learning Analysis of Speech and Eye-Movement Data. *Front Hum Neurosci, 15*, 716670. doi:10.3389/fnhum.2021.716670

Jänig, W. (2006). Functional anatomy of the peripheral sympathetic and parasympathetic nervous system. In *Integrative Action of the Autonomic Nervous System* (pp. 13-34).

Johnson, J. C. S., Marshall, C. R., Weil, R. S., Bamiou, D. E., Hardy, C. J. D., & Warren, J. D. (2021). Hearing and dementia: from ears to brain. *Brain, 144*(2), 391-401. doi:10.1093/brain/awaa429

Johnson, J. C. S., McWhirter, L., Hardy, C. J. D., Crutch, S. J., Marshall, C. R., Mummery, C. J., . . . Warren, J. D. (2021). Suspecting dementia: canaries, chameleons and zebras. *Pract Neurol*. doi:10.1136/practneurol-2021-003019

Johnston, B., Bowman, F., Carduff, E., Donmez, F., Lowndes, A., & McKeown, A. (2022). 'Playlist for Life' at the end of life: a mixed-methods feasibility study of a personalised music listening intervention in the hospice setting. *Pilot and Feasibility Studies, 8*(1). doi:10.1186/s40814-022-00983-8

Johnston, E., Rasmusson, X., Foyil, B., Shopland, P., & Wang, S. (2017). Witnesses to transformation: Family member experiences providing individualized music to their relatives with dementia. *Cogent Education, 4*(1). doi:10.1080/2331186x.2017.1362888

Joshi, A., Mendez, M. F., Kaiser, N., Jiminez, E., Mather, M., & Shapira, J. S. (2014). Skin Conductance Levels May Reflect Emotional Blunting in Behavioral Variant Frontotemporal Dementia. *J Neuropsychiatry Clin Neurosci, 26*(3), 227-232.

Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. *Neuron, 89*(1), 221-234. doi:10.1016/j.neuron.2015.11.028

Joubert, C., Davidson, P. S. R., & Chainay, H. (2018). When Do Older Adults Show a Positivity Effect in Emotional Memory? *Exp Aging Res, 44*(5), 455-468. doi:10.1080/0361073X.2018.1521498

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys Life Rev, 10*(3), 235-266. doi:10.1016/j.plrev.2013.05.008

Juslin, P. N., Barradas, G., & Eerola, T. (2015). From Sound to Significance: Exploring the Mechanisms Underlying Emotional Reactions to Music. *Am J Psychol, 128*(3), 281-304. doi:10.5406/amerjpsyc.128.3.0281

Juslin, P. N., Harmat, L., & Eerola, T. (2013). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music, 42*(4), 599-623. doi:10.1177/0305735613484548

Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research, 33*(3), 217-238. doi:10.1080/0929821042000317813

Juslin, P. N., Liljeström, S., Laukka, P., Västfjäll, D., & Lundqvist, L. (2011). Emotional reactions to music in a nationally representative sample of Swedish adults. *Musicae Scientiae, 15*(2), 174-207. doi:10.1177/1029864911401169

Keady, J. D., Campbell, S., Clark, A., Dowlen, R., Elvish, R., Jones, L., . . . Williams, S. (2020). Re-thinking and re-positioning 'being in the moment' within a continuum of moments: introducing a new conceptual framework for dementia studies. *Ageing and Society, 42*(3), 681-702. doi:10.1017/s0144686x20001014

Keady, J. D., & Gilliard, J. (2002). Testing times: the experience of neuropsychological assessment for people with suspected Alzheimer's disease. *The person with Alzheimer's disease: Pathways to understanding the experience*, 3-28.

Kontos, P., & Grigorovich, A. (2018). Rethinking musicality in dementia as embodied and relational. *J Aging Stud, 45*, 39-48. doi:10.1016/j.jaging.2018.01.006

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2007). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music, 36*(1), 101-126. doi:10.1177/0305735607082623

Kurita, A., Takase, B., Okada, K., Horiguchi, Y., Abe, S., Kusama, Y., & Atarasi, H. (2006). Effects of Music Therapy on Heart Rate Variability in Elderly Patients with Cerebral Vascular Disease and Dementia. *Journal of Arrhythmia, 22*(3), 161-166. doi:10.1016/s1880-4276(06)80014-1

Kwon, N. H., S.; Blaylock, n.; O'Connell, H.; Hacen, N.; Gwin, J. (2022, 15 September 2022). *Detecting Anxiety and Depression from Phone Conversations using x-vectors.* Paper presented at the Workshop on Speech, Music and Mind, Virtual.

Lai, D. L. L., Crutch, S. J., West, J., Harding, E., Brotherhood, E. V., Takhar, R., . . . Camic, P. M. (2020). Development of the Video Analysis Scale of Engagement (VASE) for people with advanced dementia. *Wellcome Open Res, 5*, 230. doi:10.12688/wellcomeopenres.16189.3

Lai, H. (2004). Music preference and relaxation in Taiwanese elderly people. *Geriatric Nursing, 25*(5), 286-291. doi:https://doi.org/10.1016/j.gerinurse.2004.08.009

Lampe, L., Huppertz, H. J., Anderl-Straub, S., Albrecht, F., Ballarini, T., Bisenius, S., . . . Germany, F. C. (2023). Multiclass prediction of different dementia syndromes based on multi-centric volumetric MRI imaging. *Neuroimage Clin, 37*, 103320. doi:10.1016/j.nicl.2023.103320

Larsen, R. S., & Waters, J. (2018). Neuromodulatory Correlates of Pupil Dilation. *Front Neural Circuits, 12*, 21. doi:10.3389/fncir.2018.00021

Lartillot, O., & Toiviainen, P. (2007, September 10-15). *A Matlab toolbox for musical feature extraction from audio.* Paper presented at the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France.

Lehmann, M., Barnes, J., Ridgway, G. R., Wattam-Bell, J., Warrington, E. K., Fox, N. C., & Crutch, S. J. (2011). Basic visual function and cortical thickness patterns in posterior cortical atrophy. *Cereb Cortex, 21*(9), 2122-2132. doi:10.1093/cercor/bhq287

Lewis, A., Philip, K. E. J., Lound, A., Cave, P., Russell, J., & Hopkinson, N. S. (2021). The physiology of singing and implications for 'Singing for Lung Health' as a therapy for individuals with chronic obstructive pulmonary disease. *BMJ Open Respir Res, 8*(1). doi:10.1136/bmjresp-2021-000996

Lineweaver, T. T., Bergeson, T. R., Ladd, K., Johnson, H., Braid, D., Ott, M., . . . Brimmer, T. (2022). The Effects of Individualized Music Listening on Affective, Behavioral,

Cognitive, and Sundowning Symptoms of Dementia in Long-Term Care Residents. *Journal of Aging and Health, 34*(1), 130-143. doi:10.1177/08982643211033407

Liu, Y., Wang, Z., & Yu, G. (2021). The Effectiveness of Facial Expression Recognition in Detecting Emotional Responses to Sound Interventions in Older Adults With Dementia. *Front Psychol, 12*, 707809. doi:10.3389/fpsyg.2021.707809

Luengo-Fernandez, R., & Landeiro, F. (in preparation). The Economic Burden of Dementia in the UK. Retrieved from https://dementiastatistics.org/statistics/cost-and-projections-in-the-uk-and-globally-3/

Lundqvist, L., Carlsson, F., & Juslin, P. N. (2009). Emotional responses to music: experience, expression and physiology. *Psychology of Music, 37*(1), 61-90.

Lustig-Barzelay, Y., Sher, I., Sharvit-Ginon, I., Feldman, Y., Mrejen, M., Dallasheh, S., . . . Rotenstreich, Y. (2022). Machine learning for comprehensive prediction of high risk for Alzheimer's disease based on chromatic pupilloperimetry. *Sci Rep, 12*(1), 9945. doi:10.1038/s41598-022-13999-0

MacPherson, S., Bird, M., Anderson, K., Davis, T., & Blair, A. (2009). An art gallery access programme for people with dementia: 'you do it for the moment'. *Aging Ment Health, 13*(5), 744-752. doi:10.1080/13607860902918207

MacRitchie, J., Floridou, G. A., Christensen, J., Timmers, R., & de Witte, L. (2023). The use of technology for arts-based activities in older adults living with mild cognitive impairment or dementia: A scoping review. *Dementia (London), 22*(1), 252-280. doi:10.1177/14713012221127359

Madison, G., & Schiölde, G. (2017). Repeated Listening Increases the Liking for Music Regardless of Its Complexity: Implications for the Appreciation and Aesthetics of Music. *Front Neurosci, 11*, 147. doi:10.3389/fnins.2017.00147

MANDDOLIN4 - Music And Neuroscience against Dementia. (2022). Retrieved from https://www.ucl.ac.uk/institute-of-advanced-studies/music-futures/funded-projects/funded-projects-202223/manddolin4-music-and-neuroscience-against

Marshall, C. R., Hardy, C. J. D., Volkmer, A., Russell, L. L., Bond, R. L., Fletcher, P. D., . . . Warren, J. D. (2018). Primary progressive aphasia: a clinical approach. *J Neurol, 265*(6), 1474-1490. doi:10.1007/s00415-018-8762-6

Maseda, A., Cibeira, N., Lorenzo-Lopez, L., Gonzalez-Abraldes, I., Bujan, A., De Labra, C., & Millan-Calenti, J. C. (2018). Multisensory Stimulation and Individualized Music Sessions on Older Adults with Severe Dementia: Effects on Mood, Behavior, and Biomedical Parameters. *Journal of Alzheimer's Disease, 63(4)*, 1415-1425.

Matcham, W., Thoman, J., Sobol, A., Sanchez, E., & Gaspar, P. (2022). Definitions and Measurement of Engagement in Persons with Dementia: A Scoping Review. *Journal of Aging and Long-Term Care, 5*(2), 27-39. doi: 10.51819/jaltc.2022.1005251

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *J Cogn, 1*(1), 16. doi:10.5334/joc.18

Maw, N. N., & Pomplun, M. (2004). *Studying Human Face Recognition with the Gaze-Contingent Window Techniqu.* Paper presented at the Proceedings of the Twenty-Sixth Annual Meeting of the Cognitive Science Society, Chicago, IL.

McDermott, O. (2018). Development and evaluation of the CHORD (CHOrus Research in Dementia) manual.

McDermott, O., Crellin, N., Ridder, H. M., & Orrell, M. (2013). Music therapy in dementia: a narrative synthesis systematic review. *Int J Geriatr Psychiatry, 28*(8), 781-794. doi:10.1002/gps.3895

McDermott, O., Orrell, M., & Ridder, H. M. (2015). The development of Music in Dementia Assessment Scales (MiDAS). *Nord J Music Ther, 24*(3), 232-251. doi:10.1080/08098131.2014.907333

McIntyre, A., Harding, E., Yong, K. X. X., Sullivan, M. P., Gilhooly, M., Gilhooly, K., . . . Crutch, S. (2019). Health and social care practitioners' understanding of the problems of people with dementia-related visual processing impairment. *Health Soc Care Community, 27*(4), 982-990. doi:10.1111/hsc.12715

Mehegan, L., & Rainville, G. C. (2020). *2020 AARP Music and Brain Health Survey*. Retrieved from https://www.aarp.org/pri/topics/health/brain-health/brain-health-and-music.html

Melander, C. A., Kikhia, B., Olsson, M., Walivaara, B. M., & Savenstedt, S. (2018). The Impact of Using Measurements of Electrodermal Activity in the Assessment of Problematic Behaviour in Dementia. *Dement Geriatr Cogn Dis Extra, 8*(3), 333-347. doi:10.1159/000493339

Mendez, M. F. (2012). Early-onset Alzheimer's disease: nonamnestic subtypes and type 2 AD. *Arch Med Res, 43*(8), 677-685. doi:10.1016/j.arcmed.2012.11.009

Mendez, M. F., Ghajarania, M., & Perryman, K. M. (2002). Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 14*(1), 33-40.

Menghini, L., Gianfranchi, E., Cellini, N., Patron, E., Tagliabue, M., & Sarlo, M. (2019). Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology, 56*(11), e13441. doi:10.1111/psyp.13441

Mengoudi, K., Ravi, D., Yong, K. X. X., Primativo, S., Pavisic, I. M., Brotherhood, E., . . . Alexander, D. C. (2020). Augmenting Dementia Cognitive Assessment with Instruction-less Eye-tracking Tests. *IEEE Journal of Biomedical and Health Informatics, 24*(11), 3066-3075. doi:10.1109/JBHI.2020.3004686

Moss, H., Lynch, J., & O'Donoghue, J. (2018). Exploring the perceived health benefits of singing in a choir: an international cross-sectional mixed-methods study. *Perspect Public Health, 138*(3), 160-168. doi:10.1177/1757913917739652

Müller, V., Delius, J. A. M., & Lindenberger, U. (2018). Complex networks emerging during choir singing. *Ann N Y Acad Sci, 1431*(1), 85-101. doi:10.1111/nyas.13940

Music For Dementia. (2021). Musical Dementia Care Pathway. Retrieved from https://musicfordementia.org.uk/news-and-media/blog/2021/03/our-journey-with-music-through-dementia/

Music For Dementia. (2022). *Power of Music*. Retrieved from https://www.musicfordementia.org.uk/resources

Nair, B. R., Browne, W., Marley, J., & Heim, C. (2013). Music and dementia. *Degener Neurol Neuromuscul Dis, 3*, 47-51. doi:10.2147/DNND.S35762

Nair, B. R., Heim, C., Krishnan, C., D'Este, C., Marley, J., & Attia, J. (2011). The effect of Baroque music on behavioural disturbances in patients with dementia. *Australas J Ageing, 30*(1), 11-15. doi:10.1111/j.1741-6612.2010.00439.x

Nesbitt, C., Gupta, A., Maly, K., Okhravi, H. R., & Jain, S. (2019). 15 Feasibility of Using Wearable Sensors to Detect Agitation in Persons with Dementia. *CNS Spectrums, 24*(1), 181-181. doi:10.1017/s1092852919000105

Newkirk, L. A., Kim, J. M., Thompson, J. M., Tinklenberg, J. R., Yesavage, J. A., & Taylor, J. L. (2004). Validation of a 26-point telephone version of the Mini-Mental State Examination. *J Geriatr Psychiatry Neurol, 17*(2), 81-87. doi:10.1177/0891988704264534

Norberg, A., Melin, E., & Asplund, K. (1986). *Reactions to music, touch and object presentation in the final stage of dementia: An exploratory study*: International Journal of Nursing Studies. Vol.23(4), 1986, pp. 315-323.

North, A. C. (2010). Individual Differences in Musical Taste. *The American Journal of Psychology, 123*(2), 199–208. Retrieved from https://doi.org/10.5406/amerjpsyc.123.2.0199

O'Hare, M. (2020). Podcast - Music and Dementia. Retrieved from https://www.dementiaresearcher.nihr.ac.uk/podcast-music-and-dementia/

O'Neill, A. (2019). Behind the filming of Our Dementia Choir with Vicky McClure Retrieved from https://www.alzheimers.org.uk/blog/filming-our-dementia-choir-vicky-mcclure

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In *Handbook of psychophysiology, 2nd ed.* (pp. 533-575). New York, NY, US: Cambridge University Press.

Okada, K., Kurita, A., Takase, B., Otsuka, T., Kodani, E., Kusama, Y., . . . Mizuno, K. (2009). Effects of music therapy on autonomic nervous system activity, incidence of heart failure events, and plasma cytokine and catecholamine levels in elderly patients with cerebrovascular disease and dementia. *Int Heart J, 50*(1), 95-110. doi:10.1536/ihj.50.95

Omar, R., Hailstone, J. C., & Warren, J. D. (2012). Semantic Memory for Music in Dementia. *Music Perception, 29*(5), 467-477. doi:10.1525/mp.2012.29.5.467

Opstad, H. (n.d.). Music for Life: Out of the Ordinary Retrieved from https://wigmore-hall.org.uk/news/music-for-life-out-of-the-ordinary

Osman, S. E., Tischler, V., & Schneider, J. (2016). 'Singing for the Brain': A qualitative study exploring the health and well-being benefits of singing for people with dementia and their carers. *Dementia (London), 15*(6), 1326-1339. doi:10.1177/1471301214556291

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346-1353. doi:10.1111/j.1469-8986.2011.01217.x

Owens, A. P. (2020). The Role of Heart Rate Variability in the Future of Remote Digital Biomarkers. *Front Neurosci, 14*, 582145. doi:10.3389/fnins.2020.582145

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*(1-2), 185-198. doi:10.1016/s1071-5819(03)00017-x

Patel, A. D. (2013). Sharing and Nonsharing of Brain Resources for Language and Music. In M. A. Arbib (Ed.), *Language, Music and the Brain*. Cambridge, MA.

Pavisic, I. M., Firth, N. C., Parsons, S., Rego, D. M., Shakespeare, T. J., Yong, K. X. X., . . . Primativo, S. (2017). Eyetracking Metrics in Young Onset Alzheimer's Disease: A Window into Cognitive Visual Functions. *Front Neurol, 8*, 377. doi:10.3389/fneur.2017.00377

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PLoS One, 6*(11), e27241. doi:10.1371/journal.pone.0027241

Petrovsky, D., Cacchione, P. Z., & George, M. (2015). Review of the effect of music interventions on symptoms of anxiety and depression in older adults with mild dementia. *International Psychogeriatrics, 27*(10), 1661-1670. doi:10.1017/S1041610215000393

Pham, N. T. T., Graff-Radford, J., Machulda, M. M., Spychalla, A. J., Schwarz, C. G., Senjem, M. L., . . . Whitwell, J. L. (2022). Regional white matter hyperintensities in posterior cortical atrophy and logopenic progressive aphasia. *Neurobiol Aging, 119*, 46-55. doi:10.1016/j.neurobiolaging.2022.07.008

Philip, K. E., Lewis, A., Jeffery, E., Buttery, S., Cave, P., Cristiano, D., . . . Hopkinson, N. S. (2020). Moving singing for lung health online in response to COVID-19: experience from a randomised controlled trial. *BMJ Open Respir Res, 7*(1). doi:10.1136/bmjresp-2020-000737

Pierrot-Deseilligny, C., Rivaud, S., Gaymard, B., Muri, R., & Vermersch, A. I. (1995). Cortical control of saccades. *Ann Neurol, 37*(5), 557-567. doi:10.1002/ana.410370504

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology, 47*(3), 560-569. doi:10.1111/j.1469-8986.2009.00947.x

Playlist for Life. (2019). Red Flag Songs. Retrieved from https://www.playlistforlife.org.uk/red-flag-songs/

Polk, M., & Kertesz, A. (1993). Music and language in degenerative disease of the brain. *Brain and cognition, 22*, 98-117.

Popay, J. R., H,; Sowden, A.; Petticrew,M,; Arai, L,; Rodgers, M,; Britten, N.; Roen, K.; Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*. Retrieved from https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf

Qiu, S., Joshi, P. S., Miller, M. I., Xue, C., Zhou, X., Karjadi, C., . . . Kolachalama, V. B. (2020). Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain, 143*(6), 1920-1933. doi:10.1093/brain/awaa137

Qiu, S., Miller, M. I., Joshi, P. S., Lee, J. C., Xue, C., Ni, Y., . . . Kolachalama, V. B. (2022). Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun, 13*(1), 3404. doi:10.1038/s41467-022-31037-5

Quoniam, N., Ergis, A. M., Fossati, P., Peretz, I., Samson, S., Sarazin, M., & Allilaire, J. F. (2003). Implicit and explicit emotional memory for melodies in Alzheimer's disease and depression. *Ann N Y Acad Sci, 999*, 381-384. doi:10.1196/annals.1284.047

Raglio, A., Bai, S., Figini, C., Manera, M., Imbriani, C., Fundarò, C., & Buonocore, M. (2022). Music listening for people living with dementia: a systematic review. *G Ital Med Lav Erg, 44*(1), 84-92.

Raglio, A., Maestri, R., Robbi, E., Pierobon, A., La Rovere, M. T., & Pinna, G. D. (2022). Effect of Algorithmic Music Listening on Cardiac Autonomic Nervous System Activity: An Exploratory, Randomized Crossover Study. *J Clin Med, 11*(19). doi:10.3390/jcm11195738

Raglio, A., Pavlic, E., & Bellandi, D. (2018). Music Listening for People Living With Dementia. *J Am Med Dir Assoc, 19*(8), 722-723. doi:10.1016/j.jamda.2018.05.027

Ranscombe, P. (2019). Creating a personal music playlist to help treat dementia. *Lancet Neurol*. doi:10.1016/S1474-4422(19)30186-3

Ranscombe, P. (2021). Our Dementia Choir hits the right notes. *The Lancet Neurology, 20*(3). doi:10.1016/s1474-4422(19)30247-9

Ranson, J. M., Bucholc, M., Lyall, D., Newby, D., Winchester, L., Oxtoby, N. P., . . . Llewellyn, D. J. (2023). Harnessing the potential of machine learning and artificial intelligence for dementia research. *Brain Inform, 10*(1), 6. doi:10.1186/s40708-022-00183-3

Rao, C. B., Peatfield, J. C., McAdam, K. P. W. J., Nunn, A. J., & Georgieva, D. P. (2021). A Focus on the Reminiscence Bump to Personalize Music Playlists for Dementia. *Journal of multidisciplinary healthcare, 14*, 2195-2204. doi:10.2147/JMDH.S312725

Regalia, G., Resnati, D., & Tognetti, S. (2023). Sensors on the Wrist. In *Encyclopedia of Sensors and Biosensors* (pp. 1-20).

Requena-Komuro, M. C., Jiang, J., Dobson, L., Benhamou, E., Russell, L., Bond, R. L., . . . Hardy, C. J. (2022). Remote versus face-to-face neuropsychological testing for dementia research: a comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals. *BMJ Open, 12*(11), e064576. doi:10.1136/bmjopen-2022-064576

Ridder, H. M., & Aldridge, D. (2005). *Individual Music Therapy with Persons with Frontotemporal Dementia: Singing Dialogue. [References]*: Nordic Journal of Music Therapy. Vol.14(2), 2005, pp. 91-106.

Ridha, B. H., Crutch, S., Cutler, D., Frost, C., Knight, W., Barker, S., . . . Rossor, M. N. (2018). A double-blind placebo-controlled cross-over clinical trial of DONepezil In Posterior cortical atrophy due to underlying Alzheimer's Disease: DONIPAD study. *Alzheimers Res Ther, 10*(1), 44. doi:10.1186/s13195-018-0363-1

Riganello, F., Candelieri, A., Quintieri, M., Conforti, D., & Dolce, G. (2010). Heart rate variability: an index of brain processing in vegetative state? An artificial intelligence, data mining study. *Clin Neurophysiol, 121*(12), 2024-2034. doi:10.1016/j.clinph.2010.05.010

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves *BMC Bioinformatics, 12*, 77. doi:https://doi.org/10.1186/1471-2105-12-77

Romano Bergstrom, J., Duda, S., Hawkins, D., & McGill, M. (2014). Physiological Response Measurements. In *Eye Tracking in User Experience Design* (pp. 81-108).

Ruksenaite, J., Volkmer, A., Jiang, J., Johnson, J. C., Marshall, C. R., Warren, J. D., & Hardy, C. J. (2021). Primary Progressive Aphasia: Toward a Pathophysiological Synthesis. *Curr Neurol Neurosci Rep, 21*(3), 7. doi:10.1007/s11910-021-01097-z

Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality & Social Psychology, 39*(6), 1161-1178.

Sabry, F., Eltaras, T., Labda, W., Alzoubi, K., & Malluhi, Q. (2022). Machine Learning for Healthcare Wearable Devices: The Big Picture. *J Healthc Eng, 2022*, 4653923. doi:10.1155/2022/4653923

Sacks, O. (2007). *Musicophilia: Tales of Music and the Brain*.

Sakamoto, M., Ando, H., & Tsutou, A. (2013). Comparing the effects of different individualized music interventions for elderly individuals with severe dementia. *Int Psychogeriatr, 25*(5), 775-784. doi:10.1017/S1041610212002256

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci, 14*(2), 257-262. doi:10.1038/nn.2726

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The Rewarding Aspects of Music Listening Are Related to Degree of Emotional Arousal. *PLoS One, 4*(10). doi:10.1371/journal.pone.0007487

Samson, S., Dellacherie, D., & Platel, H. (2009). Emotional power of music in patients with memory disorders: clinical implications of cognitive neuroscience. *Ann N Y Acad Sci, 1169*, 245-255. doi:10.1111/j.1749-6632.2009.04555.x

Särkämö, T. (2018). Music for the ageing brain: Cognitive, emotional, social, and neural benefits of musical leisure activities in stroke and dementia. *Dementia (London), 17*(6), 670-685. doi:10.1177/1471301217729237

Särkämö, T. (2020). Musical Leisure Activities to Support Cognitive and Emotional Functioning in Aging and Dementia. In S. G. J. T. A Baird (Ed.), *Music and Dementia* (pp. 103-121). New York, NY: Oxford University Press

Särkämö, T. (2022). Singing the blues away: reduction of depression in dementia by recreational choir singing. *Lancet Healthy Longev, 3*(3), e124-e125. doi:10.1016/S2666-7568(22)00033-2

Särkämö, T., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2016a). Clinical and Demographic Factors Associated with the Cognitive and Emotional Efficacy of Regular Musical Activities in Dementia. *J Alzheimers Dis, 49*(3), 767-781. doi:10.3233/JAD-150453

Särkämö, T., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2016b). Pattern of Emotional Benefits Induced by Regular Singing and Music Listening in Dementia. *J Am Geriatr Soc, 64*(2), 439-440. doi:10.1111/jgs.13963

Särkämö, T., Laitinen, S., Tervaniemi, M., Numminen, A., Kurki, M., & Rantanen, P. (2012). Music, Emotion, and Dementia: Insight From Neuroscientific and Clinical Research. *Music and Medicine, 4*(3), 153-162. doi:10.1177/1943862112445323

Särkämö, T., & Sihvonen, A. J. (2018). Golden oldies and silver brains: Deficits, preservation, learning, and rehabilitation effects of music in ageing-related neurological disorders. *Cortex, 109*, 104-123. doi:10.1016/j.cortex.2018.08.034

Särkämö, T., Tervaniemi, M., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2014). Cognitive, emotional, and social benefits of regular musical activities in early dementia: randomized controlled study. *Gerontologist, 54*(4), 634-650. doi:10.1093/geront/gnt100

Schladt, T. M., Nordmann, G. C., Emilius, R., Kudielka, B. M., de Jong, T. R., & Neumann, I. D. (2017). Choir versus Solo Singing: Effects on Mood, and Salivary Oxytocin and Cortisol Concentrations. *Front Hum Neurosci, 11*, 430. doi:10.3389/fnhum.2017.00430

Schott, J. M., & Crutch, S. J. (2019). Posterior Cortical Atrophy. *CONTINUUM (MINNEAP MINN), 21*(1 (Dementia)), 52–75.

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *Int J Psychophysiol, 71*(1), 50-56. doi:10.1016/j.ijpsycho.2008.07.009

Shakespeare, T. J., Kaski, D., Yong, K. X., Paterson, R. W., Slattery, C. F., Ryan, N. S., . . . Crutch, S. J. (2015). Abnormalities of fixation, saccade and pursuit in posterior cortical atrophy. *Brain, 138*(Pt 7), 1976-1991. doi:10.1093/brain/awv103

Shakespeare, T. J., Pertzov, Y., Yong, K. X., Nicholas, J., & Crutch, S. J. (2015). Reduced modulation of scanpaths in response to task demands in posterior cortical atrophy. *Neuropsychologia, 68*, 190-200. doi:10.1016/j.neuropsychologia.2015.01.020

Shakespeare, T. J., Yong, K. X., Frost, C., Kim, L. G., Warrington, E. K., & Crutch, S. J. (2013). Scene perception in posterior cortical atrophy: categorization, description and fixation patterns. *Front Hum Neurosci, 7*, 621. doi:10.3389/fnhum.2013.00621

Sihvonen, A. J., Ripolles, P., Leo, V., Rodriguez-Fornells, A., Soinila, S., & Särkämö, T. (2016). Neural Basis of Acquired Amusia and Its Recovery after Stroke. *J Neurosci, 36*(34), 8872-8881. doi:10.1523/JNEUROSCI.0709-16.2016

Sihvonen, A. J., Särkämö, T., Leo, V., Tervaniemi, M., Altenmuller, E., & Soinila, S. (2017). Music-based interventions in neurological rehabilitation. *Lancet Neurol, 16*(8), 648-660. doi:10.1016/S1474-4422(17)30168-0

Sikka, R., Cuddy, L. L., Johnsrude, I. S., & Vanstone, A. D. (2015). An fMRI comparison of neural activity associated with recognition of familiar melodies in younger and older adults. *Front Neurosci, 9*, 356. doi:10.3389/fnins.2015.00356

Singh, N. A., Arani, A., Graff-Radford, J., Senjem, M. L., Martin, P. R., Machulda, M. M., . . . Whitwell, J. L. (2022). Distinct brain iron profiles associated with logopenic progressive aphasia and posterior cortical atrophy. *Neuroimage Clin, 36*, 103161. doi:10.1016/j.nicl.2022.103161

Sittler, M. C., Worschech, F., Wilz, G., Fellgiebel, A., & Wuttke-Linnemann, A. (2021). Psychobiological mechanisms underlying the health-beneficial effects of music in people living with dementia: A systematic review of the literature. *Physiol Behav, 233*, 113338. doi:10.1016/j.physbeh.2021.113338

Slattery, C. F., Agustus, J. L., Paterson, R. W., McCallion, O., Foulkes, A. J. M., Macpherson, K., . . . Warren, J. D. (2019). The functional neuroanatomy of musical memory in Alzheimer's disease. *Cortex, 115*, 357-370. doi:10.1016/j.cortex.2019.02.003

Spiro, N. (2010). Music and dementia: observing effects and searching for underlying theories. *Aging Ment Health, 14*(8), 891-899. doi:10.1080/13607863.2010.519328

Stapleton, K. (2020). Review of Music and Dementia: From Cognition to Therapy. by Amee Baird, Sandra Garrido, and Jeanette Tamplin. *The Journal of Sociology & Social Welfare, 57*(3), 219-222.

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int J Psychophysiol, 52*(1), 77-86. doi:10.1016/j.ijpsycho.2003.12.005

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: disorders of musical listening. *Brain, 129*(Pt 10), 2533-2553. doi:10.1093/brain/awl171

Strohmaier, S., Homans, K. M., Hulbert, S., Crutch, S. J., Brotherhood, E. V., Harding, E., & Camic, P. M. (2021). Arts-based interventions for people living with dementia: Measuring 'in the moment' wellbeing with the Canterbury Wellbeing Scales. *Wellcome Open Res, 6*, 59. doi:10.12688/wellcomeopenres.16596.3

Sun, Y., Baird, A., Gelding, R., de Wit, B., & Thompson, W. F. (2021). Can music enhance awareness in unresponsive people with severe dementia? An exploratory case series using behavioral, physiological and neurophysiological measures. *Neurocase, 27*(4), 354-365. doi:10.1080/13554794.2021.1966045

Suzuki, M., Kanamori, M., Nagasawa, S., Tokiko, I., & Takayuki, S. (2007). Music therapy-induced changes in behavioral evaluations, and saliva chromogranin A and immunoglobulin A concentrations in elderly patients with senile dementia. *Geriatrics & Gerontology International, 7*(1), 61-71. doi:10.1111/j.1447-0594.2007.00374.x

Suzuki, M., Kanamori, M., Watanabe, M., Nagasawa, S., Kojima, E., Ooshiro, H., & Nakahara, D. (2004). Behavioral and endocrinological evaluation of music therapy for elderly patients with dementia. *Nurs Health Sci, 6*(1), 11-18. doi:10.1111/j.1442-2018.2003.00168.x

Szabadi, E. (2018). Functional Organization of the Sympathetic Pathways Controlling the Pupil: Light-Inhibited and Light-Stimulated Pathways. *Front Neurol, 9*, 1069. doi:10.3389/fneur.2018.01069

Takahashi, T., & Matsushita, H. (2006). Long-Term Effects of Music Therapy on Elderly with Moderate/Severe Dementia. *Journal of Music Therapy, 43*(4), 317–333. doi:https://doi.org/10.1093/jmt/43.4.317

Tang-Wai, D. F., Graff-Radford, N. R., Boeve, B. F., Dickson, D. W., Parisi, J. E., Crook, R., . . . Petersen, R. C. (2004). Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology, 63*(7), 1168-1174.

Tang, H. Y., & Vezeau, T. (2010). The use of music intervention in healthcare research: a narrative review of the literature. *J Nurs Res, 18*(3), 174-190. doi:10.1097/JNR.0b013e3181efe1b1

Tasker, A. (2021). Hull firm develops music app to cut staff and patient stress. Retrieved from https://www.bbc.co.uk/news/uk-england-humber-57071627

Tekin, K., Sekeroglu, M. A., Kiziltoprak, H., Doguizi, S., Inanc, M., & Yilmazbas, P. (2018). Static and dynamic pupillometry data of healthy individuals. *Clin Exp Optom, 101*(5), 659-665. doi:10.1111/cxo.12659

Thomas, G., Crutch, S. J., & Camic, P. M. (2018). Measuring physiological responses to the arts in people with a dementia. *Int J Psychophysiol, 123*, 64-73. doi:10.1016/j.ijpsycho.2017.11.008

Thompson, Z., Baker, F. A., Tamplin, J., & Clark, I. N. (2021). How Singing can Help People With Dementia and Their Family Care-Partners: A Mixed Studies Systematic Review With Narrative Synthesis, Thematic Synthesis, and Meta-Integration. *Front Psychol, 12*, 764372. doi:10.3389/fpsyg.2021.764372

Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., . . . Kalman, J. (2018). A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Curr Alzheimer Res, 15*(2), 130-138. doi:10.2174/1567205014666171121114930

Tsoi, K. K. F., Chan, J. Y. C., Ng, Y.-M., Lee, M. M. Y., Kwok, T. C. Y., & Wong, S. Y. S. (2018). Receptive Music Therapy Is More Effective than Interactive Music Therapy to Relieve Behavioral and Psychological Symptoms of Dementia: A Systematic Review and Meta-Analysis. *Journal of the American Medical Directors Association, 19*(7), 568-576.e563. doi:https://doi.org/10.1016/j.jamda.2017.12.009

Turker, S., Reiterer, S. M., Seither-Preisler, A., & Schneider, P. (2017). "When Music Speaks": Auditory Cortex Morphology as a Neuroanatomical Marker of Language Aptitude and Musicality. *Front Psychol, 8*, 2096. doi:10.3389/fpsyg.2017.02096

UK Music. (2021). *This Is Music Report*. Retrieved from https://www.ukmusic.org/research-reports/this-is-music-2021/

Valdiglesias, V., Maseda, A., Lorenzo-Lopez, L., Pasaro, E., Millan-Calenti, J. C., & Laffon, B. (2017). Is Salivary Chromogranin A a Valid Psychological Stress Biomarker During Sensory Stimulation in People with Advanced Dementia? *Journal of Alzheimer's Disease, 55(4)*, 1509-1517.

Valentine, E., & Evans, C. (2001). The effects of solo singing, choral singing and swimming on mood and physiological indices. *British Journal of Medical Psychology, 74*(1), 115-120. doi:10.1348/000711201160849

van't Hooft, J. J., Fieldhouse, J. L. P., Singleton, E. H., Jaschke, A. C., Warren, J. D., Tijms, B. M., & Pijnenburg, Y. A. L. (2022). Music appreciation phenotypes in patients with frontotemporal dementia: A pilot study. *Int J Geriatr Psychiatry, 37*(9). doi:10.1002/gps.5793

van't Hooft, J. J., Pijnenburg, Y. A. L., Sikkes, S. A. M., Scheltens, P., Spikman, J. M., Jaschke, A. C., . . . Tijms, B. M. (2021). Frontotemporal dementia, music perception and social cognition share neurobiological circuits: A meta-analysis. *Brain Cogn, 148*, 105660. doi:10.1016/j.bandc.2020.105660

van den Bosch, I., Salimpoor, V. N., & Zatorre, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Front Hum Neurosci, 7*, 534. doi:10.3389/fnhum.2013.00534

van der Steen, J. T., Smaling, H. J., van der Wouden, J. C., Bruinsma, M. S., Scholten, R. J., & Vink, A. C. (2018). Music-based therapeutic interventions for people with dementia. *Cochrane Database Syst Rev, 7*(7), CD003477. doi:10.1002/14651858.CD003477.pub4

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., . . . Iwatsubo, T. (2023). Lecanemab in Early Alzheimer's Disease. *N Engl J Med, 388*(1), 9-21. doi:10.1056/NEJMoa2212948

Van Gerven, P. W. M., Paas, F., Van Merrienboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167-174. doi:10.1111/j.1469-8986.2003.00148.x

van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in hearing, 23*, 2331216519832483. doi:10.1177/2331216519832483

Vanstone, A. D., Sikka, R., Tangness, L., Sham, R., Garcia, A., & Cuddy, L. L. (2012). Episodic and Semantic Memory for Melodies in Alzheimer's Disease. *Music Perception, 29*(5), 501-507. doi:10.1525/mp.2012.29.5.501

Vickhoff, B., Malmgren, H., Astrom, R., Nyberg, G., Ekstrom, S. R., Engwall, M., . . . Jornsten, R. (2013). Music structure determines heart rate variability of singers. *Front Psychol, 4*, 334. doi:10.3389/fpsyg.2013.00334

Vink, A., & Hanser, S. (2018). Music-Based Therapeutic Interventions for People with Dementia: A Mini-Review. *Medicines (Basel), 5*(4). doi:10.3390/medicines5040109

Võ, M. L., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology, 45*(1), 130-140. doi:10.1111/j.1469-8986.2007.00606.x

Voices Now. (2017). The Big Choral Census. Retrieved from tinyurl.com/yckp8faw

Waitzman, E. (2022). Primary and community care: Improving patient outcomes. Retrieved from https://lordslibrary.parliament.uk/primary-and-community-care-improving-patient-outcomes/

Walker, N., Crutch, S. J., West, J., Jones, F. W., Brotherhood, E. V., Harding, E., & Camic, P. M. (2021). Singing and music making: physiological responses across early to later stages of dementia. *Wellcome Open Res, 6*, 150. doi:10.12688/wellcomeopenres.16856.2

Warran, K., Burton, A., & Fancourt, D. (2022). What are the active ingredients of 'arts in health' activities? Development of the INgredients iN ArTs in hEalth (INNATE) Framework. *Wellcome Open Res, 7*, 10. doi:10.12688/wellcomeopenres.17414.2

Warrington, E. K. (1996). The Camden Memory Test Battery. In: Brighton, UK: Psychology Press.

Warrington, E. K., James, M., & Thames Valley Test Company. (1991). The Visual Object and Space Perception Battery.

Waters, B., Sousa, L., Orrell, M., & McDermott, O. (2022). Analysing the use of music to facilitate social interaction in care home residents with dementia: Narrative synthesis systematic review. *Dementia (London), 21*(6), 2072-2094. doi:10.1177/14713012221100625

Weise, L., Jakob, E., Topfer, N. F., & Wilz, G. (2018). Study protocol: individualized music for people with dementia - improvement of quality of life and social participation for people with dementia in institutional care. *BMC Geriatr, 18*(1), 313. doi:10.1186/s12877-018-1000-3

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *J Exp Psychol Hum Percept Perform, 42*(8), 1061-1065. doi:10.1037/xhp0000226

Whitehouse, P. J., Vella Burrows, T., & Stephenson, D. (2018). Global perspectives on dementia and art: An international discussion about changing public health policy. *Dementia (London), 17*(6), 785-797. doi:10.1177/1471301217739737

Wilkins, R. W., Hodges, D. A., Laurienti, P. J., Steen, M., & Burdette, J. H. (2014). Network science and the effects of music preference on functional brain connectivity: from Beethoven to Eminem. *Sci Rep, 4*, 6130. doi:10.1038/srep06130

Willems, S., Adam, S., & Van der Linden, M. (2002). Normal mere exposure effect with impaired recognition in Alzheimer's disease. *Cortex, 38*(1), 77-86. doi:10.1016/s0010-9452(08)70640-3

Willison, J. R., & Warrington, E. K. (1992). Cognitive retardation in a patient with preservation of psychomotor speed. *Behavioural neurology, 5*(2), 113-116.

Wong, A., Burford, S., Wyles, C. L., Mundy, H., & Sainsbury, R. (2008). Evaluation of strategies to improve nutrition in people with dementia in an assessment unit. *The Journal of Nutrition Health and Aging, 12*(5), 309-312. doi:10.1007/BF02982660

Woods, B., O'Philbin, L., Farrell, E. M., Spector, A. E., & Orrell, M. (2018). Reminiscence therapy for dementia. *Cochrane Database Syst Rev, 3*(3), CD001120. doi:10.1002/14651858.CD001120.pub3

World Health Organisation. (2023). Dementia. Retrieved from https://www.who.int/news-room/fact-sheets/detail/dementia

Yildirim Biçer, G., Zor, K. R., & Küçük, E. (2022). Do static and dynamic pupillary parameters differ according to childhood, adulthood, and old age? A quantitative study in healthy volunteers. *Indian J Ophthalmol, 70*(10), 3575-3578. doi:10.4103/ijo.IJO_1254_22

Yong, K. X. X., Graff-Radford, J., Ahmed, S., Chapleau, M., Ossenkoppele, R., Putcha, D., . . . Harding, E. (2023). Diagnosis and Management of Posterior Cortical Atrophy. *Curr Treat Options Neurol, 25*(2), 23-43. doi:10.1007/s11940-022-00745-0

Zeeman, M., Figeys, M., Brimmo, T., Burnstad, C., Hao, J., & Kim, E. S. (2023). Task-Evoked Pupillary Response as a Potential Biomarker of Dementia and Mild Cognitive Impairment: A Scoping Review. *Am J Alzheimers Dis Other Demen, 38*, 15333175231160010. doi:10.1177/15333175231160010

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software, 11*(10), 1-17. doi:doi:10.18637/jss.v011.i10

Zeileis, A., Köll, S., & Graham, N. (2020). Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *Journal of Statistical Software, 95*(1), 1-36. Retrieved from doi:10.18637/jss

Zeilig, H., West, J., & van der Byl Williams, M. (2018). Co-creativity: possibilities for using the arts with people with a dementia. *Quality in Ageing and Older Adults, 19*(2), 135-145. doi:10.1108/qaoa-02-2018-0008

Zhang, L., Arandjelovic, O., Dewar, S., Astell, A., Doherty, G., & Ellis, M. (2020). *Quantification of Advanced Dementia Patients' Engagement in Therapeutic Sessions: An Automatic Video Based Approach using Computer Vision and Machine Learning.* Paper presented at the Annu Int Conf IEEE Eng Med Biol Soc. 2020, Montreal, QC, Canada

Zhao, S., Bury, G., Milne, A., & Chait, M. (2019). Pupillometry as an Objective Measure of Sustained Attention in Young and Older Listeners. *Trends in hearing, 23*, 2331216519887815-2331216519887815. doi:10.1177/2331216519887815

Zhou, J., Greicius, M. D., Gennatas, E. D., Growdon, M. E., Jang, J. Y., Rabinovici, G. D., . . . Seeley, W. W. (2010). Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain, 133*(Pt 5), 1352-1367. doi:10.1093/brain/awq075

9   APPENDICES

Appendix A: Dementia-music-physiology studies: Experimental design and sampling

Authors	Dementia type	Dementia severity	Intervention type	Physiological measurements
Passive music listening				
Norberg, Melin, & Asplund (1986) *Sweden*	AD	Severe	Once-twice daily individual stimulation Recorded music, touch and object presentation 12-day intervention	HR RR
Maseda et al. (2018) *Spain*	UD	Severe and very severe (6-7 Global Deterioration Scale, GDS)	30-minute individual MT or MSSE sessions (2/week) 12 weeks 24 sessions MT session followed Gerdner (2013) protocol	HR SpO2
Garrido et al. (2018) *Australia*	AD (n=3) VD (n=3) AID (n=2) KD (n=1) MCI (n=1) UD (n=89)	Range (0-25; x̄ = 7.7 Mini mental state examination (MMSE)	Three personalised playlists 2-4 songs in each Full songs presented	FAU (fEMG)
Garrido et al. (2019) *Australia*	AD (n=3) VD (n=3) AID (n=2) KD (n=1) MCI (n=1) UD (n=89)	Range (0-25; x̄ = 7.7 MMSE)	Three personalised playlists Full songs presented	EDA FAU (fEMG)
Gök Ugur et al. (2019) *Turkey*	AD	Range (0-28; MT x̄ = 9.93 ± 7.98; control x̄ = 12.23 ± 8.76 MMSE)	30-minute MT individualised sessions listening to Turkish music (3/week) 12 weeks	HR RR SBP DBP
Corrêa et al. (2020) *Brazil*	UD (probable dementia)	Severe	4 x music listening sessions (20 minutes each), 1 per week	HR % cardiac coherence SBP DBP
Barradas et al. (2021) *Portugal*	AD	Mild to moderate (1-2 CDR; x̄ = 18.00 ± 3.98 MMSE)	5 instrumental music pieces; from Juslin et al. (2014; 2015)	EDA fEMG

Liu et al. (2021) *China*	UD	Mild [21-27 MMSE inclusion criteria]	Cross-sectional presentation of 1-minute excerpts of music and two environmental sounds (birdsong and stream)	FER (fEMG)
Benhamou et al. (2021) *United Kingdom*	FTD [bvFTD n = 21; svPPA n = 12; nfvPPA n = 12; tAD n = 19]	Mild-moderate	48 x piano excerpts comprising 4 experimental conditions (unedited, single note deviation, single note deviation + key/tonality deviation, single note deviation + acoustic deviation)	Maximum pupil dilation response
Sun et al. (2021) *Australia*	UD	Severe [Non-responsive; 0-2/4 on Guy's Advanced Dementia Schedule (GADS)]	Cross-sectional presentation of 6 1-minute excerpts of familiar (favourite, as informed by family member) and unfamiliar (modern release) songs, and control condition of white noise + amplitude envelope of favourite song segments combined. Other conditions included presentation of personal memory and news story	HR ST EDA
da Rocha et al. (2022) *Brazil*	AD (n = 10) Mixed (n = 2) Other (n = 2)	-	One session per week for 2 months (20 min/week) Intervention group: one music listening session of concert music, including: Nocturne Opus no 2. by Frederic Chopin; Adagio in G minor (best live version) by Tomaso Albinoni; Serenade for Winds (K. 361, 3rd movement) by Mozart. Control group: headphone use without music	HR DBP SBP

Opportunity for active music-making (e.g. singing)

Ridder & Aldridge (2005) *Denmark*	FTD	Severe (MMSE = 0/30)	20-30 minute individual MT sessions (5 days/week) 20 sessions	HR
Takahashi & Matsushita (2006) *Japan*	CVD (n=8) AD (n=15) PDD (n=1)	Moderate – severe dementia (mean HDS-R score = 6.0)	60-minute group MT sessions (1 day/week) 2 years	SBP[x]
Kurita et al. (2006) *Japan*	CVD	Mild - moderate (MMSE x̄ = 17 ± 6 / 30)	30 minute MT session (one-off)	HRV* measured by: LF/HR ECG data R-R pNN50 RMMSD

Okada et al. (2009) *Japan*	CVD	Severe	45-minute group MT sessions (1 day/week) 10 weeks 10 sessions	See Kurita et al. (2006)
Bourne et al (2019) *United Kingdom*	AD (n = 6) LBD (n = 1) FTD (n = 2) Mixed (n = 1)	Early-middle stages	One-off community singing group: Welcome song Bella Mama (a Torres Strait Islands song, 90 bpm1) Bei Männern (Mozart) Rio Grande (Sea shanty) Wiegenlied (Brahms) Erie Canal (Thomas S Allen) Dona nobis pacem (Traditional canon) Oh what a beautiful morning (Rodgers & Hammerstein)	HRˣ
Walker et al. (2021) *[Study 1] United Kingdom*	AD (n = 6) LBD (n = 1) FTD (n = 1) Mixed AD/FTD (n = 1)	Mild-Moderate (N = 9)	One-off community singing group: Welcome song Bella Mama (a Torres Strait Islands song, 90 bpm1) Bei Männern (Mozart, from the Magic Flute, 86 bpm) The Lion Sleeps Tonight (Wimoweh) (South African, 126 bpm) The Erie Canal (American popular, 126 bpm) Broken down and practiced and then sung in their entirety	HR EDA ST ACC

Comparisons of music listening and opportunity for active music-making (e.g. singing)

Sakamoto, Ando & Tsutou (2013) *Japan*	AD	Severe (CDR rating) MMSE in each group ~4.7 (s.d. ~4)	30-minute weekly individual sessions Passive or Interactive MT 10 weeks 10 sessions	HRV (HF and HR)
Walker et al. (2021) continued *[Study 2] United Kingdom*	AD (n = 3) VD (n = 1) Mixed AD/VD (n = 1) Atypical/mixed (n = 1)	Severe (N = 6)	2 x interactive music sessions (based on Music For Life (Rose, 1993) and control session (music listening)	HR EDA ST ACC

* Also included blood sampling for Plasma cytokines (IL-6 and TNF); Plasma adrenaline; Noradrenaline pre- and post- (24 hours) MT intervention; ˣ Also included salivary measure.

Diagnosis abbreviations (*in order of appearance*): AD = Alzheimer's disease; UD = diagnosis of dementia (specific type undetermined); VD = vascular dementia; AID = alcohol induced dementia; KD = Korsakoff's disease; MCI = mild cognitive impairment; FTD = frontotemporal dementia; bvFTD = behavioural variant frontotemporal dementia; svPPA = semantic variant primary progressive aphasia; nfvPPA = nonfluent variant primary progressive aphasia; tAD = typical Alzheimer's disease; CVD = cerebrovascular dementia; PDD = Parkinson's Disease Dementia; LBD = Lewy body dementia.

Physiological measure abbreviations (*in order of appearance*): HR = heart rate; RR = rate of respiration; SpO2 = blood oxygen saturation; FAU = face action units; EDA = electrodermal activity; SBP = systolic blood pressure; DBP = diastolic blood pressure; fEMG = facial electromyography; FER = facial expression recognition; ST = skin temperature; LF/HF = low-frequency/high frequency ratio; ECG = electrocardiogram; R-R = R wave amplitude of ECG (intervals); pNN50 = percentage of consecutive R-R interval differences ≥ 50 milliseconds; RMMD = square root of the average sum of squared differences between contiguous R-R intervals; ACC = accelerometer/movement; HR = high frequency.

Music activity abbreviations (*in order of appearance):* MT = music therapy; MSSE = multisensory stimulation environment.

Appendix B: Dementia-music-physiology studies: Data collection timing and findings

Authors	Physiological data collection timing	Main findings (statistically significant unless stated)
Passive music listening		
Norberg, Melin, & Asplund (1986) *Sweden*	Pre-post each song Pre-post overall intervention	↑ HR [Patient A: Tune 2; Patient B: Tune 5] ↔ HR pre-post intervention ↓ RR [Patient A: Tunes 4 and 6] ↓ RR pre-post intervention [Patient A only]
Maseda et al. (2018) *Spain*	Immediately before and after MT sessions	↓ HR immediately post-MT (and post-MSSE) [A] ↑ SpO2 immediately post-MT (and post-MSSE) [A]
Garrido et al. (2018) *Australia*	Continuous, in-the-moment (baseline vs. 7-minute condition blocks)	↑ Sadness FAU - dementia participants with high levels of depression and low anxiety ↑ Sadness FAU - AD participants during music listening vs. Vasc. dementia
Garrido et al. (2019) *Australia*	Continuous, in-the-moment (baseline vs. 7-minute condition blocks)	↑ EDA (fast tempo music vs. baseline) ↑ EDA (fast tempo music, low self-reported enjoyment group) ↑ EDA (slow tempo music vs. baseline) ↑ Activation of depressor anguli oris (fEMG) for music in
Gök Ugur et al. (2019) *Turkey*	Baseline collection, and at 12th week	↔ HR (no change) ↔ RR (no change) ↓ SBP (MT group) ↓ DBP (MT group)
Corrêa et al. (2020) *Brazil*	Pre-listening and post-listening	↔ HR (no change) ↔ % cardiac coherence (no change) ↓ SBP (post-intervention readings in classical music group cf. pop music group) ↓ DBP (post-intervention readings in classical music group cf. pop music group)

273

Barradas et al. (2021) *Portugal*	Before (baseline) and during music listening	↑ EDA higher during target-mechanism songs compared to baseline
		↑ EDA higher during brain stem music than baseline and contagion condition
		↑ EDA higher during memory music than baseline and contagion condition
		↑ EDA, EMG zygomaticus and corrugator during music compared with baseline
Liu et al. (2021) *China*	In-the-moment	↗ Positive correlations (subjective pleasure and FER valence) in music at all timepoints
		↗ Positive correlation (subjective arousal rating and FER valence) in music at 20s
		↗ Positive correlation (subjective feelings of 'dominance' and FER valence) in music 60 - 80s
		↗ Positive correlation (subjective feelings of acoustic comfort and FER valence) in music 60 - 80s
Benhamou et al. (2021) *United Kingdom*	500ms before deviant onset In-the-moment (segmented at deviant onset)	↑ nfvPPA and tAD groups elicited larger pupillary responses than the svPPA and bvFTD groups
		↑ Acoustic deviant > Syntactic deviant > Semantic deviant > No-deviant in pupillary responses in all groups
Sun et al. (2021) *Australia*	In-the-moment	↑ Skin temperature during music compared with personal memory
		↑ Skin temperature during music compared with news story [B]
		↔ No significant difference between the familiar/unfamiliar music response
		↔ No significant responses in EDA or HR
da Rocha et al. (2022) *Brazil*	Pre- and post-intervention (comparing first and last within-session responses)	↔ No significant changes in BP or HR

Opportunity for active music-making (e.g. singing)

Ridder & Aldridge (2005) *Denmark*	Pre intervention (30 mins daily); continuously during intervention [every 5 seconds in weeks 2-5); post-intervention and follow-up (4 weeks after MT)	↓ HR (post-MT intervention at Session 20 cf. Session 1)

Takahashi & Matsushita (2006) *Japan*	Pre-post total intervention, 1- (midpoint) and 2-year (end intervention)	↑ SBP (nonMT control group at 2 years relative to MT group) ↑ SBP (post-session relative to pre-session in MT group)
Kurita et al. (2006) *Japan*	30 mins prior MT, during MT, and 2 hours post-MT	↑ HF, R-R, pNN50, RMSSD (during vs. pre-MT) ↓ HF, R-R, pNN50, RMSSD (post vs. during MT) ↓ LF/HF (during vs. pre-MT) ↑LF/HF (post vs. during MT) [B]
Okada et al. (2009) *Japan*	Before (1 hour preceding MT) During MT (45 mins) Post MT (2 hours subsequent to MT, for 1 hour)	↔ No differences in R-R responses in MT group ↑ HF, pNN50, RMSSD in MT group (during vs. pre-MT) ↓ pNN50, RMSSD in MT group (post vs. during MT) ↓ LF/HF in MT group during MT vs. before MT [B] ↔ No differences in HRV responses in control (non-MT) group
Bourne et al (2019) *United Kingdom*	Before (2 minute 33 seconds) During (60 minutes) After (2 minutes 33 seconds)	↑ HRV in PLWD group within-choral session relative to pre- and post-session
Walker et al. (2021) *[Study 1]* *United Kingdom*	Before and in-the-moment	↓↑ HR higher (N = 5) and HR lower (N = 1) during first song vs. baseline [C] ↑ EDA higher during first song vs. baseline (N = 8) ↓↑ ST higher (N = 3) and ST lower (N=2) during first song vs. baseline [C] ↑ HR higher during fast songs compared with slow songs (N = 5) ↓↑ EDA higher in slow songs (N = 5); higher in fast songs (N = 3) [C] ↓↑ ST higher during fast songs (N = 6) and higher during slow songs (N = 1) [C]

Comparisons of music listening and opportunity for active music-making (e.g. singing)

Sakamoto, Ando & Tsutou (2013) *Japan*	Pre- and post- music intervention	↓ HR immediately post-MT (both interactive & passive groups) ↑ HF immediately post-MT (interactive, passive, and control groups) ↔ No HR/HF between-group differences between passive and interactive groups

275

Walker et al. (2021) continued *[Study 2] United Kingdom*	In-the-moment Control session (music listening) Session 1 (Interactive music group) Session 6 (interactive music group)	**Within-session comparisons (same RQs as Study 1)** *Music listening - first song vs. baseline* ↓ HR lower (N = 4) ↑ EDA higher (N = 5) ↑ ST higher (N = 4) *Interactive music sessions - first song vs. baseline* ↕ HR higher (N = 2) and HR lower (N = 2) [C] ↕ EDA higher (Session 1 and Session 6 N = 4) / lower (Session 6 N = 1) [C] ↕ ST higher (Session 1 N = 2; Session 6 N = 4) / lower (Session 1 N = 3; Session 6 N = 1) [C] *Music listening - fast tempo v. slower tempo* ↕ HR higher (N = 3) and HR lower (N = 2) [C] ↓ EDA lower (N = 6) ↓ ST lower (N = 5) *Interactive music sessions - fast tempo v. slower tempo* ↕ HR higher (Session 1 N = 1; Session 6 N = 1) and HR lower (Session 1 N = 2, Session 6 N = 4) [C] ↕ EDA higher (Session 1 N = 1; Session 6 N = 2) and lower (Session 1 N = 4; Session 6 N = 4) [C] ↕ ST higher (Session 1 N = 4) and lower (Session1 N = 1; Session 6 N = 6) [C]

A. No significant differences between the MT and MSSE groups

B. Indicates this result did not reach statistical significance

C. ↕ indicates individual differences in the physiological reactions to the music

Music activity abbreviations *(in order of appearance):* MT = music therapy; MSSE = multisensory simulation environment.

Appendix C: Music engagement responses from healthy older adults, split by gender

		Gender		P-value
		Female	Male	
N		52	38	
Music linked with an important life event		28 (54%)	13 (36%)	0.101 [c]
Description of theoretical music knowledge	Followed a music course	10 (20%)	2 (6%)	0.076 [b]
	No theoretical knowledge	30 (60%)	23 (66%)	
	Read music	2 (4%)	6 (17%)	
	Studied music	8 (16%)	4 (11%)	
Played in instrument over the life-course		35 (67%)	17 (46%)	**0.044* [c]**
Number of instruments played *(of participants who played participants)*	1	20 (59%)	6 (35%)	0.223 [b]
	2	11 (32%)	7 (41%)	
	3	3 (9%)	4 (24%)	
Current instrument playing		8 (15%)	8 (21%)	0.487 [c]
Instrument-playing as hobby		8 (100%)	8 (100%)	1.000 [b]
Instrument-playing as professional		0 (0%)	0 (0%)	1.000 [b]
Years longest instrument played for	$\tilde{x}$ [IQR]	0 [0;3.5]	0 [0;2.0]	0.508 [a]
	Min./Max.	0 / 69.0	0 / 65.0	
Highest grade attained in instrument	$\tilde{x}$ [IQR]	3.0 [0.8;5.2]	1.0 [0;7.8]	0.719 [a]
	Min./Max.	0 / 8.0	0 / 8.0	
Sung over the life-course		7 (13%)	1 (3%)	0.133 [c]
Engaged in singing currently		6 (12%)	1 (3%)	0.232 [c]
Current singing as hobby		4 (67%)	1 (100%)	1.000 [c]

[a] = Wilcoxon rank sum test; [b] = Fisher's Exact Test; [c] = Pearson's Chi-Squared test * = significant at the level of $p < 0.05$

Appendix D: Candidate's contribution to the work

Chapter 2: Literature review

- Selection of review approach

- Defining search terms and running search

- Defining inclusion/exclusion criteria

- Endnote organisation and identifying literature for inclusion

- Literature synthesis

- Reporting

Chapters 3-4: Play It Again

- Administrative assistance with large overarching grant funding application (collating relevant documentation, liaising with project Co-Is on grant content, grant formatting and typesetting)

- Ethical application and amendments (NHS and University REC pathways)

- Paradigm design in collaboration with Prof Seb Crutch & Dr Nick Firth [credited]

- Submitting project plan to Management Committee for approval

- Control participant recruitment (N = 57) and data collection (N = 55)
  N = 42 participants' recruitment and data collection undertaken by Ivanna Pavisic [credited], recruitment and data collection N = 1 by Kyriaki Mengoudi [credited] and data collection N = 2 by Ben Levett [credited] under supervision by candidate as part of student volunteer experience.

- Patient participant recruitment, demographic, physiological and neuropsychological data collection (N = 42)

- Data cleaning and pupillometry data pre-processing

- Designing statistical plan (verified by LSHTM statistician Amy MacDougall [credited])

- Running statistical analyses in R and verifying findings and interpretation with AM

- All data visualisation

- Reporting

Chapter 5: AI in Dementia Care collaboration

- Project governance (e.g. ethical approvals of data transfer and management)
- Secondary analysis conceptualisation
- Providing clinical context for collaborating colleagues at UCL Computer Science
- Data cleaning and pupillometry data pre-processing (in conjunction with YZ [credited])
- Neuropsychological data statistical analyses
- Neuropsychological data visualisation
- Reporting

Chapter 6: Our Dementia Choir

- On-site data collection for 3/6 data collection timepoints; 3 data collection timepoints acquired by Emma Harding [credited] (University of Nottingham)
- TMMSE follow-up data collection
- Data synchronisation, timestamping and segmentation
- Physiological data statistical analyses
- Data visualisation
- Reporting

END