PHD IN SPEECH, HEARING AND PHONETIC SCIENCES

PRODUCTION AND PERCEPTION OF EMOTIONAL PROSODY:

THE CASE OF AUTISM AND PARKINSON'S DISEASE

CRISTIANE CHIA TSENG HSU

Thesis submitted in partial fulfilment

of the requirements for the degree of PhD in Speech, Hearing and Phonetic Sciences

of University College London in 2016

UNIVERSITY COLLEGE LONDON

DEPARTMENT OF SPEECH, HEARING AND PHONETIC SCIENCES

‘I, Cristiane Chia Tseng Hsu, 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

The current study aims to investigate the acoustic correlates of the production and perception in emotional and attitudinal prosody of Taiwan Mandarin-speaking adolescents with autism (ASD) and Parkinson's disease (PD) patients. It has been widely accepted among speech therapists that, although from very different aetiologies, individuals with ASD or PD speak with deviant prosodic patterns, monopitch, wrong stress, and greater difficulties in expressing emotions through prosody, while their perception of emotion and attitude is also jeopardised. However, when comparing their performance with their respective peers, results from previews studies have been far from consistent, ranging from significantly poorer to similar performance level.

This disparity in findings may emerge from disease severity, task type, and research method. Therefore, in the present study, these issues were systematically controlled. Only individuals with mild to moderate ASD or PD were included in the experimental groups. Synthetic speech was used to generate auditory stimuli for the perceptual tasks, allowing precise manipulations of acoustic parameters. The use of designed sentences to elicit speech provided better comparison across speakers.

The major innovation of this study is the use of bio-informational dimension (BID) theory as the theoretical framework. BID dimensions include body-size projection, dynamicity, audibility and association. Body-size projection associates high pitch with submissiveness, friendliness, politeness, vulnerability, and low pitch with dominance, confidence, protection, aggressiveness. This research paradigm has been proven to be effective in several studies examining the perception of emotional prosody. The hypotheses of the present study were made based on body-size projection.

The main findings of this study include 1) adolescents with ASD and PD patients performed poorer than their controls in the perceptual judgement of emotional

and attitudinal prosody; 2) only their expressive skills in emotional prosody was comparable to their controls; 3) their expressive skills in attitudinal prosody was worse than their controls.

Acknowledgments

I first learnt about autism from a TV commercial in Brazil when I was fourteen. It was showing a child in the shadow, wiggling fingers, waving arms and shaking head repeatedly in a ritualistic manner.

Only in 2005, during a lecture in child language acquisition in the University of Leeds, I learnt that some individuals with autism could have speech, and their prosody is usually atypical. Then I had the idea to investigate speech prosody in autism.

Ten years have gone until I finally conducted the research and wrote this thesis. I want to thank many people who helped and contributed in several ways to make this thesis possible.

First of all, I want to thank my respondents and their family. I dedicate this thesis to all of you. Thank you for sharing personal experiences as patients, I will always keep them in mind while practising as a speech therapist.

I want to thank my supervisor, Professor Yi Xu, who have taught me, guided me, inspired me and encouraged me for the last four years. I will always remember that I should aim high for anything I want to do in my life.

I want to thank my collaborators from National Taiwan University Hospital, Dr Lu Lu of the Department of Rehabilitation and Dr Chun Hwei Tai of the Department of Neurology. Without your help I would not be able to apply for the ethics approval or recruit the patients.

I want to thank my examiners, Barry Heselwood and John Swettenham, for all the suggestions and recommendations which made my thesis more than better. The special ‘thank you’ goes to Barry who always encouraged me to do research since I was an MA student in Linguistics. Barry, you really inspire me!

I want to thank my PhD colleagues, especially Albert Lee for taking care of me throughout this PhD programme, reminding me of important issues, and cheering me up whenever I was down. I missed you so much in this final year! I hope we can meet up soon as PhD doctors!

I want to thank my dear friends, Hylda, Frank. Let us share more happiness, tears, gossips, frustrations and challenges! For years to come, be prepared for more crazy research ideas from me!

I want to thank my family, in special my mom 邱莉娟, my uncle 邱志華 and my grandpa 邱元祥. Thank you for being my eternal support and shelter, and providing me unconditional TLC – tenderness, love and care.

I want to thank my husband, Costin Leu, who stood by me during the writing up of this thesis. Thank you for everything, my love. From now on, you will have the opportunity to know a different me: without the writing-up pressure!

Finally I want to thank God. Thank You for giving me the strength and making everything happen at the right time. Under Your guidance I may make the impossible to happen, and this PhD thesis is the best proof!

Contents

List of tables

List of figures

Chapter 1 Literature review

1.1 Prosody

When listening to any speaker talking, either in one's native or in an unfamiliar language, listeners can usually perceive and distinguish some small nuances characterising phrase/sentence boundary, emotion, attitude, etc., which allow them to judge the speaker's mental state and cultural background (Ohala & Gilbert, 1978). It seems to have 'something' going on behind the words and phrases, which may slightly complement or even radically transform the superficial meaning of the sentence, facilitating the comprehension of the verbal message by the listeners under a specific context. This 'something' may be some slight changes in loudness, length or pitch in a word or in the whole utterance, which are beyond the scope of phonemic or distinctive feature analyses (Fox, 2000).

In practice, these changes have a name – generally linguists would call it prosody. Derived from the Greek musical term 'prosodía' meaning 'sung accompaniment', prosody is the term designated to describe the melodic quality of the speech (Fox, 2000). Prosody encompasses suprasegmental aspects in speech, mainly including rhythm, stress, accent, and intonation (Chao, 1968; Fox, 2000; Ladd, 1996; Pierrehumbert & Hirschberg, 1990; Selkirk, 2005). Acoustically, these aspects result from a systematic interaction between three acoustic parameters – length, loudness, and pitch (Fox, 2000).

Length is the psychoacoustic measure of syllable duration, also considered as the least controversial among the segmental features, since speech signals are physical properties that can be objectively measured (usually in milliseconds) due to their finite duration (Ladd, 1996). Loudness is a perceptual concept whose acoustic correlate is intensity, which is determined by the amplitude of the acoustic wave. Thus the greater the amplitude, the louder the speech (Fox, 2000). Pitch is a psychoacoustic concept, reflecting both the articulatory and acoustic properties during the voice production – respectively the vocal fold vibratory rate and the speech signal periodicity, and the faster the vibration, the higher the pitch (Gussenhoven, 2001a; Ladd, 1996; Ohala, 1978; Vaissière, 2008). Yet, prosody cannot be realised simply by manipulating length, intensity, or pitch in isolation, as these three acoustic parameters should be in balance to produce a harmonious natural-sounding prosody (Fox, 2000).

Prosody has been considered as suprasegmental features of speech, which are larger than segmental phonetic features, like vowels and consonants (Gussenhoven, 2001b). However, traditionally prosody has been studied as non-segmental features, and thus, regarded as the "residue" of the syllabic system – prosody is usually studied after the investigation of vowels and consonants has been conducted (Crystal, 1979). Under this view, the contribution of prosody to speech seems to be of marginal importance. Yet, under a communicative perspective, prosody also encodes distinctive meanings which can be categorised as contrastive communicative functions (Xu, 2006).

According to different speech contexts, prosody carries particular meaning and plays specifically three main functions: grammatical, pragmatic and emotional (Shriberg et al., 2001; Vaissière, 2008). Grammatical intonation incorporates all the suprasegmental features used to indicate syntactic information such as the sentence type (i.e. statement vs. question) and the phrasing units within the utterance, pragmatic intonation conveys new/old information, turn-taking and speaker's intents, and affective prosody stresses the speaker's emotion and attitude towards the listener and/or the speech content (Shriberg et al., 2001; Vaissière, 2008). The main focus of the present study is affective prosody, specifically the two aspects of emotion and attitude. More specific issues about emotion and attitude in prosody in the present study will be discussed in the section 1.4.

1.1.1 Prosody and lexical tone

Prosody is said to be present in every language, however, its analysis can be more complex in some languages than in others. This is due to the fact that the very same acoustic parameters can be employed at the same time to convey information at both word and sentence levels. For instance, pitch variation when used at word level assigns different meanings to words, but when used at sentence level conveys post-lexical information within the linguistic structure (Duanmu, 2004; Ladd, 1996; Yip, 2002).

Tone is the use of pitch as a lexical contrastive feature, known as lexical tones. In tonal languages, such as Mandarin Chinese (MC), different lexical tones convey different lexical meanings to the same phoneme sequence in a syllable (Duanmu, 2004; Ladd, 1996; Yip, 2002), as illustrated in Table 1:

Table 1 The use of pitch to convey lexical meanings in Mandarin Chinese

High(H) or Tone 1	Rising (R) or Tone 2	Low (L) or Tone 3	Falling (F) or Tone 4
H	LH	HLH	HL
mā	má	mă	mà
‘mother’	‘hemp’	‘horse’	‘to scold’

In contrast, words can be spoken with different pitch contours, while lexical meanings remain intact (Duanmu, 2004; Ladd, 1996; Yip, 2002). The use of pitch to convey communication functions at sentence level ("postlexial" meaning) is known as intonation, as illustrated in the Table 2:

Table 2 The use of pitch to convey postlexical meanings in intonation languages

High (H) Low (L)	LH
dog	dog
neutral	question

Regarded as a prosodic aspect, intonation is the linguistic term of pitch variation at suprasegmental level to convey post-lexical information within the linguistic structure (Ladd, 1996). While tone uses the very same pitch variation to assign different lexical meanings to words (Yip, 2002). Although playing distinctive linguistic roles, acoustically tone and intonation are two linguistic entities encoded into speech through the same acoustic parameter (pitch), and physiologically controlled by the same speech mechanism (Duanmu, 2004; Gussenhoven, 2004; Xu, 1997, 1999). Therefore, the analysis of prosody in tone languages has been a challenge, and the effect of interaction between tone and prosody cannot be rule out (Duamu, 2004).

Tone sandhi phenomena in MC serve to illustrate the fact that in practice lexical tone and prosody interact, instead of being two independent entities. In MC, the rather well-defined and stable four lexical tones can undergo variations under specific phonological contexts, called tone sandhi, which is also one of the evidences for prosodic interference in the lexical tones, since it does not occur in isolated words (Duanmu, 2004; Shen, 1990; Xu, 1997). However, tone sandhi only accounts for a part of the prosodic system in MC, and other prsodic aspects, i.e. yes-no question and focus, cannot be explained solely by tone sandhi system, especially when the specific phonological contexts do not occur in the sentence.

To better understand prosody in MC, Shen (1990) used stimulus sentences composed by words all with identical lexical tones, and had native MC speakers produced them. The mean pitch of each sentence was measured in four points – starting point, highest peak, lowest trough and ending point, and extracted for acoustic analysis, then based on the findings, she proposed three basic prosodic patterns in MC (Ladd, 1996; Shen, 1990):

- Tune I: starting with a mid pitch, moving toward reaching the highest peak at a mid-high pitch, then falling to end at a low pitch. In short, it has non-expended pitch range with final lowering. Tune I is the intonation pattern for statements.
- Tune II: starting with a mid-high pitch, moving toward reaching the highest peak at a high pitch, then dropping a little bit to end at a high pitch. In short, it has expended pitch range with final rising. Tune II is the intonation pattern for yes/no questions.
- Tune III: starting with a mid pitch, moving toward reaching the highest peak at mid-high pitch, then falling to end at low pitch. In short, it has expended pitch range with final lowering. Tune III is the intonation pattern for wh-question.

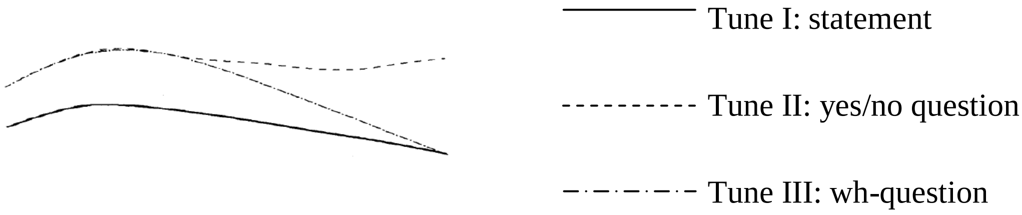

Figure 1. *Basic tunes of Mandarin Chinese (adapted from Shen, 1990)*

Similar research procedure has been adopted by Liu & Xu (2005), except for the number of measuring points for pitch – increasing from four in the whole sentence to ten in each syllable, and their findings were consistent with Shen (1990)’s – they found the greatest difference between statements and questions resided in the pitch across the whole utterance, which is higher in questions than in statements; and the main difference between yes/no questions (i.e. Shen’s Tune II) and wh-questions (i.e. Shen’s Tune III) is the pitch rise at the final syllable, which is only present in the yes/no questions (Liu & Xu, 2005).

In another study comparing the different effects of focus on statements and questions in MC, Liu and Xu (2005) found that the position of the focus influenced pitch contour of the statements and questions – pitch register of the whole sentence has been elevated with focus in the initial position. When the focus was located in the middle of the sentence, pitch register started to elevate from this point, and the difference between statements and questions emerged after the focus (Liu & Xu, 2005). The effect of the focus at the final position was similar to those without focus, showing the greatest differences between statements and questions in the last syllable (Liu & Xu, 2005). In terms of pitch range, yes/no questions were those with the greatest pitch range raise among other types of questions (wh-questions and confirmation questions) (Liu & Xu, 2005).

Note that MC in the studies by Liu and Xu (2005) and Shen (1990) refers to Beijing Mandarin (BM), the standard version of the language spoken in China. Yet, other variants of MC may differ from BM in terms of syntax, phonology and phonetics (Xu, Chen, & Wang, 2012). For instance, research comparing prosody of Taiwan Mandarin (TM), a variant of MC spoken in Taiwan, to prosody of BM and found that speakers of TM and BM expanded intensity and syllable duration on the focused word, but only BM speakers produced post-focus lowering of pitch (Chen, Xu, & Guion-Anderson, 2014; Xu et al., 2012). Therefore, findings in prosody of BM cannot be generalised to other variants of MC, and future studies in MC should be more specific about which variant of MC is being investigated.

1.2 Prosody in speech and language disorders

Prosody still remains a largely unexplored area in speech and language disorders, as it is often put aside amid ‘less important’ or paralinguistic aspects of speech. In general, speech clinicians believe that improvement of respiration, resonation, phonation and articulation will automatically facilitate prosodic skills, thus little time and effort are dedicated to diagnosis and treatment of prosody disturbances (Dworkin, 1991; Peppé, 2009). For those clinicians willing to include prosody as a target in intervention, the primary difficulties they may encounter are issues concerning the nature of prosody, measurements for evaluation of prosodic skills, transcription and description of prosodic disturbances, and the interpretation of prosodic characteristics in the face of the underlying aetiology (Crystal, 2009; Peppé, 2009).

When talking about prosody in speech, those who are used to linguistic jargons may briefly list stress, intonation, rhythm, accent and tone (Fox, 2000). In practice, there is no consensus about what is or is not prosody, and this list could be much longer. Prosody is considered as suprasegmental, as it includes all speech aspects that are not directly associated with the articulation of vowels and consonants (Gussenhoven, 2015). However, this definition is too broad for the documentation of symptoms in the clinical setting. In order to facilitate a concise description of prosodic deficits in motor speech disorders, Darley, Aronson and Brown (1969) proposed an assessment approach for evaluating speech problems in patients with motor speech disorders, based on an acoustic study of over 200 patients in seven different neurogenic groups.

In this classical study, Darley and associates measured 38 dimensions divided into seven categories – pitch, loudness, voice quality, respiration, prosody, articulation and intelligibility. As for prosody, they included ten dimensions, mainly concerning rate, stress and intervals/silence (Darley et al., 1969). Although its concept of prosody differs from linguistic perspectives, this study remains the most influential research in prosodic disorders until the present moment, cited by major text books in speech therapy, and kept as the gold standard by researchers and clinicians (Duffy, 2005; Dworkin, 1991; Nuffelen, 2011). Furthermore, these dimensions have always been perceptually judged by the clinicians, who evaluated and documented the dimensions as present or absent from symptoms of the patient (Darley et al., 1969; Duffy, 2005; Dworkin, 1991; Nuffelen, 2011).

Despite the assiduous attempt of Darley, Aronson and Brown (1969) in establishing standards for the diagnosis and treatment of prosodic impairments in motor speech disorders, the understanding of prosody in other speech and language disorders is still very limited. In order to provide an evidence-based normed assessment tool, Peppé and McCann (2003) created the Profiling Elements of Prosodic Systems – Child version (PEPS-C) – an exclusive assessment tool for evaluating prosodic skills which assesses prosody in four main aspects – interaction, affect, chunking and focus, and has been normed on 120 children aged between 5 years and 14 years. The PEPS-C has been tested in different developmental language disorders, such as autism, specific language impairment, pragmatic language impairment, hearing impairment and stuttering (Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2006; Peppé & McCann, 2003), and translated into five European languages – English, Spanish, French, Flemish and Norwegian (Peppé et al., 2010).

Although Darley, Aronson and Brown (1969) and PEPS-C were designed for different populations, both assessment tools were scored based on the perceptual judgment of the examiner, who has to make a binary choice, indicating if a certain item is present or absent in the clients'/patients' prosody. This widespread practice of evaluating prosodic skills, based on perceptual judgment, has further implications, as it is common to see in conditions of very distinct aetiology, such as Autism spectrum disorder (ASD) and Parkinson’s disease (PD), a sharing of similar speech characteristics: deviant prosodic patterns, monotonous/flat speech, difficulties in assigning stress, more difficulties in expressing emotions through prosody than in conveying linguistic functions (Cheang & Pell, 2007; Diehl, Watson, Bennetto, Mcdonough, & Gunlogson, 2009; Hubbard & Trauner, 2007; Möbes, Joppich, Stiebritz, Dengler, & Schröder, 2008; Paul, Augustyn, Klin, & Volkmar, 2005; Shriberg et al., 2001).

In order to verify the authenticity of these shared speech characteristics, the main goal of the current study is to systematically examine and describe prosody skills and deficits of two very distinct conditions – ASD, a well-known developmental language condition, and PD, a common acquired motor speech disorder. In the next sections, clinical diagnostic criteria and research findings regarding prosody of individuals with ASD and PD will be explained comprehensively.

1.2.1 Prosody in autism spectrum disorder

ASD is a lifelong neurodevelopmental condition often diagnosed before the age of 3 years, and characterised by two major clinical symptoms – ‘*deficits in social communication and social interaction*’ and ‘*restrictive repetitive behaviours, interests, and activities*’ (American Psychiatric Association, 2013). Included since 1980 in the Diagnostic and Statistical Manual of Mental Disorders, Third Edition (DSM-III), and revised in its Fourth Edition (DSM-IV-TR), Autism spectrum disorders (ASDs) included five subtypes – autistic disorder, Asperger’s syndrome (AS), pervasive developmental disorder not otherwise specified (PDD-NOS), Rett’s disorder, and childhood disintegrative disorder (CDD) (B. K. Shapiro, Menon, & Accardo, 2008).

In the latest version published in 2013, the DSM-5, ASD now includes only four out of five previous subtypes in the DSM-IV-TR – autistic disorder, AS, PDD-NOS and

CDD. The major innovation is, instead of using the collective term ASD for separate disorders, all subtypes are now considered as one single condition with different degrees of severity (American Psychiatric Association, 2013). With this modification in the diagnostic criteria, the long-time dispute about differential diagnosis between high-functioning Autism (HFA) and AS could be finally solved. Before the publication of DSM-5, these two disorders shared all the clinical features, and were differentiated only by the history of language delay, i.e., AS was characterised with no language delay before the age of 3 years, which, in practice, was frequently reported retrospectively by the parents or care givers, therefore, it was subject of inaccurate memory or biased judgement (Woodbury-Smith, Klin, & Volkmar, 2005). With the new DSM-5 criteria, there will be only different degrees of severity, and thus issues concerning subjectivity in differential diagnosis for AS and HFA may be at last ruled out.

According to studies in Europe, Asia and North America, ASD has an estimated prevalence rate of 1 to 2%, thus making it the most common neurodevelopmental disability, which causes speech and language disorders in the present day (Centers for Disease Control and Prevention, 2015; Hirtz et al., 2007). For instance, the estimated prevalence of ASD is 157 per 10,000 children aged 5 years to 9 years in the U.K. (Baron-Cohen et al., 2009), and 57 to 210 per 10,000 children aged 8 years in the United States (Centers for Disease Control and Prevention, 2014). However, the estimates in China and Taiwan are much more modest, only 11.8 per 10,000 in children aged 2 years to 14 years in China, and 2.4 to 17.3 per 10,000 in children aged 0 year to 17 years in Taiwan (Lin, Lin, & Wu, 2009; Sun et al., 2013). The lower prevalence in China and Taiwan may result from diagnostic criteria, availability of services, public awareness, and methodological flaws, as studies in ASD prevalence has always been very scarce in these countries (Cubells, 2013; Lin et al., 2009; Sun et al., 2013).

The cognitive and linguistic abilities within ASD vary drastically. According to the estimates, around 20-45% of individuals with ASD display functional speech and have above boundary-line IQ (above 70) as shown in results obtained from standardised cognitive tests (i.e. Wechsler Intelligence Scale for Children, Raven's Standard Progressive Matrices, and Coloured Progressive Matrices) (Charman et al., 2011; Edelson, 2006; Jazen, 2003; Paul, Bianchi, Augustyn, Klin, & Volkmar, 2008). These verbal individuals with ASD who have above boundary-line IQ are often regarded as high-functioning, and their diagnoses correspond to the DSM-IV-TR criteria for HFA, AS or PDD-NOS (Edelson, 2006; Jazen, 2003) and to the DSM-5 criteria for a mild form of ASD (American Psychiatric Association, 2013).

In terms of language skills, 25% to 50% of individuals with ASD are non-verbal, while virtually all verbal individuals with ASD have language impairments of various severity degrees in different aspects of language, either phonology, semantics, syntax, pragmatics or prosody (Edelson, 2006; Eigsti, de Marchena, Schuh, & Kelley, 2011; Jazen, 2003; Paul et al., 2008). Thus, even though high-functioning individuals with ASD have remaining speech skills, they often display impairments in verbal communication. Among speech-related problems, peculiarities in expressive prosody have been the focus of various studies in the speech of individuals with ASD ever since the term 'autism' was coined. Asperger and Kanner, the two pioneers in ASD, have portrayed expressive prosody of their child and adolescent subjects with ASD as 'monotonously singsong' like with an 'odd intonation' (Asperger, 1944/1990; Kanner, 1943). Although these children had good articulation and were able to imitate wording and intonation, they habitually used utterances out of context (Kanner, 1943).

Listeners, general public or clinicians, often report that they perceive individuals with ASD as speaking with less pitch variation, thus their prosody is usually regarded as sounding flat or monotonous (Bonneh et al., 2011; Nakai et al., 2014). However, results from instrumental acoustic analysis indicated that this population tends to speak with larger pitch range and pitch variation (Bonneh et al., 2011; Green and Tobin, 2009; Nakai et al., 2014). Further empirical evidence also suggested that individuals with ASD seem to rely almost exclusively on pitch to convey prosody, without sufficiently exploiting amplitude and duration cues (Diehl et al., 2009; Hubbard & Trauner, 2007). Therefore, their deviant prosody may have its origin in the imbalanced use of acoustic parameters, rather than the no-use of these parameters (Diehl et al., 2009).

Furthermore, by contrasting results from listener judgment and acoustic analysis, Nadig and Shaw (2012) found that speech and language therapy students did not judge increased pitch range of child and adolescents speakers with ASD as having larger pitch variation. Moreover, this increased pitch range did not facilitate listeners' comprehension, as these raters consistently judged speech prosody of speakers with ASD as sounding more atypical than their typically developing peers (Nadig & Shaw, 2012). They suggested that the unnecessary increase of pitch variation may lead these respondents to perceive ASD speech as of flatter prosody (Nadig & Shaw, 2012), but they did not offer any empirical evidence to support this speculation. Sharda and colleagues (2010) further associated increased pitch range and pitch variation in speakers with ASD with 'motherese', i.e., the way that mothers speak to their young children, and concluded that their prosody deficits may result from developmental language delay.

Deficiencies in expressive prosody have been considered as one of the major speech problems in ASD, along with poor pragmatic skills and faulty interactional discourse (Paul et al., 2008; Shriberg et al., 2001). Atypical intonation may elicit negative response in listeners, and thus impose even more obstacles in the social acceptance of individuals with ASDs (Paul, Shriberg, et al., 2005). Empirical evidence suggested that the prosodic performance and social communication competence are positively related in ASD; individuals with ASD who have abnormal prosody production also have lower scores in social and language tests (Depape, Chen, Hall, & Trainor, 2012; Nakai et al., 2014; Paul, Shriberg, et al., 2005).

1.2.2 Prosody in Parkinson's disease

PD is a progressive neurodegenerative condition caused by the reduction of the dopamine level in the basal ganglia circuit. This reduction is caused by the loss of neurons in the substantia nigra, a brain area responsible for the dopamine production (Edwards, Quinn, & Bhatia, 2008; Fahn J. Jankovic & Hallett, 2007; Lehéricy, Sharman, Santos, Paquin, & Gallea, 2012). As dopamine is a neurotransmitter that facilitates voluntary movements, its reduction decreases the voluntary movements and/or the inhibition of involuntary movements, resulting in a movement disorder with three canonical symptoms – bradykinesia (slow initiation of voluntary movements with gradually reducing speed and amplitude of sequential movements), resting tremor, and muscular rigidity (Edwards et al., 2008). These symptoms affect all body muscles, roughly with the same strength.

Among the elderly population, PD is the second most common neurogenic condition, only after Alzheimer disease (Hirtz et al., 2007). Based on meta-analysis of data from Europe, Asia, Australia, and North and South Americas, the prevalence of PD increases with age – 41 per 100,000 people aged 40 years to 49 years, 107 per 100,000 people aged 50 years to 59 years, 173 per 100,000 people aged 55 years to 64 years, 428 per 100,000 people aged 60 years to 69 years, 425 per 100,000 people aged 65 years to 74 years, 1,087 per 100,000 people aged 70 years to 79 years, and 1,903 per 100,000 people aged older than age of 80 years (Pringsheim, Jette, Frolkis, & Steeves, 2014).

In the developmental trajectory of the illness, the PD canonical motor symptoms gradually manifest in the voice and speech of individuals with PD, affecting all the levels of the speech mechanism – respiration, phonation, resonance and articulation (Duffy, 2005; Dworkin, 1991; Penner, Miller, Hertrich, Ackermann, & Schumm, 2001; Sapir, Ramig, & Fox, 2008). From two studies involving more than 200 PD patients, the incidence rate of speech and voice problems is 70% to 89%, yet only about 2% of PD patients receive speech therapy (Hartelius & Svensson, 1994; Logemann, Fisher, Boshes, & Blonsky, 1978). Major accounts for voice and speech problems in PD include: slow individual articulations, associated with bradykinesia in the jaw movements; accelerated and imprecise repetitive articulations, resulted from rigidity in jaw movements; reduced and/or monotonous loudness and pitch variations, caused by the rigidity and bradykinesia in the thyroarytenoid and cricothyroid muscles; unsteady voice during prolonged vowel production, linked to the resting tremor; and increasing breathiness and hoarseness, as consequences of the rigidity and bradykinesia in thyroarytenoid muscles (Duffy, 2005; Penner et al., 2001; Sapir et al., 2008).

Voice, articulation and prosody deficits are the most prominent features in the speech of PD patients (Duffy, 2005). Their voice has been described as breathy and with reduced intensity, articulation as imprecise and with slower diadochokinesis rate, and prosody as having less pitch and loudness variations (Ma, Whitehill, & Cheung, 2010; Midi et al., 2008; Rigaldie, Nespoulous, & Vigouroux, 2006; Stewart et al., 1995; Whitehill, Ma, & Lee, 2003). Moreover, all these features together contribute to the listener’s impression regarding PD patients as sounding less happy, friendly, interested, and involved than healthy elderly speakers (Jaywant & Pell, 2010). It has been believed that all these deficits were intrinsically associated with the motor symptoms of PD (Duffy, 2005). Yet, empirical findings suggest that some speech deficits experienced by

PD patients might have causes beyond the motor disorder, as not all aspects of speech were equally affected in PD (Möbes et al., 2008).

1.3 Prosody in the typical populations

The speech production mechanism is a dynamic system which undergoes continuous physiological changes from childhood to late adulthood. During the developmental course, not only the syntax structure and pragmatic skills improve, but also the motor control over articulators masters over time. The acquisition of prosodic skills seems to be intrinsically related with maturation of articulatory control (Patel & Grigos, 2006; Wong, Schwartz, & Jenkins, 2005). Then the motor control gradually deteriorates with the weakening and stiffening of the muscles during the ageing process (Linville, 2004). In this section, issues concerning prosodic changes in developmental and ageing trajectory will be described and discussed.

1.3.1 Prosody in development

From a developmental perspective, prosody follows some major markers in language development. Research findings suggest that children start acquiring prosodic patterns very early in life, even before producing any word at all (Li & Thompson, 1977). Observations based on perceptual judgements suggest that, from the birth to the age of 1 year, children’s vocalisations evolve from biological reflexive crying, differential crying, vocal play, babbling, to proto-phonology (Crystal, 1979; DePaolis, Vihman, & Kunnari, 2008; Kent & Murray, 1982; Papaeliou & Trevarthen, 2006; Stark, 1979). When reaching the age of 1 year, children are able to respond to adults’ speech in conversation-like interactions (Crystal, 1979; Stark, 1979). At the end of the age of 1 year, a reorganisation period occurs and falling patterns are acquired (Snow, 2006). Between the ages of 1 year and 2 years, diverse prosodic patterns emerge in children’s vocalisation according to the prosodic system of their mother tongue (DePaolis et al., 2008). Finally, at the age of 2 years, children start acquiring rising patterns in contrast to falling patterns (Snow, 2006).

The acquisition of prosody seems to be associated with the maturation of the articulatory movement control (Patel & Grigos, 2006; Wong et al., 2005). For instance, in English-speaking children, articulatory coordination and prosodic control mature throughout the childhood, evolving from variable articulatory movements at the age of 4 years, to consistent movement patterns to contrast different prosodic patterns, such as statements and yes/no questions, at the age of 7 years (Patel & Grigos, 2006). In terms of acoustic parameters, children at the age of 4 years seem to rely mostly on duration to contrast questions from statements, while children at the age of 11 years master the manipulation of pitch to convey the same contrast with proficiency similar to adults (Patel & Grigos, 2006).

As a general rule, most of the prosodic skills seem to be mastered around the age of 5 years, yet some of them may continue to develop through the school years till reaching adult-like proficiency (Wells, Peppé, & Goulandris, 2004). For example, studies found that children at the age of 5 years still have difficulties assigning stress in two-word phrases, whereas a question’s final rising pattern is fully mastered only by the age of 7 years or 8 years (Patel & Brayton, 2009; Wells et al., 2004). Although children of same age may master various prosodic skills at a different pace, most of them enter adolescence with adult-like prosodic skills, when the maturation of the control and coordination of the articulators and laryngeal muscles is mainly achieved (Wells et al., 2004; Wong et al., 2005).

Moreover, research findings suggest that children’s lexical tone acquisition and prosodic development may coincide in language development. Li and Thompson (1977) found that Mandarin Chinese-speaking children start acquiring tone between the ages of

1 year and 2 years, which is accomplished in a rather brief time interval, before even acquiring the full phonological repertoire. Tone development, just like prosodic development, also follows a trajectory. Young children usually acquire first high tone, followed by falling, rising, and low tones respectively (Li & Thompson, 1977). By the age of 3 years, Mandarin Chinese-speaking children are able to produce stable tones, with the exception of low tone (Wong et al., 2005). The association of articulatory control and tone production in MC-speaking children is evidenced by a relatively late acquisition age of low tone, i.e., after the age of 3 years (Wong et al., 2005). There are at least two issues associated with this later acquisition: first, the low tone in MC is a contour tone, which first falls then rises; second, as the production of contour tone demands more precise motor control of the articulators, children at the age of 3 years may not have mastered the control and coordination of laryngeal muscles to sustain vocal fold tension and produce stable pitch throughout the syllable (Wong et al., 2005; Yip, 2002).

However, this description of developmental course of vocalisation is rather controversial, as typically-developing children of the same age may master various prosodic skills at a different pace (Kent & Murray, 1982; Wells et al., 2004), yet they usually acquire the whole prosodic system by the end of childhood, when entering adolescence (Wells et al., 2004). In contrast, children with developmental speech-language disorders may follow an atypical prosodic development trajectory, and often produce disordered prosody characterised by patterns that differ from the conventional ones, which, thus, compromises their speech intelligibility (Crystal, 1975).

1.3.2 Prosody in ageing

From young adulthood to old age, people experience changes in respiration, voice and speech that are consequences of ageing. For instance, older people often have decreased lung pressure, shallower respiration during speech and reduced control on vocal loudness, as a result of stiffening of the thorax and weakening of respiratory and laryngeal muscles (Baker, Ramig, Sapir, Luschei, & Smith, 2001; Linville, 2004). Lowering of mean pitch is a common voice profile in elderly women, which is associated with the thickening of laryngeal mucosa, yet elderly men often experience raising of mean pitch (Linville, 2004; Mwangi, Spiegl, Haderlein, & Maier, 2009; Torre & Barlow, 2009). Old adults also have increasing breathiness in voice quality due to higher incidence of anterior glottal gap (Linville, 2004). In addition, laryngeal lowering in the neck, degenerative processes in oral structures, and reduction of articulatory precision contribute to altered resonance patterns, decreasing the speech intelligibility of elderly people (Linville, 2004; Torre & Barlow, 2009).

However, these general ageing changes in speech and voice seem to have marginal effects in the performance of expressive prosody by old adults. Tauber, James, and Noble (2010) found that older speakers were as capable as young adults to use prosody for producing intonational boundary and conveying different meanings of ambiguous sentences. Furthermore, the prosodic production of the elderly in parsing ambiguous sentences was more correctly rated by listeners than the performance of younger adults (Tauber et al., 2010). The researchers concluded that listeners may have taken advantage of slower speech of older speakers when judging the meaning of the utterances (Tauber et al., 2010).

Besides speech production, speech comprehension is also affected by the normal ageing process, as working memory capacity, hearing acuity and neural processing speed decline significantly with increased age. Past research exploring prosody perception found that older people often performed poorer than young adults in emotion identification tasks. For instance, Lima, Alves, Scott, & Castro, (2014) showed that older adults were less accurate than younger adults when identifying vocal emotions, but this performance disparity was not associated with the emotional valence. The findings in Dupuis and Pichora-Fuller (2015) confirmed the tendency of older people making more mistakes than younger people when judging emotion from semantically neutral sentences. Furthermore, Paulmann, Pell and Kotz (2008) demonstrated that adults with a mean age of 42 years also presented similar deficits in perception of speech emotion, thus they concluded that the decline in vocal emotional perception may have started as early as middle age.

One may argue that the auditory acuity plays a major role in speech perception of ageing populations, as older people often present some age-related hearing loss. Yet, Dupuis and Pichora-Fuller (2015) found no significant correlations between the vocal emotion identification accuracy and measures of auditory acuity. Furthermore, Mitchell and Kingston (2014) found that emotion perception of their elderly participants is more closely associated with pitch discrimination ability, rather than with loudness and duration. Therefore, at least for vocal emotion comprehension, age-related hearing loss seems not the only reason for explaining older adults' poorer performance.

1.4 Emotion and attitude in prosody

Emotional prosody is present in every language, and the patterns of its contours differ subtly from language to language, with some differences resulted from cultural influences. Although emotional prosody is rather universal among languages, it is hard to find one-to-one emotion-prosody correlates, because it can vary significantly from speaker to speaker, due to the fact that there are no two speakers with the same degree of reaction for the very same emotion (Wells, Peppé & Goulandris, 2004).

Conventionally, vocal expression of emotion has been described and compared in terms of the three basic acoustic parameters – pitch, intensity and duration. For instance, Williams & Stevens (1972) in their classical research described the pitch variation under four emotional states – neutral, anger, fear and sorrow, in an experiment performed by actors. They discovered that: 1) anger has higher pitch (1/2 octave above), greater pitch range, increased intensity, voicing irregularity (weak first formant), extreme articulation; 2) fear has pitch similar to neutral emotion, with occasional peaks, unusual shape of pitch contour at the peaks (bumps and breaks), longer duration, precise articulation; and 3) sorrow has lower pitch, narrower pitch range, decreased articulatory rate, longer duration (pause), whisper voice. However, this type of research paradigm, which merely describes the average values of acoustic changes in different emotional expression, is not sufficient for establishing a prosodic pattern for a specific emotion, as it does not adopt any theory of emotion, which gives a definition for emotion and serves as the guidance for the interpretation of results (Scherer, 2003).

One of the biggest challenges in the research field of emotion and attitude is the conceptual definition of “emotion” and “attitude” (Olufemi, 2012; Shiota & Kalat, 2012). Although being a part of our everyday vocabulary, it seems that there is no clear definition for emotion and attitude in scientific research (Mulligan & Scherer, 2012). Traditionally, emotion has been studied in many different ways – some researchers have regarded emotions as distinct mental states, some have considered them as different physiological responses to outer stimuli, some have studied them as cognitive processing, while others have investigated them as social attributes (Manstead, 2005; Parkinson, 1996; Tracy & Randles, 2011). In the current study, both emotion and attitude are viewed as an adaptive mechanism for an organism to deal with ever-changing environmental stimuli by using appropriate physiological and psychological actions (Scherer, 1984).

Traditionally, emotional prosody has been investigated alongside attitudinal prosody under the label of 'affective prosody' for both being conveyed by similar variations in pitch, loudness and duration (Mitchell & Ross, 2013). However, this practice is rather problematic, as emotion and attitude are two concepts that although they largely overlap, each of them has its own nature and dynamics. Emotions are innate, whereas attitudes are socially learnt (Olufemi, 2012; Shiota & Kalat, 2012). Emotions are episodic, as they have a beginning and an end, yet their duration is hard to be defined, while attitudes are relatively stable but can be modified (Mulligan & Scherer, 2012; Olufemi, 2012). Both emotion and attitude are reactions, feelings and beliefs towards a person, an event or an object that must be external, or outside of the individual who is experiencing the emotion or attitude (Olufemi, 2012; Shiota & Kalat, 2012). Both emotion and attitude have identical components – cognitive appraisal of the situation, subjective feeling towards the experience, neurophysiological changes during the event, and behaviours and actions motivated by the emotion or attitude (Mulligan & Scherer, 2012; Olufemi, 2012; Scherer, 1984; Shiota & Kalat, 2012).

In speech prosody, empirical research has demonstrated that happiness (emotion) and friendliness (attitude) have similar prosodic profiles, while they differ mainly in terms of the values of the parameters along the same acoustic dimensions. These findings suggest that speakers probably avoid sounding too happy when communicating friendliness, since this might be socially inappropriate (Noble & Xu, 2011), as attitudes are more subtle and are said to be more socially controlled than emotions (Scherer, Banse, & Wallbott, 2001; Scherer, 2003).

In the sections 1.4.1 and 1.4.2, research findings in production and perception of emotional prosody regarding individuals with ASD and PD patients will be presented and discussed. Yet, the focus of discussion will be directed to emotional prosody, as until the present moment there is virtually no study examining directly attitudinal prosody in ASD or PD.

1.4.1 Perception of emotional prosody in ASD

Further empirical evidence suggested that deficits in emotional prosody of individuals with ASD might extend beyond speech production, affecting also their competence in decoding prosody. As early as in his first publication on ASD in 1943, Kanner portrayed his 11 child patients as being unable to ‘form affective contact with people’ (Kanner, 1943). Although emotional impairment and difficulties in perceiving emotion through facial expressions and voice are clearly reflected in these individuals’ behaviours, supporting empirical evidence of emotional impairment is often controversial and subtle (Begeer, Koot, Rieffe, Meerum Terwogt, & Stegge, 2008; Heaton et al., 2012).

Past research investigating perception and expression of emotion has found that individuals with ASD cannot use facial and vocal cues to identify emotions at the same level of proficiency as age-matched normal controls (Begeer et al., 2011; Frith & Happé, 1999; Holroyd & Baron-Cohen, 1993; Spek, Scholte, & van Berckelaer-Onnes, 2010). Some of the studies also found the correlation between their emotional performance and their deficits in Theory of Mind (ToM) (Baron-Cohen, 2000; Kristen, Rossmann, & Sodian, 2014). ToM is the cognitive ability that people are equipped with to understand and infer their own and others’ thoughts, beliefs, attitudes, emotions, etc. (Apperly, 2012; Premack & Woodruff, 1978). However, this correlation is not sufficient to establish a cause-effect relationship between emotional abilities and ToM, as understanding the mechanism and dynamics of emotion is not within ToM’s scope.

Deficits in emotion processing in ASD seem to be pervasive across different sensory domains, with various levels of severity (Globerson, Amir, Kishon-Rabin, &

Golan, 2015; Heaton et al., 2012; Philip et al., 2010). However, the overall findings in perception of facial and vocal emotion of this population are mixed, ranging from significantly poorer performance to compatible competence. In studies where the differences are found between the experimental and the control groups, the deficits of individuals with ASD are usually emotion- and/or task-specific. For instance, in facial emotion judgment, when compared to normal controls, the difficulties of individuals with ASD usually emerge in identifying negative valence emotions (sadness, anger and fear), perceiving complex emotions (guilt, shame and envy), detecting more subtle emotional expressions, and recognising emotions under brief exposure to stimuli (Ashwin, Chapman, Colle, & Baron-Cohen, 2006; Baron-Cohen, Wheelwright, & Jolliffe, 1997; Harms, Martin, & Wallace, 2010; Rump, Giovannelli, Minshew, & Strauss, 2009).

Similarly, the findings in perception of emotional prosody are also equivocal. Results from the Reading the Mind in Voice Test, an assessment tool designed to capture subtle deficits in ToM, show that, when compared to typical controls, high-functioning adults with ASD perform inferiorly in tasks involving identifying the emotion in the recorded utterances (Golan, Baron-Cohen, Hill, & Rutherford, 2007; Rutherford, Baron-Cohen, & Wheelwright, 2002). Likewise in studies following the six basic emotions paradigm – happiness, sadness, anger, surprise, fear and disgust (Ekman & Friesen, 1976), adults with ASD also have poorer judgment in emotion through speech and non-verbal vocalisations when compared to their typical peers (Globerson et al., 2015; Heaton et al., 2012). Further research found evidence implying that the difficulties of individuals with ASD in judging emotion through prosody might be selective. For instance, they seem to be particularly challenged while identifying happiness and anger from prosodic cues.

Conversely, in tasks for identifying emotions in meaningless utterances, pseudo-sentences and low-pass filtered sentences with eliminated verbal content, no significant group differences are found between children and adolescents with ASD and age-matched controls (Brennand, Schepman, & Rodway, 2011; Grossman, Bemis, Plesa Skwerer, & Tager-Flusberg, 2010; Le Sourn-Bissaoui, Aguert, Girard, Chevreuil, & Laval, 2013). Some empirical evidence suggests that the focus of attention of children and adolescents with ASD play a main role in the performance on the tasks. They are able to apply their knowledge in emotion once they are focused on the prosody of the utterances (Woynaroski et al., 2013); they are also less influenced by the emotional valence of the utterance, while their typical peers tend to prefer those with positive emotions (Brooks & Ploog, 2013).

Recently Wang and Tsao (2015) examined the association between emotional prosody perception and pragmatic language skills in school-aged high-functioning children with ASD in Taiwan by using recordings from a professional actress as auditory stimuli. Their results showed children with ASD have selective impairment in emotional prosody perception: they were as good as their age-matched peers in identifying the negative emotions (sadness and anger), but experienced difficulties with the positive ones (happiness) (Wang & Tsao, 2015). These results suggested children and adolescents with ASD who are speakers of other languages rather than English also experience challenges when understanding prosody.

1.4.2 Production of emotional prosody in ASD

Prosodic impairments in individuals with ASD seem to be pervasive across different functions of prosody – grammatical, pragmatic and emotional (McCann & Peppé, 2003). Nevertheless, their difficulties with emotional prosody have received ample attention from researchers and clinicians, mainly because of the widespread belief that individuals with ASD are unable to ‘form affective contact with people’ (Kanner, 1943), as well as they are ‘mind blind’ (Baron-Cohen, 1990; Frith, 2001). Furthermore, nearly all individuals with ASD experience challenges in some, if not all, components of emotional competence – understanding, responding, perception and expression (Begeer et al., 2008). For instance, children with ASD can understand other’s emotion through cues extracted from situations and desires, but not from beliefs (Baron-Cohen, 1991). People with ASD are also reported to fail to regulate their emotions appropriately and effectively according to social context, their responses are frequently interpreted as evidence of high irritability, poor anger control, mood dysregulation and aggression (Samson et al., 2014). Deficits in emotion perception of individuals with ASD are found across different sensory domains with various severity degrees (Globerson et al., 2015; Heaton et al., 2012; Philip et al., 2010). Finally, compared to typically developing peers, children with ASD know when and how to show their emotions only when supported with enough contextual cues (Begeer et al., 2011).

However, the findings in emotional prosody production of individuals with ASD were incongruent. Studies reported the expressive emotional skills of this population varying from significantly poorer than average individuals to similarly proficient. This disparity seems to result from the use of different research methods involving a wide range of experimental tasks. For example, Hubbard and Trauner (2007) explored acoustic parameters of expressive emotional prosody in ASD. By using acoustic analysis and subjective judgement from independent blinded raters, they investigated the prosodic performance of children with ASD in imitative and spontaneous speech. Their findings indicated that, compared to average controls, these children have increased pitch range, and random loudness and syllable duration. Although children with ASD in this study could produce variations in these three acoustic parameters, their speech was perceived as abnormal by the raters (Hubbard & Trauner, 2007).

In order to establish a structured protocol for examining prosodic skills, Peppé and McCann (2003) developed the Profiling Elements of Prosodic Systems – Children, PEPS-C, to assess children's prosodic abilities in different communication areas – interaction, affect, chunking and focus, with the test outcomes perceptually judged by trained examiners. Past research using PEPS-C as a research tool often gave a brief description of affective prosody production of tested children with ASD. Diehl and Paul (2013) found children with ASD produced longer utterances to express their dislike, but no significant differences were found for acoustic parameters compared to the typically developed children. In contrast, Peppé et al. (2011) suggested that children with ASD might have difficulties in conveying emotions through prosody, since their performance sounded inaccurate to the examiners. Although PEPS-C is intended to assess various aspects of prosody, the performance of the tested children is evaluated by perceptual judgement of the examiner, thus the assessment results may be biased.

Chen and Liu (2010) conducted a study involving acoustic analysis of speech prosody of TM-speaking children with ASD. They used structured-storytelling with picture books to elicit emotional speech (neutral, angry, happy and sad). Mean pitch, pitch range, mean intensity and syllable duration were extracted from the recordings using the software Praat (Boersma & Weenink, 2013). Overall, they found no significant differences in performance of emotions between the experimental and the control groups. Yet, they described that children with high-functioning autism or Asperger syndrome had lower mean pitch when expressing anger and sadness, smaller pitch range for joy, anger and sadness, whereas their performance in intensity and duration was similar to their typically developing peers (Chen & Liu, 2010).

Beside what have been debated about the nature of prosodic impairments in ASD, these deficits seem to go hand in hand with social-communicative competence. Generally, individuals with ASD who show poorer performance on prosody production also score lower in social and language tests (Depape, Chen, Hall, & Trainor, 2012; Nakai et al., 2014; Paul, Shriberg, et al., 2005). Emotional competence, i.e. the expression, perception, responding to, and understanding of emotions, is essential for the development of functional social interaction skills, thus being able to recognise other people's emotions and respond accordingly are essential to establish social bonds (Begeer et al., 2008). Restricted emotional competence of individuals with ASD to extract relevant emotional information from contextual cues may restrain them from developing more complex social skills, and this deficit imposes challenges to their everyday social interaction (Da Fonseca et al., 2009), as the misinterpretation of others' emotions may lead to inappropriate responses, and impose more social barriers for the individual with ASD, preventing them from participating in and contributing to social interaction.

The incongruent findings in ASD can be explained partly by the complexity of the tasks used in the experiments. As a common rule, experiments designed for older adolescents and adults are usually more complex than those for children, thus the findings in emotion recognition is more mixed for older individual with ASD than children with ASD (Harms et al., 2010). Moreover, the group differences could also have resulted from different research methods. For example, compared to their typically developing peers, children with ASD perform poorer when identifying emotion from single word recordings, while there are no significant between-group differences when utterances are used as stimuli (Grossman et al., 2010; McCann, Peppé, Gibbon, O'Hare, & Rutherford, 2007; Paul, Augustyn, et al., 2005; Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007).

Another possible reason for the disparity in research findings is that studies in this field rarely incorporate any specific theory of emotion to motivate the research questions and hypotheses (Scherer, 2003). Past research in emotion perception of individuals with ASD either only reported the results without a clear theoretical framework (such as Globerson et al., 2015; Grossman & Tager-Flusberg, 2012; Heikkinen et al., 2010; Le Sourn-Bissaoui et al., 2013; Ploog, Brooks, Scharf, & Aum, 2014; Wang & Tsao, 2015), or exploited ToM as an attempt to make sense of the findings (such as Chevallier, Noveck, Happé, & Wilson, 2009; Golan et al., 2007; Rutherford et al., 2002). According to Baron-Cohen (2000), although some tests for ToM have been developed, they are not an exclusive means to diagnose ASD, as other neurogenic conditions – Parkinson's disease, Huntington's disease, multiple sclerosis, dementia, traumatic brain injury, just to name a few, might also develop deficit in ToM. Ultimately, ToM is not a theory exclusively for understanding emotions (Mitchell & Phillips, 2015).

In summary, the literature in perception and production of ASD is vast and abundant, yet past findings are contradictory. With a better examination of research methodology, most of the studies did not incorporate a specific research paradigm, therefore, it is not possible to formulate or test specific hypotheses that offer a more systematic view of emotional prosody deficits in individuals with ASD. This atheoretical research approach in studying vocal emotions could be problematic, for there were no hypotheses to be tested and guide the findings, which could then compromise the interpretation of outcomes (Scherer, 2003).

1.4.3 Perception in emotional prosody in PD

Contrary to James Parkinson's observational conclusion that PD individuals' "senses are uninjured" (Parkinson, 1817), the recent increasing empirical and clinical evidence suggests that the disorder may have already started long before the onset of motor symptoms (Halliday, Lees, & Stern, 2011; Langston, 2006). Based on their post-mortem study of brain tissues of PD patients, Braak and associates (2002) proposed a new staging system for PD – it consists of 6 stages associated to the presence of Lewy bodies in different areas of the nervous system with the clinical symptoms. Lewy body is an abnormal aggregate of presynaptic protein that develops inside neurons (William, Gregory, & Frederick, 2009). The presence of Lewy bodies in several brain areas indicates a more advanced stage of PD. In stage 1&2 PD, the involvement advances from periphery to medulla oblongata, resulting in olfactory deficits; in stage 3&4 PD, the brainstem is infiltrated, thus sleep and motor disturbances emerge; finally in stage 5&6 PD, the limbic system and neocortical areas are also compromised, causing cognitive impairments (Braak et al., 2002). Thus, the emergence of non-motor symptoms, such as olfactory dysfunction, constipation, REM sleep behaviour disorder, cognitive impairment, psychosis, anxiety, depression, and apathy, may also be disease markers of different stages in PD (Halliday et al., 2011; Langston, 2006).

Among the non-motor symptoms, apathy and depression are closely related as co-morbidities in PD, especially in patients of more advanced stages (Oguru, Tachibana, Toda, Okuda, & Oka, 2009). These two psychiatric conditions contribute to the lack of motivation, withdrawal behaviour and reduced emotional expression often found in PD patients (den Brok et al., 2015; Dujardin et al., 2007; Voss & Hegeman Richard, 2011), and they may compromise long-term outcomes for their negative impacts on the treatment, therefore, giving each of them the right treatment can improve the patient's quality of life and ease the caregiver's burden (den Brok et al., 2015).

Although the debates about the nature of emotional dysfunction in PD patients are still ongoing, some clinical researchers proposed that apathy involves the frontostriatal circuit, a brain area responsible for the executive function and cognitive ability, whereas depression is engaged with the limbic system, an area in the brain intrinsically linked to emotions (Blonder & Slevin, 2011; den Brok et al., 2015; Dujardin et al., 2007; Oguru et al., 2009; Voss & Hegeman Richard, 2011). Nevertheless, differential diagnoses may be challenging, for both present overlapping signs in PD, such as lack of motivation, withdrawal behaviour and reduced affective expression (den Brok et al., 2015; Dujardin et al., 2007; Voss & Hegeman Richard, 2011).

It is widely accepted nowadays among clinicians and researcher that PD also presents emotional dysfunction as a part of core symptoms, yet the empirical findings in emotion perception of PD patients were very divergent, with their performance level varying from compatible with to worse than the performance of average elderly. For instance, PD patients staging from 2 to 5 on Hoehn & Yahr's severity measure (1967) performed similarly in tasks consisting of labelling the emotional facial expression from a naming list (Adolphs, Schul, & Tranel, 1998; Pell & Leonard, 2005). Yet, PD patients staging from 1 to 3 had more difficulties in naming the emotional facial expressions portraying anger, sadness, disgust and fear, regardless of the expression’s intensity and position (upright or inverted) (Dujardin et al., 2004; Narme, Bonnet, Dubois, & Chaby, 2011).

Similarly, results in emotional perception in prosody were far from consistent. Mitchell & Bouças (2009) reported no emotional prosody decoding deficits in PD patients with stages 1 to 3 in tasks that involve emotion judgment in utterances with emotionally matched or conflicted semantic contents. Dara, Monetta, & Pell (2008) manipulated intensity in both meaningful sentences and pseudo-sentences to investigate the ability of PD patients with stages 1 to 4 to judge emotions as positive or negative by using. They found that compared to healthy elderly people, PD patients had more difficulties in rating anger, disgust and fear (Dara et al., 2008). Breitenstein, Van Lancker, Daum, & Waters (2001) asked patients with PD in stage 2 who were native English speakers but not fluent in German to judge emotions from German sentences. They found that PD with poorer cognition skills were more affected by the language of the utterance, and all PD patients performed worse than healthy older adults in tone and syllable duration discrimination tasks (Breitenstein et al., 2001).

There were also studies that used electroencephalography (EEG) to detect variations in event-related brain potential during the emotional prosody judgment task. (Shröder et al., 2006) used a single utterance ‘Ana’ produced with different emotions as auditory stimuli, and they identified decreased mismatch negativity (MMN), “a component of the auditory event-related potential (ERP) that is elicited during passive listening by an infrequent change in a repetitive series of sounds” (Shröder et al., 2006), for sad and happy utterances in PD patients in stages 1 and 2. The decreased MMN suggests that PD patients have difficulties in processing emotions in prosody (Shröder et al., 2006). However, Garrido-Vásquez et al. (2013) compared PD patients with asymmetrical motor involvement to average elderly using sentences and pseudo-sentences, and discovered that PD patient with left-side motor involvement had alterations in ERP when listening to sentences with prosody conveying disgust, anger and happy. Both studies concluded that the differences in performance between groups present in specific emotion valence were not resulted from hearing or auditory impairment, otherwise all emotions would be affected more or less equally.

In order to clarify if there is an effect of different sensory modalities, researchers contrasted the results of emotion perception in facial expression and speech prosody, yet, the findings were even more random. Buxton, MacDonald, & Tippett (2013) found PD patients of stages 2 and 3 were significantly less accurate than healthy elderly in identifying emotions – disgust, sadness and joy, through facial expression and speech. Conversely, Ariatti, Benuzzi, & Nichelli (2008) found that judging emotion from prosody is more difficult than from facial expression for PD patients in stages 1 to 3, especially in identifying disgust. Finally, Kan, Kawamura, Hasegawa, Mochizuki, & Nakamura (2002) reported the exact opposite results, since PD patients of stages 2 and 3 in their research performed better in prosody tasks than in facial expression tasks, with the accuracy in prosody perception similar to typical healthy older adults.

1.4.4 Production of emotional prosody in PD

Further acoustic investigations revealed that the prosodic expression ability of PD patients is associated with the task type. For instance, Möbes et al. (2008) had mild to moderate PD patients (mean stage = 1.4, in a scale of stages 1 to 5 according to Hoehn and Yahr's (1967) severity measure) perform three tasks – produce sustained vowel [aː], say ‘Anna’ with happy, neutral and sad emotions, and imitate a professional actor’s speech recording for these emotions. They found the pitch and loudness ranges of PD patients’ speech were significantly smaller than healthy controls when producing non-imitative emotional prosody, and the reduced pitch range in PD patients was especially evident for the happy emotion (Möbes et al., 2008). Cheang and Pell (2007) confirmed this tendency of task-related performance, as they found significant difference between PD speakers in stage 2 and healthy older adults in emotional prosody and contrastive stress tasks, where the experimental group presented lower pitch and loudness. In summary, motor symptoms seem to be insufficient to explain the selectivity of speech deficit found in expressive emotional prosody in PD.

One reason for these different results in perception and production of emotional prosody in PD may be associated with the complexity of experimental tasks, the illness severity and the received medical therapy (Assogna, Pontieri, Caltagirone, & Spalletta, 2008). The research methodology in the studies may be the other possible explanation. A closer examination concerning research method revealed that most of studies investigating prosodic skills among PD patients were set to collect evidence to formulate theories. These studies were not intended to test any hypothesis, as all of them did not follow any specific emotion theory as a research framework. And this atheoretical research approach often compromises the interpretation of outcomes (Scherer, 2003).

1.5 Bio-informational dimensions and body size projection

Emotional expressions are the fundamentals of social interaction through which we can interpret and infer others' intentions (Manstead, 2005; Parkinson, 1996).Yet, this ability had implications in evolution, as those who failed to interpret or express emotion might have their survival jeopardised, e.g., misinterpreting a hostile expression may cost individuals their lives (Darwin, 1872).

Morton (1977) proposed, based on observation of a large number of calls by mammals and birds and their associated behaviours, that many animal calls are made to manipulate the acoustic cues to project a large or small body size in order to influence the hearer – when being aggressive, they use low-pitched and rough voice to sound large so as to dominate the opponents; when being sociable, they use high-pitched and pure-tone-like voice to sound small so as to attract the hearer. Ohala, (1984) extended this theory to humans, and suggested that the vocal tract length (VTL) can also be manipulated to achieve additional 'body-size projection' effect: the longer the VTL, the larger the person sounds, and vice versa.

The relevance of the body-size projection theory to human emotional perception has been tested in a series of empirical studies with different languages by using vowels, digits, and sentences with synthetically manipulated pitch, vocal tract length (VTL) and voice quality as stimuli (Chuenwattanapranithi, Xu, Thipakorn, & Maneewongvatana, 2008; Noble & Xu, 2011; Xu, Kelly, & Smillie, 2013; Xu, Lee, Wu, Liu, & Birkholz, 2013). Results showed that stimuli synthesised with longer VTL (smaller formant dispersion) and lower pitch were perceived as spoken by a larger person and sounding angry, while those with shorter VLT (wider formant dispersion) and higher pitch were perceived as from a smaller person and sounding happy (Chuenwattanapranithi et al., 2008; Xu, Kelly, & Smillie, 2013).

Xu, Kelly, & Smillie (2013) then proposed that these manipulations are along a set of bio-informational dimensions (BID), which include body-size projection, dynamicity, audibility and association. Body-size projection, the scheme for using acoustic cues to convey body size and emotion, is associated with voice quality, VTL and pitch; dynamicity indicates how vigorous the vocalisation is; audibility controls how far the vocalisation can be heard; and association is the use of sounds accompanying certain biological functions (such as vomiting) as emotional signals (Xu, Kelly, & Smillie, 2013).

Given their demonstrated effectiveness, the BIDs, especially the body-size projection dimension, could be used for assessing a person’s sensitivity to emotional prosody with greater precision than previous research methods. And this would be especially useful for assessing emotion perception of atypical populations such as individuals with ASD and PD patients.

Chapter 2 Research questions

The present study is designed to investigate the perception of emotional prosody in TM-speaking high-functioning adolescents with ASD by comparing them with typically developing adolescents and average young adults, and in TM-speaking mild to moderate PD patients by comparing them with healthy elderly and young adults. Synthetically manipulated utterances were used as auditory stimuli in perception tasks. Acoustic parameters extracted from production tasks were further analysed. Acoustic parameters manipulated in synthetic utterances and those extracted from respondents' speech recordings are associated to the BID theory. This innovative research method allows us to control emotionally relevant acoustic cues independently, and to directly assess the contribution of each of them to emotional perception and production.

In order to investigate perception and production of emotional and attitudinal prosody by adolescent with ASD and PD patients, two main research questions have been proposed:

1. Do speakers with ASD or PD perceive the acoustic parameters in the same way as typical speakers when interpreting size code, emotional prosody and speech attitude?
2. Do they manipulate the acoustic parameters in the same way as typical speakers when conveying emotional prosody?

Based on past findings, the hypothesis for these research questions is that the performance of adolescents with ASD and PD patients in perception and production of emotional and attitudinal prosody is not at the same level of aptitude as their respective control groups.

Chapter 3 Research method

The information concerning the participants, auditory stimuli, acoustic parameters manipulated in the synthetic speech and extracted from speech samples, and experimental procedures were described in this section. The research paradigm of the present study is similar to those used in Chuenwattanapranithi et al. (2008), Xu, Kelly, et al. (2013), Xu & Kelly (2010), and Xu, Lee, et al. (2013), but with two main innovations: the inclusion of the speech production experiment, and the use of Mandarin as the target language. Ethical approvals for the current study were granted by University College London (UCL) Research Ethics Committee (3842/001) in the UK and National Taiwan University (NTU) Hospital Research Ethics Committee (201205111RIC) in Taiwan.

3.1 Participants

All the participants in the current study were native speakers of TM, currently living and studying or working in Taipei or New Taipei City. All the participants and a parent of those who were under the age of 20 years signed the informed consent forms. All the participants under the age of 20 years were accompanied by a school teacher or a parent throughout the experimental procedure, to prevent them from experiencing any anxiety with an unknown experimenter.

3.1.1 Participants for the experiment involving ASD

Ten adolescents with high-functioning autism or Asperger syndrome were recruited as the experimental group, and ten typically developing adolescents and ten average young adults served as two different control groups.

The ten adolescents with high-functioning autism or Asperger syndrome (AA group) (8 males and 2 females), recruited from Autism Society Taiwan, and Autism Parent’s Association in Taiwan, had been formerly diagnosed according to DSM-IV-TR criteria by clinicians from teaching hospitals in Taipei, Taiwan, and at the testing time they held the Physically and Mentally Disabled Manual (a primary document issued by Taiwanese Ministry of Interior to those who are eligible to receive social welfares, special medical care, and/or special education) which were registered as having ‘mild form of autism’. As confirmed by their parents, none of them had other learning difficulties or medical conditions that were not related to ASD, and all of them had normal hearing and normal or corrected-to-normal vision.

The ten typically developing adolescents (TA group) (2 females and 8 males), recruited from high schools and junior high schools, and the ten young adults (YA group) (5 females and 5 males) who had college degrees, self-reported (and reported by their parents for those who were under the age of 20 years) as never having experienced or currently experiencing learning difficulties or medical conditions that may compromise their communication skills, and had normal hearing and normal or corrected-to-normal vision.

The Peabody Picture Verbal Test – Revised – Chinese version (PPVT-R) (Dunn & Dunn, 1981) and the Test of Nonverbal Intelligence – Third Edition – Chinese version (TONI-3) (Brown, Sherbenou, & Johnsen, 1997) were administered to both AA and TA groups to assess their receptive language skills and their nonverbal IQ, respectively. The Autism Quotient (AQ) score – Chinese version (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Simon Baron-Cohen, Hoekstra, Knickmeyer, & Wheelwright, 2006) was administered to the three groups to assess for ASD.

Note that the PPVT-R and the TONI 3 have been normalised to the ages of 16 years and 18 years respectively, while some of participants from the AA and TA groups and the whole YA group were beyond these ages. The main reason for using these tests is that there are virtually no language or IQ tests in Traditional Chinese that covers all the participants' age range in the present study. Moreover, neither IQ nor language tests were administered to the YA group, as all the participants of this group graduated from college or graduate school, were employed in fulltime jobs, and reported having no difficulties in learning, therefore it was assumed that they all had average or above-average IQ (i.e. IQ≥70) and normal language skills.

Results from the assessments and basic information from the subjects were analysed with independent t-tests by using SPSS, v. 22.0 (IBM Corp., 2013). Effect sizes were also calculated with Cohen's d using G*Power3 (Faul, Erdfelder, Lang, & Buchner, 2007) to determine the degree of association between variables. Effect size values between 0.2 and 0.5 are considered a small effect, between 0.5 and 0.8 are considered a medium effect, and above 0.8 are considered a large effect.

There were no significant differences between the AA and TA groups in age ($t(18) = 0.58$, $p = 0.570$, effect size, $d = 0.26$), and performance in PPVT-R ($t(18) = -2.03$, $p = 0.058$, effect size, $d = 0.91$) and in TONI-3 ($t(18) = -0.97$, $p = 0.346$, effect size, $d = 0.43$). The AQ score was significantly different between the AA and TA groups, with the AA group scoring higher than the TA group ($t(18) = 6.12$; $p < 0.001$, effect size, $d = 2.74$), and between the AA and YA group, with the AA outscored the YA group ($t(18) = 7.32$; $p < 0.001$, effect size, $d = 3.27$). The participants' profile is summarised in the Table 3.

Table 3. Group means, standard deviations, and ranges of age, PPVT-R, TONI-3, and AQ scores by AA, TA and YA groups

	AA *(n = 10)*		TA *(n = 10)*		YA *(n = 10)*	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age (years)	15.97 (2.30)	13.35-18.91	15.41 (2.04)	13.65-18.50	26.26 (1.52)	23.68-28.01
PPVT-R	114.10 (7.23)	99-125	119.10 (2.92)	115-124	N/A	N/A
TONI-3	45.60 (9.91)	25-55	49.10 (5.72)	38-56	N/A	N/A
AQ	31.30 (7.54)	21-45	15.70 (2.83)	12-20	12.00 (3.56)	7-18

3.1.2 Participants for the experiments involving PD

Eleven elderly PD patients were recruited as the experimental group, and eleven healthy elderly people and eleven average young adults served as the two different control groups.

The eleven elderly patients with idiopathic PD (5 female and 6 males), with severities between stages 2 and 3 according to Hoehn and Yarh's severity measure (staging from 0 to 5, with stage 5 being the severest) (Hoehn & Yahr, 1967), were recruited for the PD group. The participants of this group were diagnosed and recruited by Dr Chung Hwei Tai, a neurologist specialised in movement disorders from the NTU Hospital. All of them were receiving dopaminergic medication at the moment of recruitment, and were tested with medication on. These patients never experienced (prior to onset of PD) nor currently experiencing other neurological pathologies, cognitive deficits or depression, and have normal hearing and normal or corrected-to-normal vision.

The eleven healthy elderly (5 female and 6 males) were recruited for the HE group, and eleven young adults (6 females and 5 males) who had college degrees or masters degree were recruited for the YA group. The participants of both groups self-reported as never having experienced nor currently experiencing any neurological pathologies, cognitive deficits or depression, and have normal hearing and normal or corrected-to-normal vision.

The Montreal Cognitive Assessment (MoCA) Test Full – Chinese (Taiwan) version (downloadable at: http://www.mocatest.org/paper-tests/moca-test-full/) (Kandiah et al., 2014; Nasreddine et al., 2005) was administered to the PD and HE groups to assess their cognitive skills. They were also tested with a simplified version of a pure-tone audiometry test to estimate their hearing threshold at 500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz. This audiometry test was a Matlab script written by Professor Stuart Rosen from UCL. All the pure-tone stimuli were delivered by the TDH39 headphones, attached to a laptop computer (HP Pavillion dm3) where the test and stimuli were controlled. These headphones were calibrated according to the recommendations in BS EN ISO 389-1:2000 (identical to ISO 389-1:1998) (British Society of Audiology, 2011), using Ono Sokki CF-350Z, which is a portable dual channel FFT (fast Fourier Transform) analyzer. The details of the calibration are described in the Table 23 in the Appendix A. Three measurements were extracted from the pure-tone audiometry test results – 500-Hz hearing level (HL) which is the hearing level of the participant’s better ear at 500 Hz; 3-frequency HL which is the hearing level of the better ear across 500 Hz, 1,000 Hz and 2,000 Hz; and 4,000-Hz HL which is the hearing level of the better ear at 4,000 Hz.

Note the YA group did not take the MoCA test and the pure-tone audiometry, since all the participants of this group self-reported not having cognitive or hearing problems, in addition, they were aged between 23 years to 29 years, and this age rage is much below the age of 60 years, the threshold for suffering from age-related problems, such as presbycusis (age-related hearing loss) and dementia, therefore it was assumed that they all had average IQ and normal language skills.

Results from the assessments and basic information from the respondents were analysed with independent t-tests using SPSS, v. 22.0 (IBM Corp., 2013). Effect sizes were also calculated with Cohen's d using G*Power3 (Faul et al., 2007) to determine the degree of association between variables. Effect size values between 0.2 and 0.5 are considered a small effect, between 0.5 and 0.8 are considered a medium effect, and above 0.8 are considered a large effect.

There were no significant differences between the PD and HE groups in age ($t(20) = -1.64$, $p = 0.119$, effect size, $d = 0.70$). The performance in MoCA was significantly different ($t(20) = -3.20$, $p = 0.007$, effect size, $d = 1.36$) between the two groups, the HE group achieved higher mean score (mean = 70.51) than the PD group (mean = 66.95). No significant differences were found between these groups in pure-tone audiometry test results – 500 Hz HL ($t(20) = 0.26$, $p = 0.800$, effect size, $d = 0.11$), 3 Frequency HL ($t(20) = -1.00$, $p = 0.921$, effect size, $d = 0.04$) and 4,000 Hz HL ($t(20) = -0.89$, $p = 0.388$, effect size, $d = 0.38$). The participants' profile is summarised in the Table 4.

Table 4. Group means, standard deviations, and ranges for age, years of diagnosis, years of medication, MoCA, and pure-tone audiometry by PD, HE and YA groups

	YA *(n = 11)*		PD *(n = 11)*		HE *(n = 11)*	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Age (years)	26.57 (1.77)	23.68-29.67	66.95 (4.21)	61.48-75.89	70.51 (5.83)	61.10-80.71
Years of diagnosis	N/A	N/A	6.77 (4.34)	2-13	N/A	N/A
Years of medication	N/A	N/A	5.59 (3.38)	2-10	N/A	N/A
MoCA	N/A	N/A	27.73 (1.56)	26-30	29.36 (0.67)	28-30
500-Hz HL (dB)	N/A	N/A	24 (4.37)	20-60	24 (3.93)	20-65
3-frequency HL (dB)	N/A	N/A	23 (3.69)	20-60	23 (3.42)	20-65
4,000-Hz HL (dB)	N/A	N/A	29 (15.18)	20-70	34 (10.98)	20-70

3.2 Material

3.2.1 Perception tasks

The auditory stimuli were generated from the TM utterance “uo yu a yi you yue” (“I have (an) appointment with aunty”) (Table 5), spoken by a male native speaker of TM, aged 37 years, recorded in an anechoic chamber at UCL, UK, in an emotionally ‘neutral’ voice. A single stimulus utterance was used in the experiment, because individuals with PD usually have attention deficit, and too many sentences may distract them from the changes in the acoustic parameters, which are the main target of the experiment. The content of this utterance has no emotional or attitudinal connotations, and so potential semantic interference with the emotional prosody is minimal.

Table 5. Stimulus utterance for the perception tasks

Sentence in TM	我 與 阿 姨 有 約
Pinyin transcription	ǔo yǔ ā yí yǒu yūe
IPA transcription	uo˨˩˦ y˨˩˦ a˥ i˧˥ iou˨˩˦ ye˥
Meaning	“I have (an) appointment with aunty”

This spoken utterance was then used as a model to create three entirely artificial versions of the same sentence synthesized by VocalTractLab, an articulatory speech synthesiser (Birkholz, Kröger, & Neuschaefer-Rube, 2011). The three synthetic utterances differed only in voice quality, ranging from pressed to modal (normal) and to breathy voice. The specifications of phonetic parameters of VocalTRactLab to generate the synthetic utterances can be found in the Table 24 in the Appendix B, and spectrograms of synthetic utterances with different acoustic manipulations can be seen in the Figures 90 to 101 in the Appendix C.

These synthetic utterances were then further modified in terms of formant shift ratio, pitch shift and pitch range using a script for Praat, a software program for speech analysis and processing (Boersma & Weenink, 2013).

Through this stimulus preparation process, four BID-based acoustic parameters were manipulated: voice quality, formant shift ratio, pitch shift and pitch range. Voice quality was controlled by the amount of glottal closure. When the glottis is tightly closed, the voice quality is pressed; when it is left slightly open during the closing phase of each vocal cycle, the voice quality is breathy. Pressed voice projects a large body size and so should sound angry, whereas breathy voice projects a small body size and so should sound happy. Formant shift ratio is an acoustic parameter that controls simulated vocal tract length (VTL). Small formant shift ratios simulate long VTL. A small formant shift ratio leads listener to hear a large and angry person, and a large formant shift ratio leads to an opposite perception. Pitch shift controls pitch height. A positive shift raises pitch, and negative shift lowers pitch. A low pitch is linked to a large body size and angry voice, and a high pitch is linked to a small body size and happy voice (Chuenwattanapranithi et al., 2008). Pitch range is the difference between the maximum and minimum pitch. This is the only parameter that does not directly signal body size. Instead, it signals how energetic a person is. The smaller the pitch range, the less energetic the speaker sounds, and the larger the pitch range, the more energetic the speaker sounds (Xu, Kelly, et al., 2013; Xu, Lee, et al., 2013).

These acoustic manipulations of the auditory stimuli were similar to the ones employed in Xu, Lee, et al. (2013). Each acoustic parameter was manipulated in three levels. The total number of stimuli was: 3 voice quality x 3 formant shift ratio x 3 pitch shift x 3 pitch range = 81 stimuli. The manipulations of the acoustic parameters and the predictions based on the body size projection are summarised in Table 6:

Table 6. Manipulation of voice quality, formant shift ratio, pitch shift and pitch range, and projected body size and emotion.

Body size projection	Emotion	Voice quality	Formant shift ratio	Pitch shift	Pitch range
small	happy	breathy	1.1	2 ST⁽*⁾	2
↓	↓	modal	1	0 ST	1
large	angry	pressed	0.9	-2 ST	0.5

⁽*⁾ ST = semitone

3.2.2 Production tasks

Two TM sentences (Table 7), “mao mi mai mao mao” (“(The) cat buys (a) woollen hat”) and “ni na na niou nai” (“Nina carries/fetches milk”) were displayed on the laptop monitor with instructions for participants to read out the sentences with the emotion happy, neutral or angry, or attitude friendly, neutral or serious. The stimuli and instructions were shown randomly to the participants using a JavaScript. Each of the tasks was divided into three blocks of six stimuli each (2 sentences X 3 emotions or attitudes), yielding a total of 18 stimuli per task.

Table 7. Stimulus sentences for the production tasks

Sentences in TM	貓 咪 買 毛 帽	倪 娜 拿 牛 奶
Pinyin transcription	māo mī mǎi máo mào	ní nà ná níou nǎi
IPA transcription	mau˥ mi˥ mai˨˩˦ mau˧˥ mau˥˩	ni˧˥ na˥˩ na˧˥ niou˧˥ nai˨˩˦
Meaning	“(The) cat buys (a) woollen hat”	“Nina carries/fetches milk”

Two types of acoustic parameters were extracted from the data to be statistically analysed – the traditional measurements and the BID-associated measurements. All the parameters and predictions based on BID theory are summarised in Table 8. The traditional measurements (though some are also included in BID) used here were mean pitch (Hz), pitch range (Hz), intensity (dB) and mean syllable duration (ms), which have been extensively exploited and reported in past research on expressive prosody of individuals with ASD. Mean pitch and pitch range were also calculated in semitones (ST), as mean pitch (ST) and excursion size respectively, this is to minimise the effects of the difference in pitch between female and male speakers. As demonstrated in past research, between-sex differences in pitch range largely disappear when expressed in semitones (Traunmüller & Eriksson, 1995).

The measurements uniquely associated to the BID theory were those related to VTL and voice quality. Formant dispersion (Fitch, 1997) was used to describe VTL, as the less dispersed the formants, the longer the VLT (Ohala, 1984). In the present study, formant dispersion 3 and formant dispersion 5, both expressed in Hz, are calculated as the average distance between adjacent formants up to F3 or F5, respectively. Other things being equal, the smaller the formant dispersion, the longer the VTL, and the larger the projected body size (Ohala, 1984).

The voice quality (VQ)-related acoustic parameters were H1-H2, H1*-H2*, H1-A1, H1-A3, centre of gravity, energy below 500 Hz, and energy below 1,000 Hz. All of them measure spectral tilt in one way or another. Spectral tilt refers to the flatness of the overall spectrum. Other things being equal, the more tilted (downward) the spectrum toward higher frequency, the breathier the voice. The less tilted the spectrum, the flatter it is, and the more pressed is the voice. A breathy voice, because it is more pure tone like, projects a small body size (Morton, 1977), and hence sounding happy (Xu et al., 2013). A pressed voice, in contrast, projects a large body size and hence sounding angry (Xu et al., 2013).

H1-H2 measures spectral tilt at the left (low frequency) edge of the spectrum. It is the amplitude difference between the first (H1) and second harmonics (H2). H1-A1 is the amplitude difference between H1 and the first formant (A1), and so it measures spectral tilt at a slightly higher frequency position. H1-A3, which measures spectral tilt at an even higher frequency, is the amplitude difference between H1 and the third formant (A3). H1*-H2* is the corrected H1-H2. Iseli & Alwan (2004) proposed a formula for H1*-H2* based on magnitude correction, which is “achieved by moving the influence of vocal tract resonances” (Iseli & Alwan, 2004). Past research found that H1-H2 is correlated to the open quotient, as female speakers have higher open quotient than male speakers (Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995). Thus, H1*-H2* allows better comparison between different vocal tract configurations, e.g. speakers of different gender and age (Iseli, Shue, & Alwan, 2007). Centre of gravity (CG) measures the tilt of the entire spectrum (Surendran & Levow, 2008), and in this study it has been reported in Hz and ST. Energy below 500 Hz and energy below 1,000 Hz divide the spectrum into two bands and measures the energy of the lower band up to 500 Hz or 1,000 Hz relative to the energy above the dividing frequency. When converted into semitones, the cut-off points of 500 Hz and 1,000 Hz correspond to 107.59 ST and 119.59 ST, respectively. The formula used to convert Hz into ST is:

$ST = 12 \log_2 (\text{mean } f0)$, where mean f0 is expressed in Hz.

All the acoustic parameters mentioned above were summarised in the Table 8.

Table 8. Acoustic parameters and predictions based on the body size projection

Body size projection		small size	large size
Emotion		happy	angry
Traditional acoustic parameters	Mean pitch (Hz and ST)	large value	small value
	Pitch range (Hz)	large value	small value
	Excursion size (ST)	large value	small value
Vocal tract length		short length	long length
VTL-related acoustic parameters	Formant dispersion 3 (Hz)	large value	small value
	Formant dispersion 5 (Hz)	large value	small value
Voice quality		breathy	pressed
VQ-related acoustic parameters	H1-H2 (dB)	negative intensity	positive intensity
	H1*-H2* (dB)	negative intensity	positive intensity
	H1-A1 (dB)	negative intensity	positive intensity
	H1-A3 (dB)	negative intensity	positive intensity
	Centre of gravity (Hz and ST)	small value	large value

Table 8. Acoustic parameters and predictions based on the body size projection

Body size projection		small size	large size
Emotion		happy	angry
	Energy below 500 Hz (dB)	high value	low value
	Energy below 1,000 Hz (dB)	high value	low value

3.3 Procedure

All participants performed the tasks individually in a single session in a soundproof room (AA group), in a quiet classroom in the school (TA group), in a quiet room in the hospital (PD group), or in a quiet room in the participant’s home (YA and HE group). Production tasks were performed before perception tasks to minimise possible learning effects. No time limits were imposed during the experiment, yet all the participants completed perception and production tasks within an hour.

3.3.1 Perception tasks

The tasks were run by the ExperimentMFC module of Praat on a laptop computer (HP Pavillion dm3). The auditory stimuli were delivered via Sennheiser HD 265 “linear” headphones. The participants could adjust the playback volume to a comfortable level. They sat in front of the laptop wearing the headphones. All the participants of the AA, TA and YA groups were instructed to click on the choices shown on the laptop monitor using a mouse. Yet, due to their poorer dexterity with the mouse, the elderly participants were instructed to say the choices aloud, and the experimenter clicked on the alternatives accordingly.

The stimuli were presented to the participants in random order in three independent perception tasks, involving judgment of body size, emotion and attitude of the speaker. In total, each of them made 243 judgments (81 stimuli X 3 tasks) which took 20 to 30 minutes to complete. They were allowed to take a break once every 27 stimuli, but all of them managed to finish the tasks without interruption. Each stimulus was presented only once in order to reduce testing time, since elderly people might have attention issues with lengthy tests. After listening to each stimulus, participants performed a three-choice task. The choices were small, medium and large in body size projection task. In the emotion task, the choices were happy, neutral and angry. And in the attitude task, the choices were friendly, neutral and serious. They were instructed to make judgments instinctively without thinking too hard.

3.3.2 Production tasks

Recordings were made with a laptop computer (HP Pavilion dm3 Notebook PC) using Praat speech analysis software (Boersma & Weenink, 2013). A condenser microphone (Fostex MC32) was connected to the laptop through a mixer (Zenyx 320USB). The microphone was placed at an angle of 45-degrees and at a distance of 10 to 15cm away from the participant's mouth. The recordings were digitized into wave-format at 44.1 kHz, and stored on the laptop hardware.

The participants were asked to produce two TM sentences (Table 4). They were instructed to read out the sentence with the emotion happy, neutral or angry, or attitude friendly, neutral or serious, as shown on the laptop monitor. The stimuli and instructions were shown randomly to the participants using a JavaScript. Emotion and attitude productions were run in two separate tasks. Each of the tasks was divided into three blocks of six stimuli each (2 sentences X 3 emotions or attitudes), yielding a total of 18 stimuli per task. Participants were allowed to repeat each trial until a satisfactory production was elicited. The best productions according to the participants were kept and used for the acoustic analysis.

The recordings were annotated and segmented into five syllables using Praat, and were further analysed with ProsodyPro_BID (Xu, 2014), a custom-written Praat script developed for extracting a large set of acoustic parameters from the recorded speech. The script also generated a set of acoustic parameters associated with emotions and attitudes based on findings of previous studies (Xu, Kelly, et al., 2013; Xu, Lee, et al., 2013). In the next section, each of these parameters will be explained in detail.

Chapter 4 Results

4.1 Perception of emotion and attitude by adolescents with ASD, typically developing adolescents and young adults

The responses of participants in the AA, TA and YA groups were automatically collected by Praat (Boersma & Weenink, 2013) as tables and saved by the experimenter. The responses obtained in the three experiments were coded as follows:

- Small as 0% large, medium as 50% large, and large as 100% large.
- Happy as 0% angry, neutral as 50%, and angry as 100% angry.
- Friendly as 0% serious, neutral as 50% serious, and serious as 100% serious.

The coded responses of each experiment were first analysed with four 3x4 mixed design ANOVAs, one for each acoustic parameters – voice quality, formant shift ratio, pitch shift and pitch range, using SPSS version 22.0 (IBM Corp., 2013), with the three groups (AA, TA and YA) as between-subjects factor and three manipulation levels as within-subjects factors. In order to find the main effects of each acoustic parameter on each group, the responses of each group in each experiment were also independently analysed with 12 independent one-way ANOVAs. A *p*-value < 0.050 was taken as significant.

4.1.1 Judgement of body size by AA, TA and YA groups

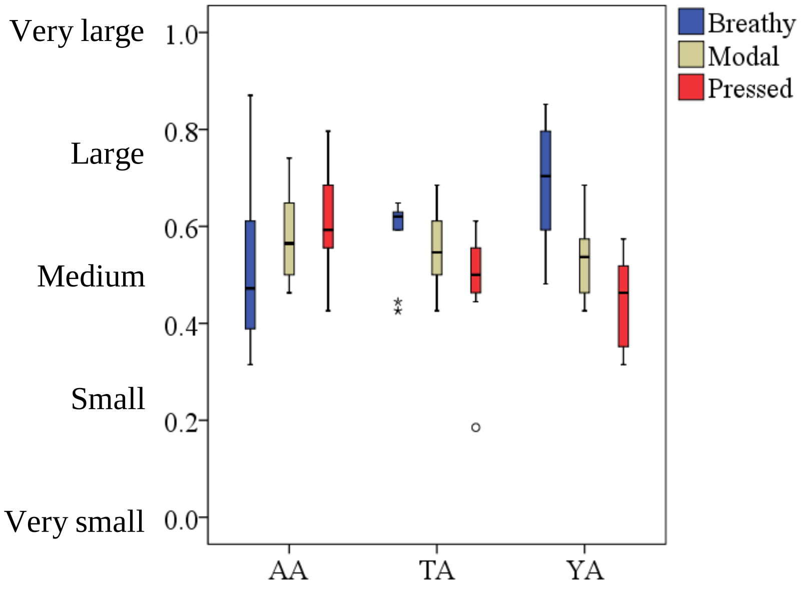

Figure 2: *Effects of voice quality on body size judgment by AA, TA and YA groups*

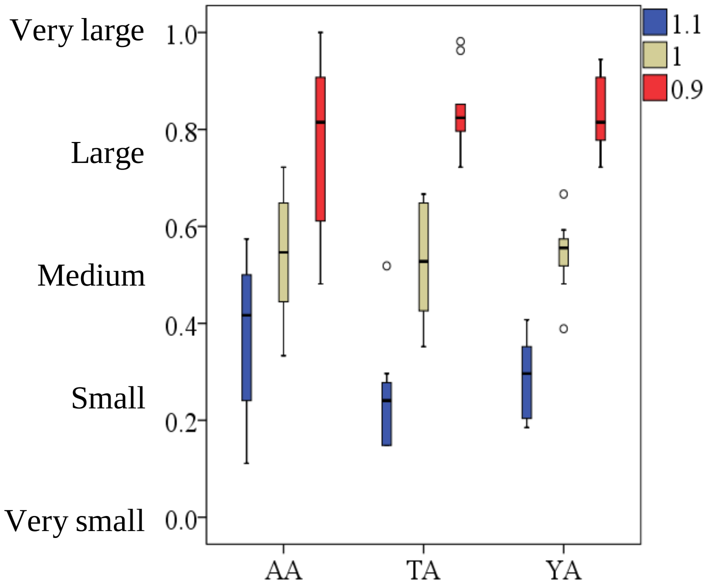

Figure 3: *Effects of formant shift ratio on body size judgment by AA, TA and YA groups*

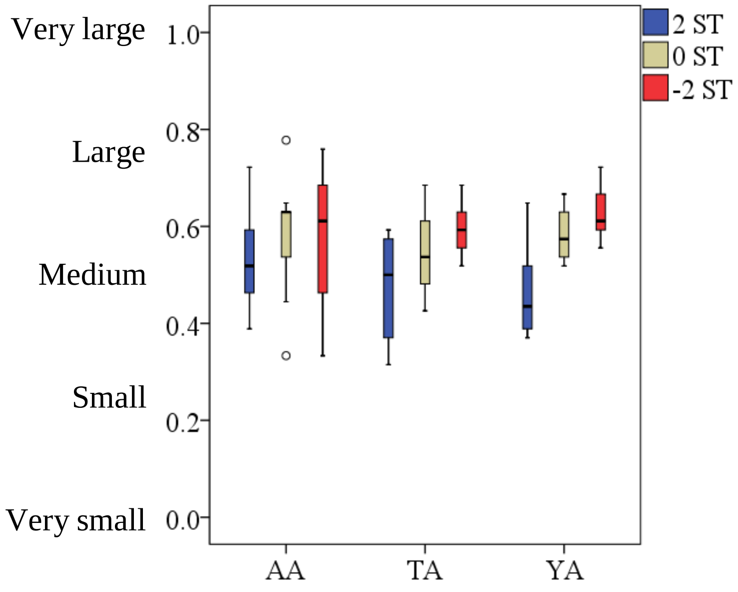

Figure 4: *Effects of pitch shift on body size judgment by AA, TA and YA groups*

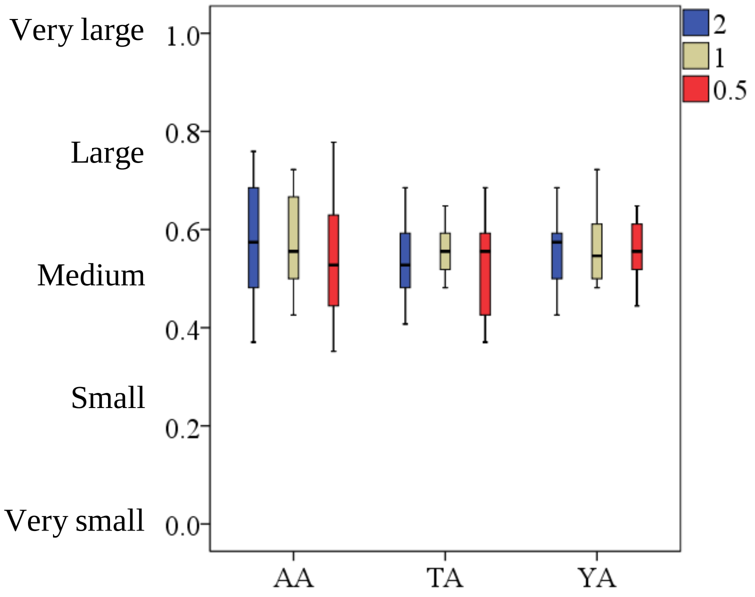

Figure 5: *Effects of pitch range on body size judgment by AA, TA and YA groups*

Figure 2 shows the relationship between voice quality and judgment of body size. This acoustic parameter had a significant effect (F(2, 18) = 14.47, $p < 0.001$) only for the YA group. A closer examination revealed that they perceived breathy voice as from a larger person, and pressed voice as from a smaller one. This result, however, differed from the findings of Xu, Lee, et al. (2013).

Figure 3 displays relationship between formant shift ratio and judgment of body size. All three groups rated utterances with a smaller formant shift ratio as having a larger body size, while a larger formant shift ratio as having a smaller size. However, the AA group showed greater overlap than the other two groups between the three body size judgments. ANOVA results confirmed that formant shift ratio had a significant effect for all three groups: the AA group (F(2, 18) = 16.96, $p < 0.001$), the TA group (F(2, 18) = 79.05, $p < 0.001$), and the YA group (F(2, 18) = 140.09, $p < 0.001$).

The relationship between pitch shift and judgment of body size is shown in Figure 4. Although all the three groups rated downward pitch shift as from a larger person, and upward pitch shift as from a smaller one, this acoustic parameter had significant effect only for the TA (F(2, 18) = 11.10, $p = 0.001$) and YA groups (F(2, 18) = 24.08, $p < 0.001$). Pitch range is the only acoustic parameter with no significant effect among the three groups (Figure 5).

Overall, no main group effect has been found in body size judgment, indicating that the performance of the three groups did not differ much in this task. Among the four acoustic parameters manipulated, the main effect of voice quality (F(2, 54) = 5.33, $p = 0.008$); formant shift ratio (F(2, 54) = 147.49, $p < 0.001$); and pitch shift (F(2, 54) = 19.94, $p < 0.001$) reached statistical significance. Thus these three acoustic parameters were most relevant for the participants of the current study while judging the body size of the speaker – the AA group judged pressed voice quality, smaller formant shift ratio, downward pitch shift and wider pitch range as from large speakers, while the TA and YA group associated breathy voice, smaller formant shift ratio, downward pitch shift and wider pitch range with large body size. Only the interaction of group and voice quality was significant (F(4, 54) = 6.86, $p < 0.001$), which reflects the fact mentioned above that the three groups responded differently to this acoustic parameter. The results of body judgement by the AA, TA and YA groups are summarised in the Table 9.

Table 9. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in body size judgment by AA, TA and YA groups

			AA *(n=10)*		TA *(n=10)*		YA *(n=10)*	
		size	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	small	0.53	0.10	0.52	0.12	0.63	0.11
	modal		0.52	0.07	0.56	0.06	0.51	0.06
	pressed	large	0.54	0.15	0.51	0.09	0.42	0.10
	F-ratio		*n.s.*		*n.s.*		F(2, 18) = 14.47	
	p-value						$p < 0.001$	
Formant shift ratio	1.1	small	0.36	0.16	0.25	0.11	0.29	0.08
	1.0		0.55	0.13	0.52	0.12	0.54	0.07
	0.9	large	0.74	0.18	0.83	0.09	0.84	0.08
	F-ratio		F(2, 18) = 16.96		F(2, 18) = 79.05		F(2, 18) =140.09	
	p-value		$p < 0.001$		$p < 0.001$		$p < 0.001$	

Table 9. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in body size judgment by AA, TA and YA groups

			AA *(n=10)*		TA *(n=10)*		YA *(n=10)*	
		size	Mean	SD	Mean	SD	Mean	SD
Pitch shift	+2 ST	small	0.53	0.09	0.48	0.10	0.46	0.09
	0 ST		0.58	0.12	0.54	0.08	0.59	0.06
	-2 ST	large	0.58	0.15	0.60	0.05	0.63	0.05
	F-ratio p-value		*n.s.*		F(2, 18) = 11.10 $p = 0.001$		F(2, 18) = 24.08 $p < 0.001$	
Pitch range	2.0	small	0.57	0.12	0.54	0.08	0.56	0.08
	1.0		0.59	0.10	0.56	0.06	0.56	0.08
	0.5	large	0.53	0.14	0.52	0.10	0.56	0.06
	F-ratio p-value		*n.s.*		*n.s.*		*n.s.*	

4.1.2 Judgement of emotion by AA, TA and YA groups

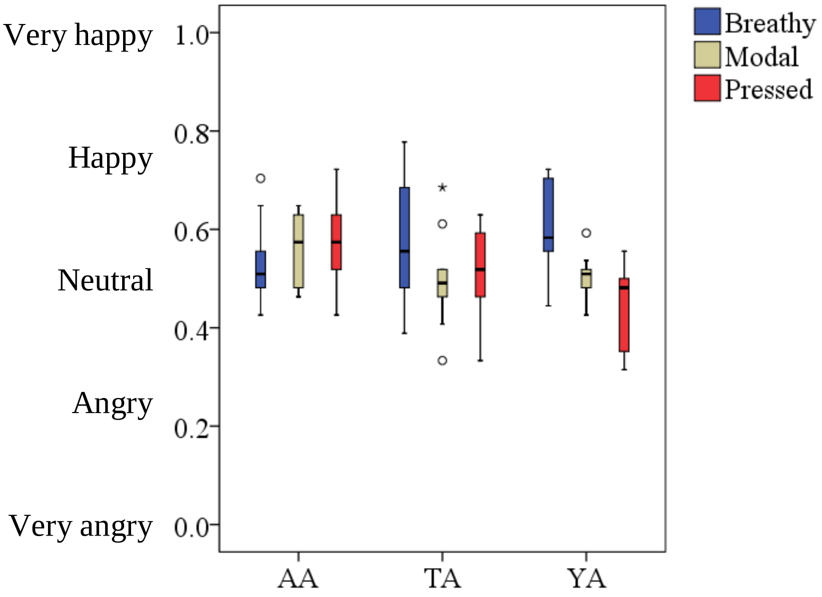

Figure 6: *Effects of voice quality on emotion judgment by AA, TA and YA groups*

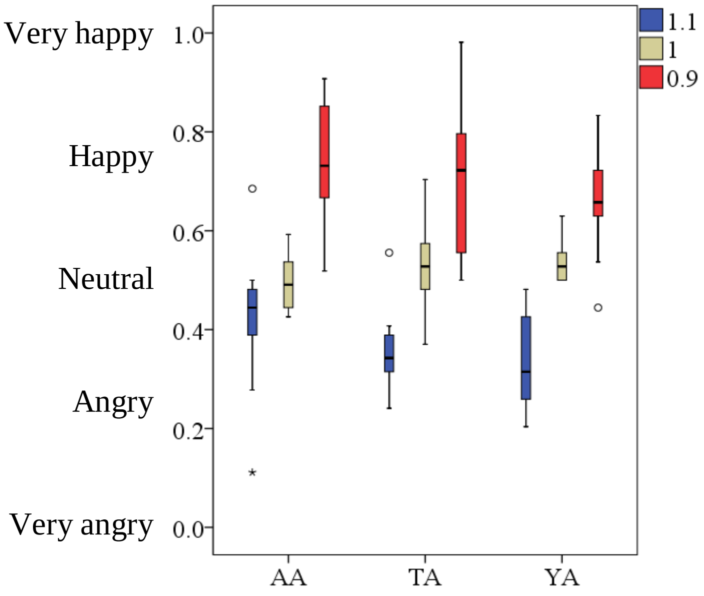

Figure 7: *Effects of formant shift ratio on emotion judgment by AA, TA and YA groups*

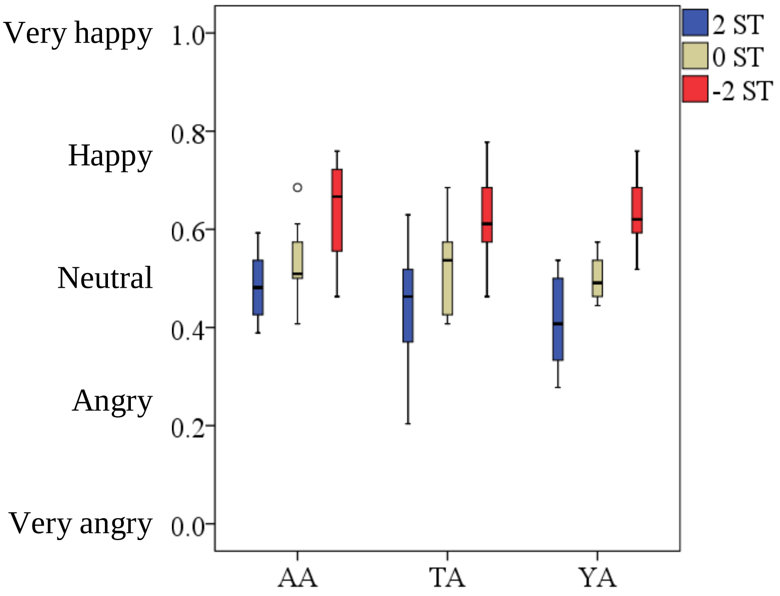

Figure 8: *Effect of pitch shift on emotion judgment by AA, TA and YA groups*

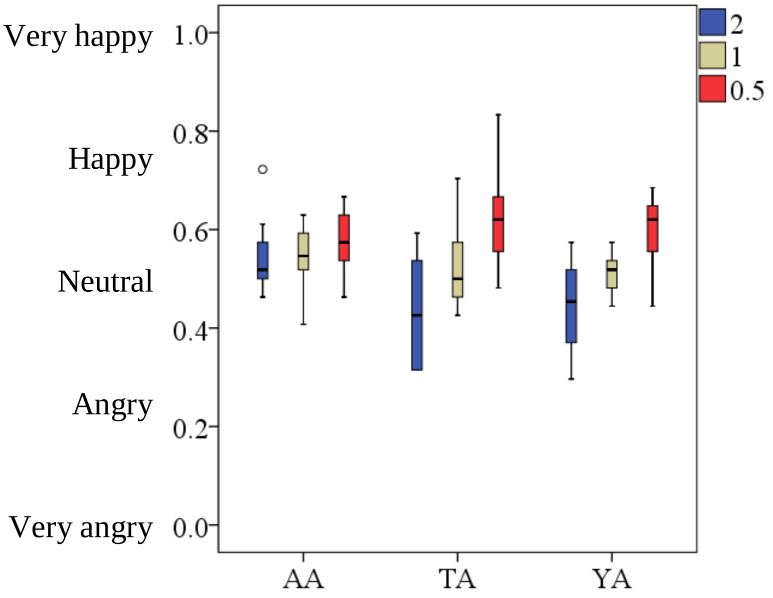

Figure 9: *Effects of pitch range on emotion judgment by AA, TA and YA groups*

Figure 5 shows the relationship between voice quality and judgment of emotion. The effect of voice quality reached statistical significance for both the TA (F(2, 18) = 3.72, *p* = 0.045) and YA groups (F(2, 18) = 8.89, *p* = 0.002). Yet, similar to body size judgment, these groups rated pressed voice as anger and breathy as happiness, which contradicts the findings by Chuenwattanapranithi et al. (2008) and Xu, Lee, et al. (2013).

Figure 7 displays the relationship between formant shift ratio and judgment of emotion. All of them rated smaller formant shift ratio as sounding angry and greater formant shift ratio as happy-sounding. The overlap between the three emotion judgments is small across all groups, and ANOVA results show that formant shift ratio had a significant effect on every group: the AA group (F(2, 18) = 17.11, $p < 0.001$), the TA group (F(2, 18) = 35.79, $p < 0.001$), and YA group (F(2, 18) = 25.93, $p < 0.001$).

Figure 8 shows the the relationship between pitch shift and judgment of emotion. In general, downward pitch shift is perceived as anger, and upward pitch shift as happiness. Once again, ANOVA results show different significance levels across the three groups: the AA group (F(2, 18) = 11.10, $p = 0.001$), the TA group (F(2, 18) = 19.57, $p < 0.001$), and the YA group (F(2, 18) = 24.13, $p < 0.001$).

Figure 9 displays the relationship between pitch range and judgment of emotion. The AA group had greater overlap than the other two groups among the three emotion types. ANOVA results show that pitch range has significant effect for the TA group (F(2, 18) = 19.55, $p < 0.001$) and the YA group (F(2, 18) = 8.60, $p = 0.002$). Both groups judged smaller pitch ranges as anger, and larger pitch ranges as happiness.

Similarly to the body size judgment, there was no main group effect for the emotion judgment, suggesting that the three groups performed similarly overall. The main effect of the four acoustic parameters reached statistical significance: voice quality (F(2, 54) = 5.27, $p = 0.008$); formant shift ratio (F(2, 54) = 70.00, $p < 0.001$); pitch shift (F(2, 54) = 51.61, $p < 0.001$) and pitch range (F(2, 54) = 25.33, $p < 0.001$). These results show that the participants were sensitive to the changes in the acoustic parameters when making emotion judgment – the AA group judged pressed voice quality, smaller formant shift ratio, downward pitch shift and narrower pitch range as sounding angry, while the TA and YA group associated breathy voice, smaller formant shift ratio, downward pitch shift and narrower pitch range with anger. There are also significant interactions between group and voice quality (F(4, 54) = 4.55, *p* = 0.003), and between group and pitch range (F(4, 54) = 3.59, *p* = 0.011), thus the three groups responded differently to these two acoustic parameters. The results of emotion judgement by the AA, TA and YA groups are summarised in the Table 10.

Table 10. Means and standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in emotion judgment by AA, TA and YA groups

			AA (*n* = 10)		TA (*n* = 10)		YA (*n* = 10)	
		emotion	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	happy	0.53	0.09	0.58	0.12	0.60	0.09
	modal	neutral	0.56	0.07	0.50	0.10	0.50	0.05
	pressed	angry	0.57	0.09	0.51	0.09	0.44	0.08
	F-ratio		*n.s.*		F(2, 18) = 3.72		F(2, 18) = 8.89	
	p-value				*p* = 0.045		*p* = 0.002	
Formant shift ratio	1.1	happy	0.42	0.15	0.36	0.08	0.34	0.10
	1	neutral	0.49	0.57	0.52	0.10	0.54	0.05
	0.9	angry	0.74	0.12	0.70	0.15	0.70	0.11
	F-ratio		F(2, 18) = 7.11		F(2, 18) =35.79		F(2, 18) = 5.93	
	p-value		*p* < 0.001		*p* < 0.001		*p* < 0.001	

Table 10. Means and standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in emotion judgment by AA, TA and YA groups

			AA (*n* = 10)		TA (*n* = 10)		YA (*n* = 10)	
		emotion	Mean	SD	Mean	SD	Mean	SD
Pitch shift	2 ST	happy	0.48	0.07	0.44	0.12	0.41	0.09
	0 ST	neutral	0.53	0.08	0.52	0.09	0.50	0.04
	-2 ST	angry	0.65	0.10	0.62	0.09	0.64	0.07
	F-ratio		F(2, 18) =11.10		F(2, 18) = 19.57		F(2, 18) =24.13	
	p-value		*p* = 0.001		*p* < 0.001		*p* < 0.001	
Pitch range	2	happy	0.54	0.08	0.43	0.11	0.44	0.09
	1	neutral	0.54	0.07	0.52	0.09	0.51	0.05
	0.5	angry	0.58	0.07	0.63	0.12	0.59	0.08
	F-ratio		*n.s.*		F(2, 18) = 19.55		F(2, 18) = 8.60	
	p-value				*p* < 0.001		*p* = 0.002	

4.1.3 Judgement of attitude by AA, TA and YA groups

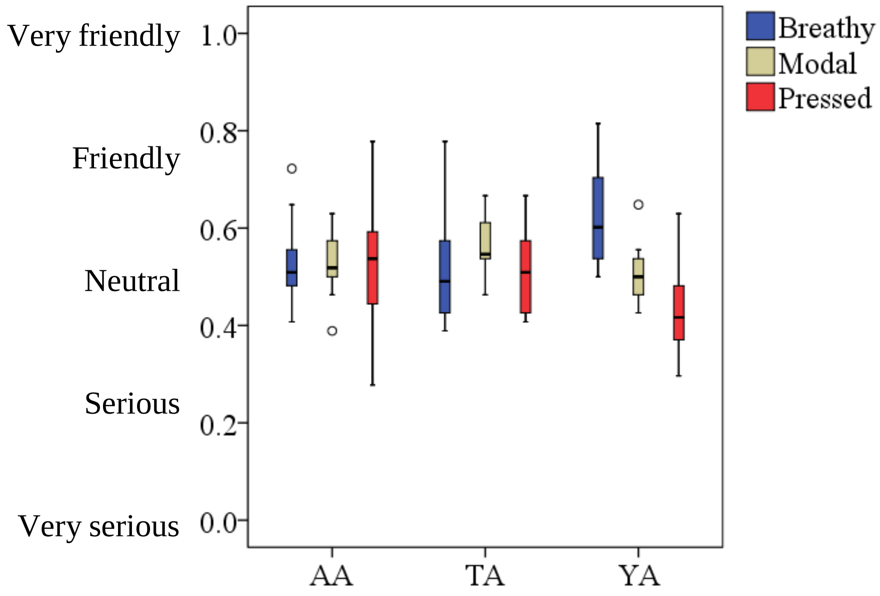

Figure 10: *Effects of voice quality on attitude judgment by AA, TA and YA groups*

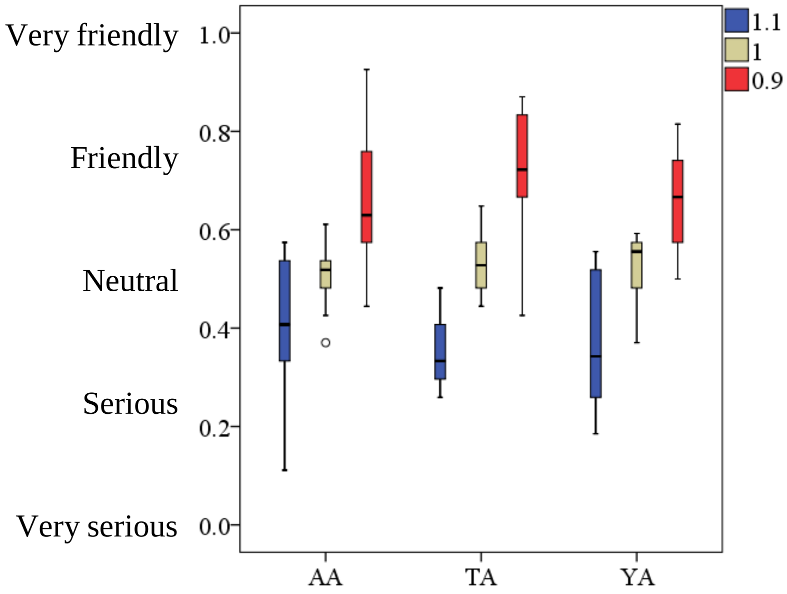

Figure 11: *Effects of formant shift ratio on attitude judgment by AA, TA and YA groups*

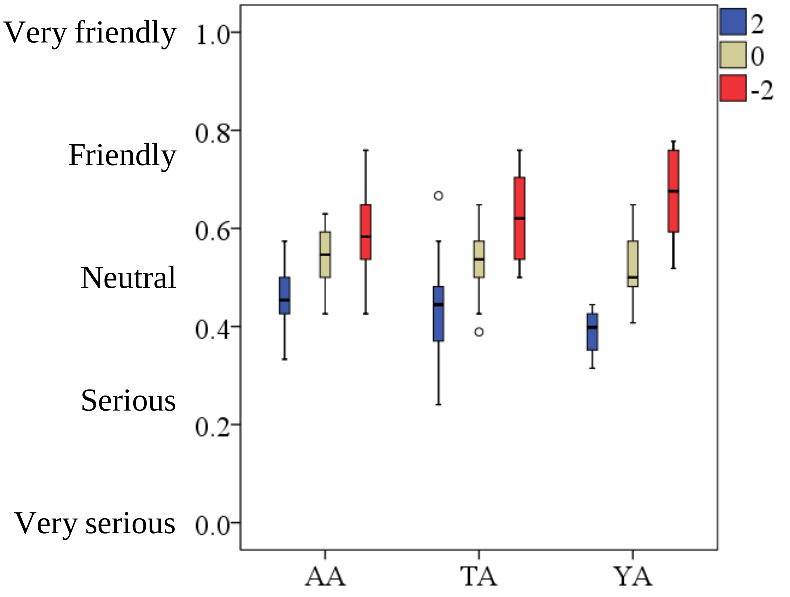

Figure 12: *Effects of pitch shift on attitude judgment by AA, TA and YA groups*

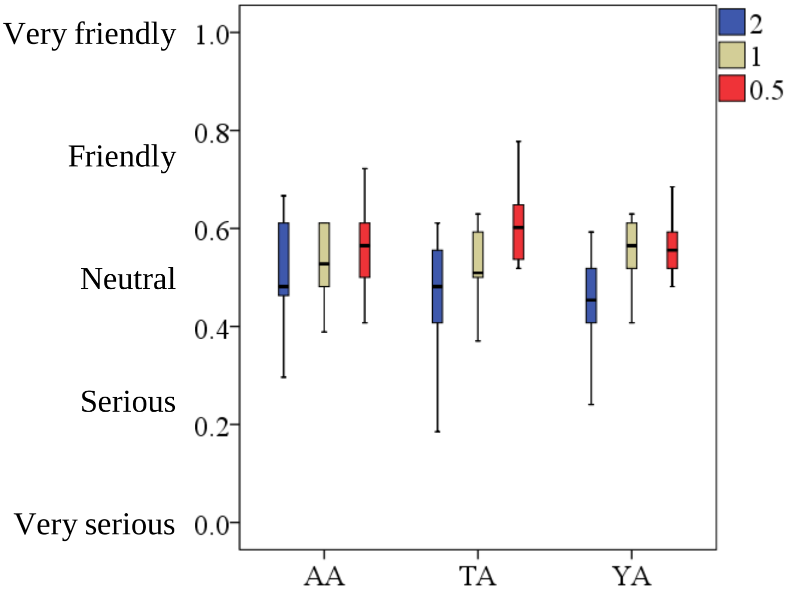

Figure 13: *Effects of pitch range on attitude judgment by AA, TA and YA groups*

Figure 10 describes the relationship between voice quality and judgment of attitude. Once again, only the YA group showed significant effect of voice quality in the attitude judgment (F(2, 18) = 9.24, *p* = 0.002). Yet, they perceived breathy voice as more serious and pressed voice as friendlier.

Figure 11 shows relationship between formant shift ratio and judgment of attitude. All the groups rated smaller formant shift ratio as from a serious person, and larger formant shift ratio as from a friendly person. There is a small overlap between the three attitudes for all three groups. ANOVA results show that formant shift ratio had a significant effect for every group: the AA group ($F(2, 18) = 8.70$, $p = 0.002$), TA group ($F(2, 18) = 22.85$, $p < 0.001$), and the YA group ($F(2, 18) = 15.16$, $p < 0.001$).

The relationship between pitch shift and judgment of attitude is shown in Figure 12. All three groups perceived downward pitch shift as sounding more serious, and upward pitch shift as friendlier. ANOVA results show that the effect of pitch shift is significant for all the three groups: the AA group ($F(2, 18) = 11.33$, $p = 0.001$), the TA group ($F(2, 18) = 8.37$, $p = 0.003$), and the YA group ($F(2, 18) = 31.42$, $p < 0.001$).

Figure 13 shows the relationship between pitch range and judgment of attitude. There is a greater overlap between the three attitude judgments based on pitch range for the AA group. ANOVA results show that the effect of this acoustic parameter is significant only for the TA group ($F(2, 18) = 5.20$, $p = 0.016$) and the YA group ($F(2, 18) = 5.77$, $p = 0.012$), but not for the AA group.

Similar to the body-size and emotion judgment, there was no significant main group effect for the attitude judgment, indicating the three groups did not differ much in their performance. The main effect of three out of four manipulated acoustic parameters reached statistical significance: formant shift ratio ($F(2, 54) = 43.07$, $p < 0.001$); pitch shift ($F(2, 54) = 43.74$, $p < 0.001$); and pitch range ($F(2, 54) = 9.84$, $p < 0.001$). Therefore, the participants of the current study were all sensitive to the changes of these parameters – the AA group judged pressed voice quality, smaller formant shift ratio, downward pitch shift and narrower pitch range as sounding serious, while the TA and YA group associated breathy voice, smaller formant shift ratio, downward pitch shift and narrower pitch range with seriousness. Once again, only the interaction between group and voice quality has reached significance (F(4, 54) = 3.55, *p* = 0.012), indicating that the three groups responded differently to this parameter. The results of attitude judgement by the AA, TA and YA groups are summarised in the Table 11.

Table 11. Means and standard deviations, F ratio and P-value for voice quality, formant shift ratio, pitch shift, and pitch range in attitude judgment by AA, TA and YA groups

			AA (*n* = 10)		TA (*n* = 10)		YA (*n* = 10)	
		Attitude	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	friendly	0.53	0.10	0.52	0.12	0.63	0.11
	modal	neutral	0.52	0.07	0.56	0.06	0.51	0.06
	pressed	serious	0.54	0.15	0.51	0.09	0.51	0.06
	F-ratio		*n.s.*		*n.s.*		F(2, 18) = 9.24	
	p-value						*p* = 0.002	
Formant shift ratio	1.1	friendly	0.40	0.14	0.35	0.08	0.37	0.13
	1	neutral	0.51	0.06	0.53	0.07	0.53	0.07
	0.9	serious	0.67	0.16	0.69	0.16	0.66	0.10
	F-ratio		F(2, 18) = 8.70		F(2, 18) = 22.85		F(2, 18) = 5.16	
	p-value		*p* = 0.002		*p* < 0.001		*p* < 0.001	

Table 11. Means and standard deviations, F ratio and P-value for voice quality, formant shift ratio, pitch shift, and pitch range in attitude judgment by AA, TA and YA groups

			AA (*n* = 10)		TA (*n* = 10)		YA (*n* = 10)	
		Attitude	Mean	SD	Mean	SD	Mean	SD
Pitch shift	2 ST	friendly	0.40	0.14	0.35	0.08	0.37	0.13
	0 ST	neutral	0.51	0.07	0.53	0.07	0.53	0.07
	-2 ST	serious	0.67	0.16	0.69	0.16	0.66	0.10
	F-ratio		F(2, 18) =11.33		F(2, 18) = 8.37		F(2, 18) =31.42	
	p-value		*p* = 0.001		*p* = 0.003		*p* < 0.001	
Pitch range	2	friendly	0.51	0.11	0.47	0.13	0.45	0.10
	1	neutral	0.52	0.08	0.52	0.08	0.55	0.07
	0.5	serious	0.56	0.08	0.61	0.08	0.57	0.07
	F-ratio		*n.s.*		F(2, 18) = 5.20		F(2, 18) = 5.77	
	p-value				*p* = 0.016		*p* = 0.012	

4.2 Production of emotion and attitude by adolescents with ASD, typically developing adolescents and young adults

All the speech samples from the participants were analysed with 16 independent 3x3 mixed design ANOVAs, using SPSS version 22.0 (IBM Corp., 2013), with group (AA, TA and YA) as a between-subject factor and different types of emotions (happy, neutral, angry) or attitudes (friendly, neutral, serious) as a within-subject factor for each acoustic parameter (mean pitch (Hz), mean pitch (ST), pitch range, excursion size, intensity, mean syllable duration, formant dispersion 3, formant dispersion 5, H1-H2, H1*-H2* H1-A1, H1-A3, centre of gravity (Hz), centre of gravity (ST), energy below 500 Hz and energy below 1000 Hz). The speech samples were also analysed with 48 independent one-way ANOVAs for individual group, aiming to find the main effects of individual emotion on each acoustic parameter per group. A *p*-value < 0.050 was taken as significant.

4.2.1 Expression of emotion by AA, TA and AA groups

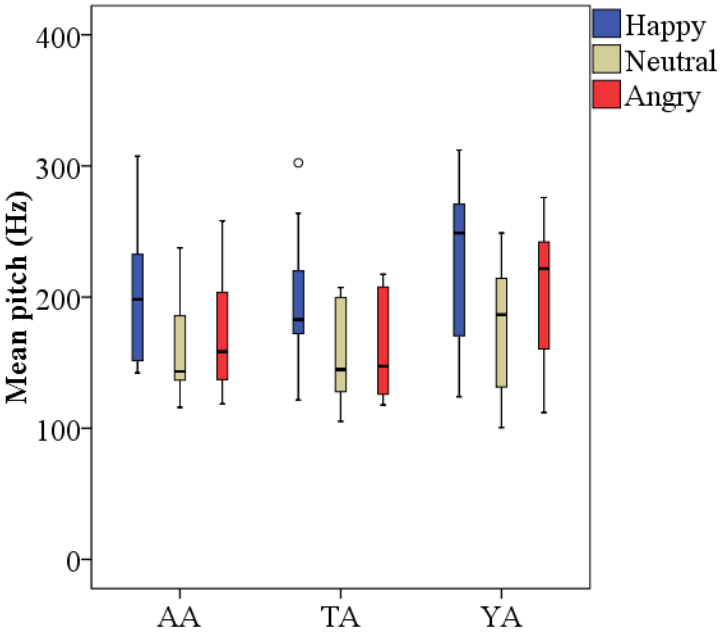

Figure 14: *Effects of mean pitch (Hz) on emotion expression by AA, TA and YA groups*

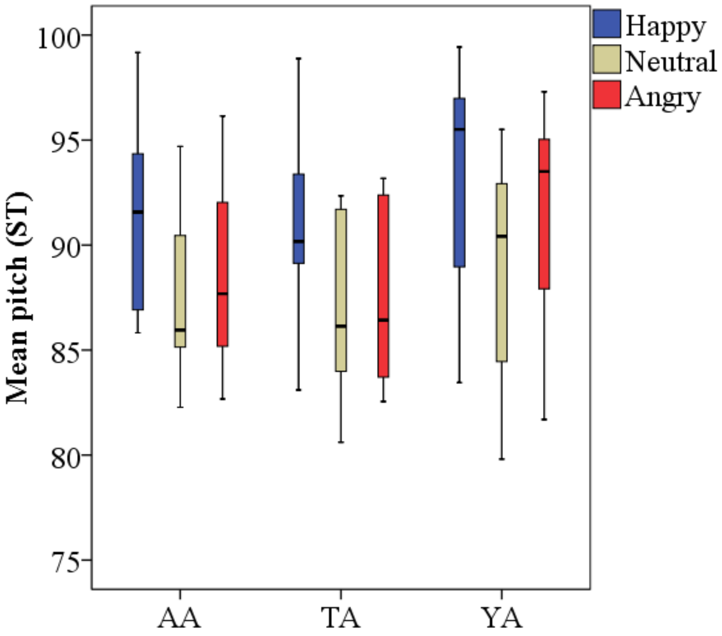

Figure 15: *Effects of mean pitch (ST) on emotion expression by AA, TA and YA groups*

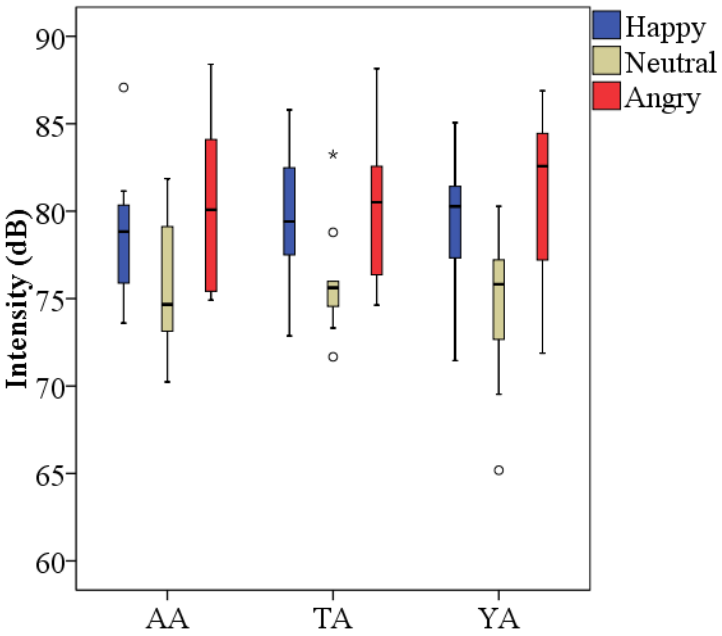

Figure 16: *Effect of intensity on emotion expression by AA, TA and YA groups*

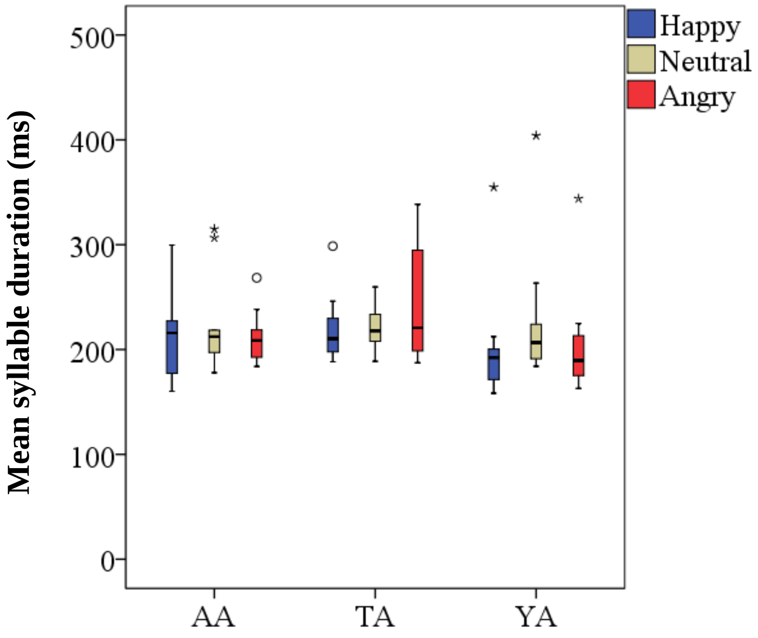

Figure 17: *Effect of mean syllable duration on emotion expression by AA, TA and YA groups*

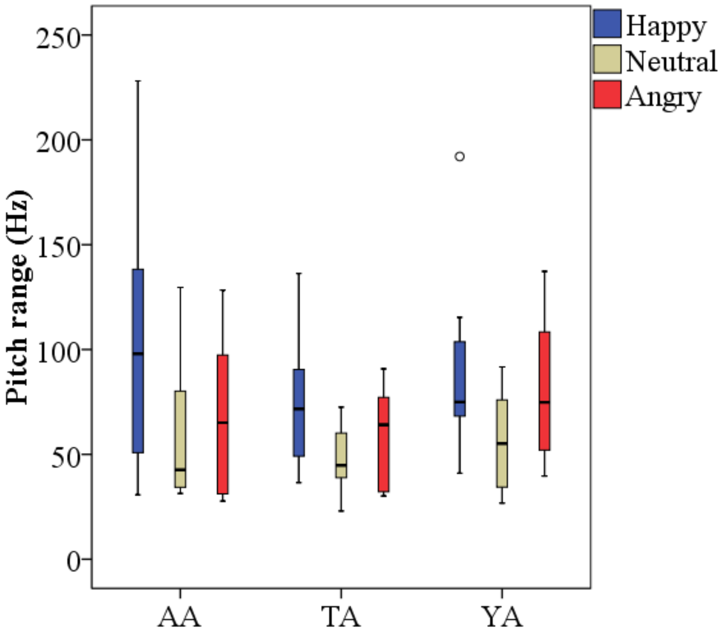

Figure 18: *Effect of pitch range on emotion expression by AA, TA and YA groups*

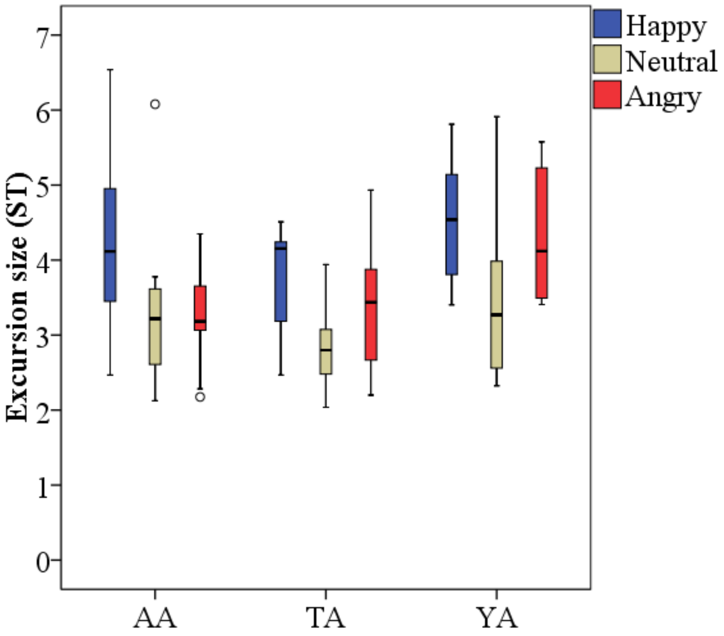

Figure 19: *Effect of excursion size on emotion expression by AA, TA and YA groups*

Figure 14 shows the relationship between pitch (Hz) and expression of emotion. All three groups produced anger with lower mean pitch than happiness. According to the prediction of body-size projection, lower pitch is associated with a larger and angry-sounding person, thus all the participants performed as predicted when conveying anger with lower mean pitch than happiness. Furthermore, ANOVA results demonstrate that all the three groups have a significant effect on mean pitch (Hz): the AA group (F(2, 18) = 6.91, $p = 0.006$), the TA group (F(2, 18) = 12.26, $p < 0.001$), and the YA group (F(2, 18) = 33.34, $p < 0.001$).

Figure 15 demonstrates the relationship between pitch (ST) and expression of emotion. Similar to mean pitch (Hz), all three groups produced anger with lower mean pitch (ST) than happiness. ANOVA results demonstrate that all the three groups have a significant effect on mean pitch (ST): the AA group (F(2, 18) = 6.91, $p = 0.006$), the TA group (F(2, 18) = 12.26, $p < 0.001$), and the YA group (F(2, 18) = 33.34, $p < 0.001$).

Figure 16 depicts the relationship between intensity and expression of emotion. All the three groups expressed anger with higher intensity than happiness, and neutral emotion with the lowest intensity. This indicates that anger is more audible than happiness, and both of them are more audible than neutral emotion. ANOVA results show that all the three groups have a significant effect on intensity: the AA group ($F(2, 18) = 24.89$, $p < 0.001$), the TA group ($F(2, 18) = 13.71$, $p < 0.001$), and the YA group ($F(2, 18) = 48.79$, $p < 0.001$).

Figure 17 shows the relationship between mean syllable duration and expression of emotion. ANOVA results show that only the YA group ($F(2, 18) = 7.95$, $p = 0.003$) have a significant effect on mean syllable duration.

Figure 18 demonstrates the relationship between pitch range and expression of emotion. All the three groups performed anger with narrower pitch range than happiness, and neutral emotion with the narrowest pitch range. Although body-size does not provide particular predictions for pitch range interpretations, from past research findings in emotion, synthetic speech with wider pitch range was often perceived as from happier speakers (Murray & Arnott, 1993; Nwe, Foo, & De Silva, 2003). Thus, all the participants conveyed emotions according to these findings. ANOVA results indicate that all the three groups have a significant effects on pitch range: the AA group ($F(2, 18) = 3.60$, $p = 0.049$), the TA group ($F(2, 18) = 6.58$, $p = 0.007$) and the YA group ($F(2, 18) = 11.33$, $p = 0.001$).

Figure 19 shows the relationship between excursion size and expression of emotion. Similar to pitch range, all the three groups performed anger with narrower pitch range than happiness, and neutral emotion with the narrowest pitch range. ANOVA results indicate that all the three groups have a significant effects on pitch range: the AA group ($F(2, 18) = 3.95$, $p = 0.038$), the TA group ($F(2, 18) = 9.94$, $p = 0.001$) and the YA group ($F(2, 18) = 8.26$, $p = 0.003$).

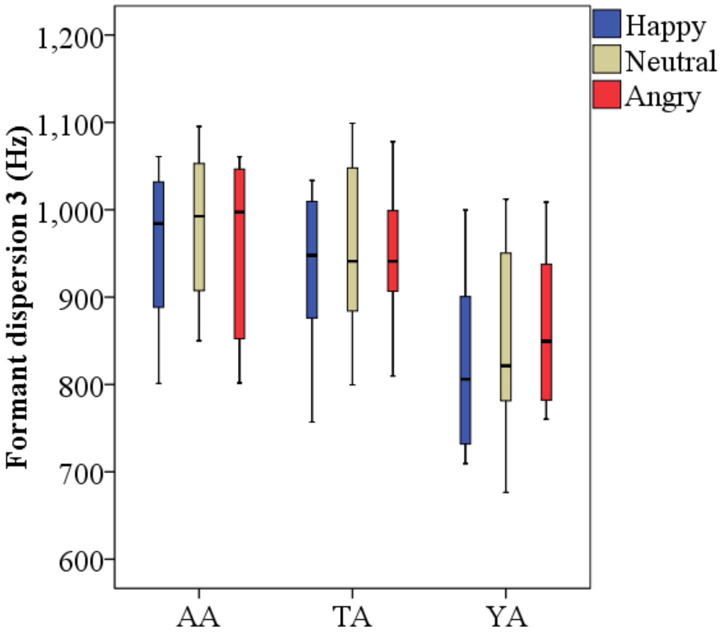

Figure 20: *Effects of formant dispersion 3 on emotion expression by AA, TA and YA groups*

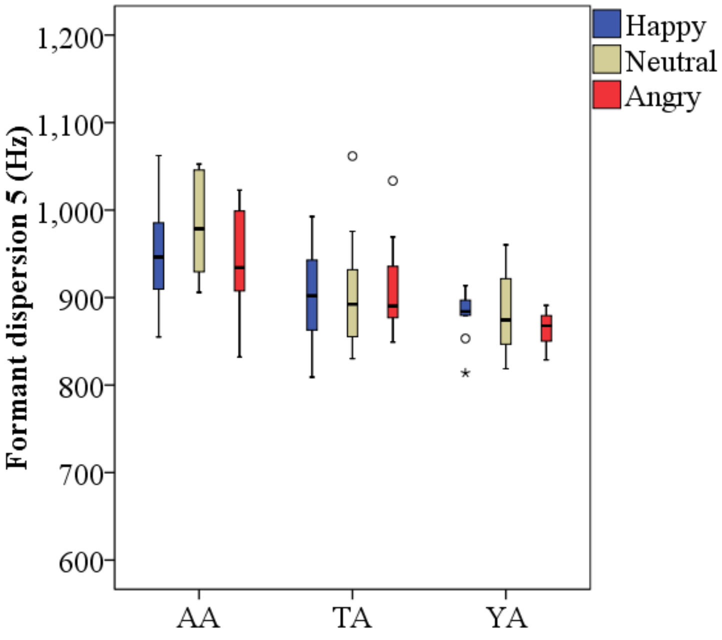

Figure 21: *Effects of formant dispersion 5 on emotion expression by AA, TA and YA groups*

Among the two categories BID-associated parameters, emotion has no significant effect on any VTL-related parameters in any of the three groups. Figures 20 and 21 show the relationship between energy formant dispersion 3, formant dispersion 5 and expression of emotion. ANOVA results confirmed that only formant dispersion 5 is affected by emotion only for the AA group (F(2, 18) = 5.97, p = 0.010). The AA group performed anger with lower formant dispersion 5 than happiness, and neutral emotion with the highest formant dispersion 5. As predicted by body size projection, lower formant dispersion characterises longer vocal tract length, the AA group expressed anger and happiness in line with these predictions.

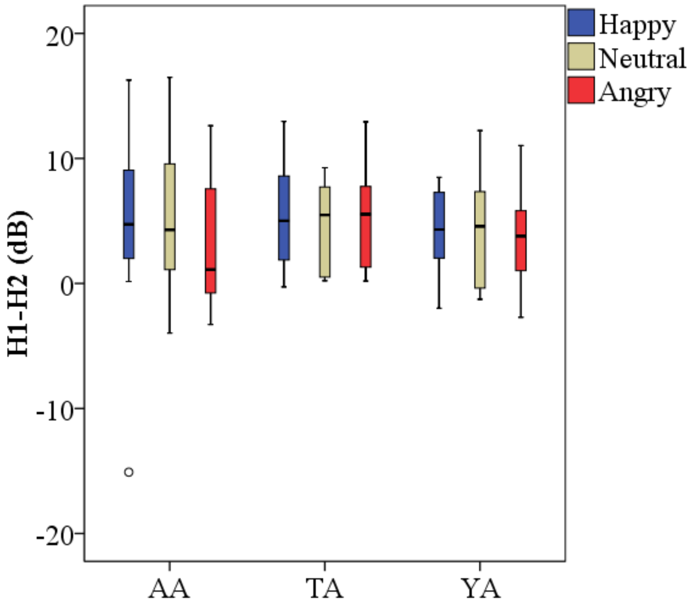

Figure 22: *Effects of H1-H2 on emotion expression by AA, TA and YA groups*

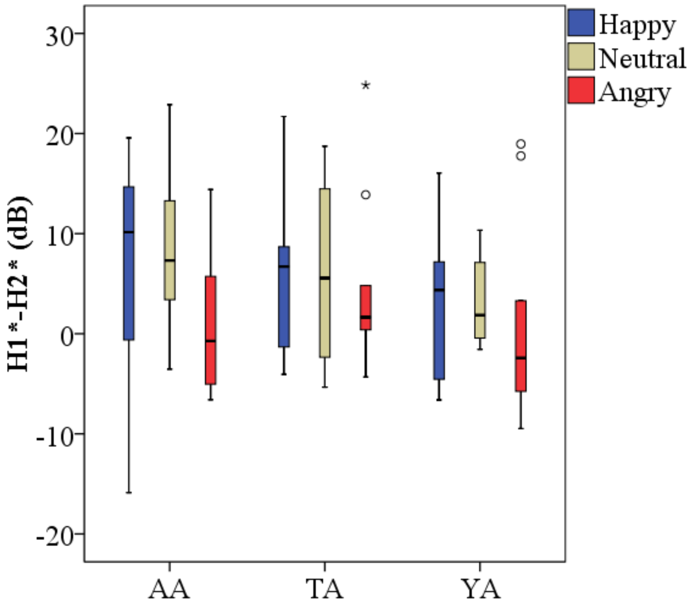

Figure 23: *Effects of H1*-H2 on emotion expression by AA, TA and YA groups*

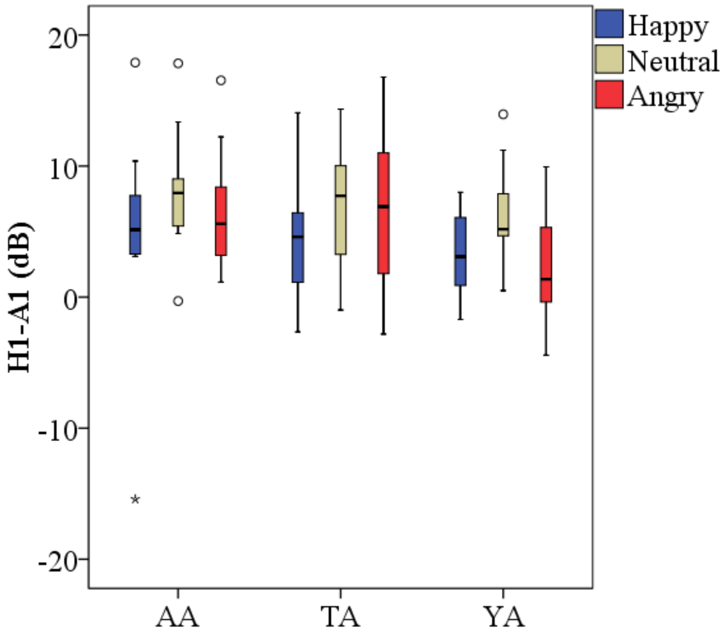

Figure 24: *Effect of H1-A1 on emotion expression by AA, TA and YA groups*

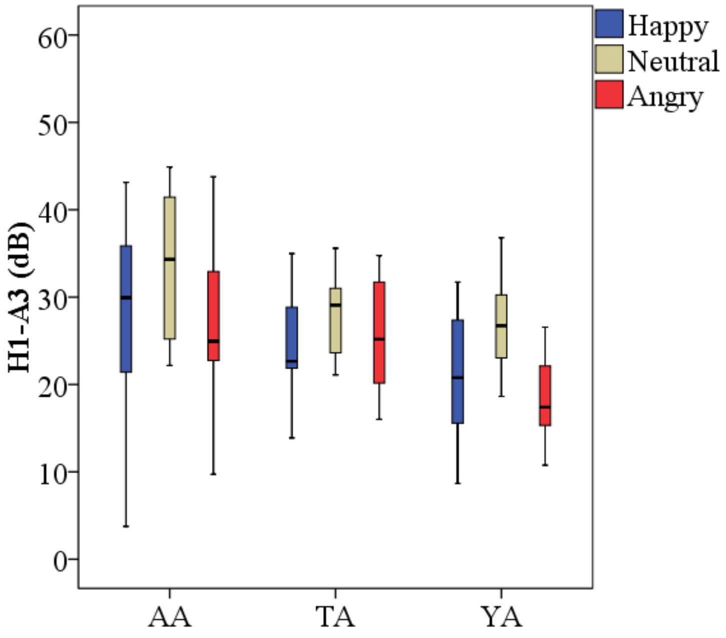

Figure 25: *Effect of H1-A3 on emotion expression by AA, TA and YA groups*

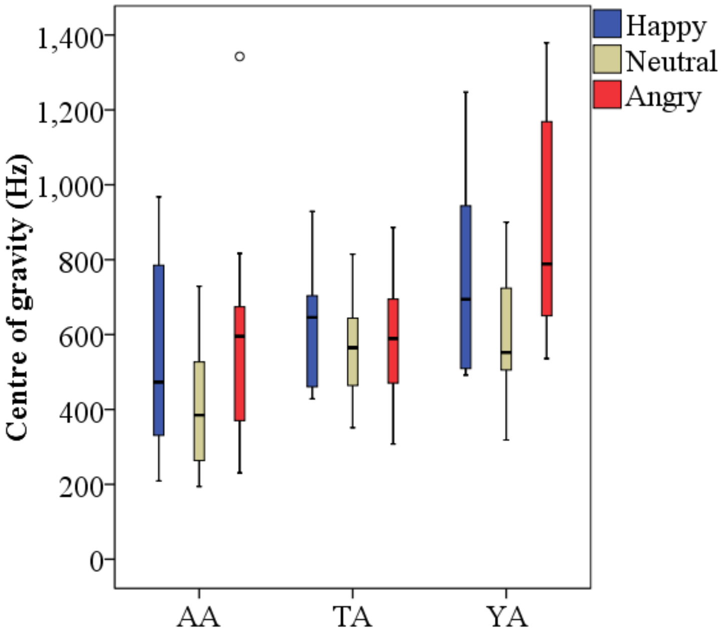

Figure 26: *Effect of centre of gravity (Hz) on emotion expression by AA, TA and YA groups*

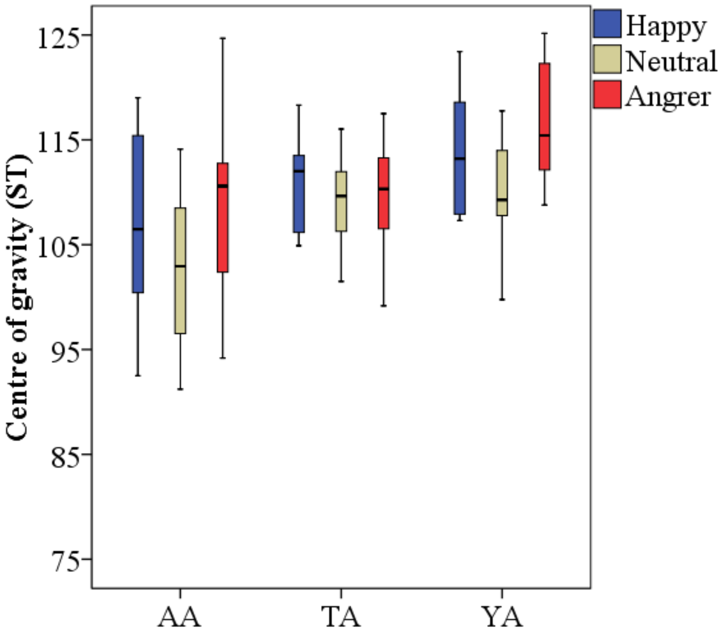

Figure 27: *Effect of centre of gravity (ST) on emotion expression by AA, TA and YA groups*

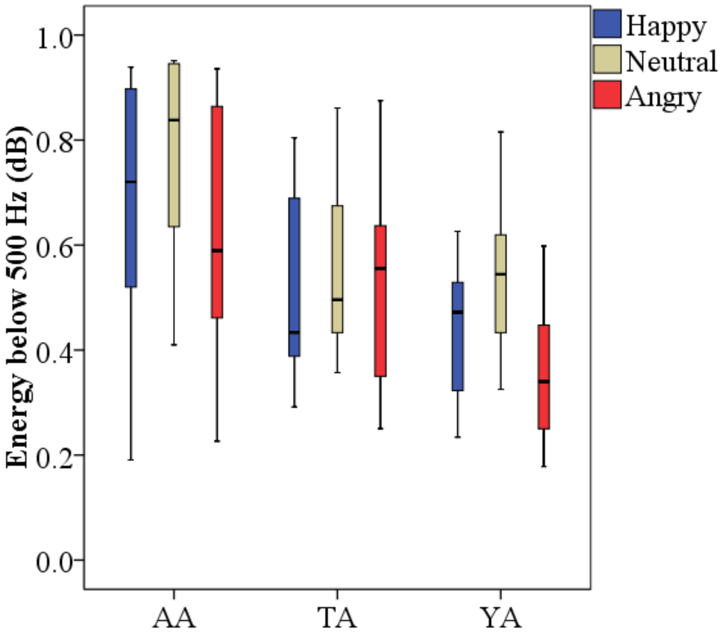

Figure 28: *Effect of energy below 500 Hz on emotion expression by AA, TA and YA groups*

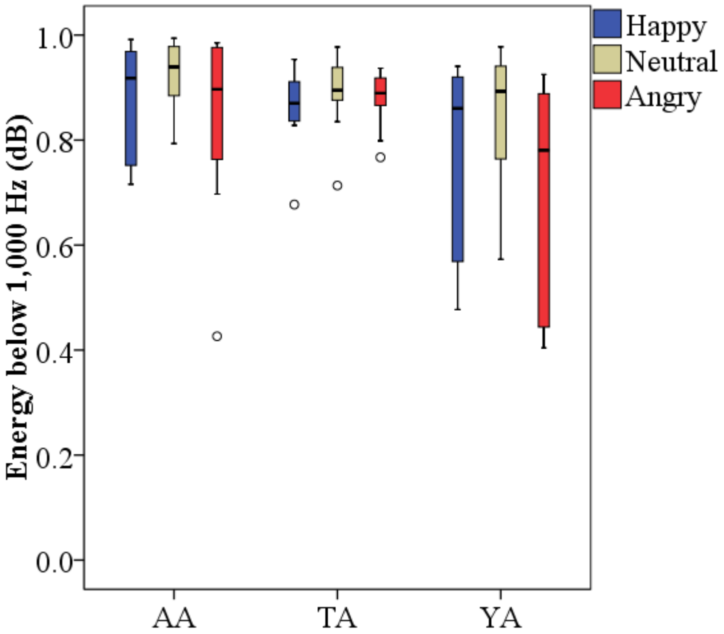

Figure 29: *Effect of energy below 1,000 Hz on emotion expression by AA, TA and YA groups*

Figures 22, 23 and 24 depict the relationship between H1-H2, H1*-H2*, H1-A1 and expression of emotion. ANOVA results confirmed that only the AA group has a significant effect on H1*-H2* (F(2, 18) = 3.79, $p = 0.042$), and the YA group has a significant effect on H1-A1 (F(2, 18) = 5.58, $p = 0.013$). The AA group expressed anger with lower H1*-H2* than happiness. Similarly the YA group also conveyed anger with lower H1-A1 than happiness. Since lower H1-H2 and lower H1-A1 more pressed voice, which characterise a larger or angry speaker, the AA and YA groups performed the emotions in accordance with the predictions of body size projection.

Figure 25 shows the relationship between H1-A3 and expression of emotion. The AA and YA groups produced anger with lower mean H1-H3 values than happiness. Yet, the TA group performed the opposite. As lower H1-A3 characterises more pressed voice quality, which is associated with larger or angry sounding speakers, thus only the AA and the YA groups expressed anger and happiness according to the principles of body size projection. ANOVA results confirmed that all the three groups have a significant effect on H1-A3: the AA group (F(2, 18) = 5.11, *p* = 0.017), the TA group (F(2, 18) = 4.16, *p* = 0.033), and the YA group (F(2, 18) = 11.66, *p* < 0.001).

Figure 26 shows the relationship between CG (Hz) and expression of emotion. The AA and YA groups expressed anger with higher mean CG (Hz) values than happiness, yet the TA group has the opposite performance. Since higher CG (Hz) indicates more pressed voice quality, which is associated with larger or angry-sounding speakers, only the AA and YA groups conveyed anger and happiness according to the predictions of body-size projection. ANOVA results confirmed that only these two groups have a significant effect on CG (Hz): the AA group (F(2, 18) = 4.03, *p* = 0.036) and the YA group (F(2, 18) = 16.98, *p* < 0.001).

Figure 27 shows the relationship between CG (ST) and expression of emotion. Similar to CG (Hz), only the AA and YA groups expressed the emotions according to the predictions of body-size projection. ANOVA results confirmed that only these two groups have a significant effect on CG (Hz): the AA group (F(2, 18) = 6.30, *p* = 0.008) and the YA group (F(2, 18) = 21.67, *p* < 0.001).

Figure 28 describes the relationship between energy below 500 Hz and expression of emotion. The AA and YA groups produced anger with smaller mean energy below 500 Hz value than happiness, while the TA group did the opposite. As lower energy below 500 Hz are associated with more pressed voice quality, which is perceived as from a larger or angrier speaker, thus only the performance of the AA and YA groups complies with the hypothesis of body size projection. ANOVA results confirm that only two groups have a significant effect on energy below 500 Hz: the AA group (F(2, 18) = 8.05, *p* = 0.003) and the YA group (F(2, 18) = 17.34, *p* < 0.001).

Figure 29 shows the relationship between energy below 1,000 Hz and expression of emotion. ANOVA results confirmed that only the YA group has a significant effect on energy below 1,000 Hz (F(2, 18) = 12.89, *p* < 0.001). The YA group conveyed anger with lower energy below 500 Hz value than happiness. Since lower energy below 1,000 Hz is associated with more pressed voice quality which is perceived as from a larger or angrier speaker, the YA group expressed anger and happiness in line with the predictions of body size projection.

Within the acoustic parameters described above, the main group effects are found to be significant only in signal energy below 500 Hz (F(2, 18) = 4.46, *p* = 0.021), suggesting the three groups performed the majority of acoustic parameters similarly. The main effect of emotions are found in mean pitch (Hz) (F(2, 54) = 39.41, *p* < 0.001), mean pitch (ST) (F(2, 54) = 37.98, p < 0.001), intensity (F(2, 54) = 75.86, *p* < 0.001), pitch range (F(2, 18) = 6.58, *p* = 0.007), excursion size (F(2, 54) = 15.16, *p* < 0.001), H1-H3 (F(2, 54) = 17.86, *p* < 0.001), CG (Hz) (F(2, 54) = 15.47, *p* < 0.001), CG (ST) (F(2, 54) = 17.97, *p* < 0.001), and energy below 500 Hz (F(2, 54) = 20.61, *p* < 0.001). These results imply that the participants were producing these acoustic parameters differently for expressing each emotion – they performed anger with lower mean pitch, higher intensity, narrower pitch range, narrower excursion size, decreased H1-A3, larger CG and lower energy below 500 Hz than happiness. Significant interactions between group and emotion are found in CG (Hz) (F(4, 54) = 2.93, *p* = 0.029), CG (ST) (F(4, 54) = 3.18, *p* = 0.020), and energy below 500 Hz (F(4, 54) = 3.30, *p* = 0.017), thus the three groups performed these two acoustic parameters significantly different across the three emotions. The results of emotion expression by the AA, TA and YA groups are summarised in the Table 12.

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
Mean pitch (Hz)	happy	199.26	54.11	196.55	53.89	229.87	61.08
	neutral	158.85	36.79	154.33	37.06	173.98	52.40
	angry	170.97	46.72	164.71	43.62	204.71	57.72
	F ratio	(F(2, 18) = 6.91		F(2, 18) = 12.26		F(2, 18) = 33.34	
	p-value	*p* = 0.006		*p* < 0.001		*p* < 0.001	
Mean pitch (ST)	happy	91.12	4.53	90.86	4.62	93.49	5.21
	neutral	87.36	3.72	86.79	4.11	88.53	5.62
	angry	88.45	4.62	87.83	4.48	91.39	5.46
	F ratio	F(2, 18) = 6.91		F(2, 18) = 12.26		F(2, 18) = 33.34	
	p-value	*p* = 0.006		*p* < 0.001		*p* < 0.001	

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
Intensity	happy	78.69	4.10	79.49	4.12	79.19	4.00
	neutral	75.44	4.01	75.97	3.23	74.59	4.50
	angry	80.68	5.07	80.58	5.04	81.30	4.70
	F ratio	F(2, 18) = 24.89		F(2, 18) = 13.71		F(2, 18) = 48.79	
	p-value	$p < 0.001$		$p < 0.001$		$p < 0.001$	
Mean syllable duration	happy	215.16	44.05	219.46	34.59	202.57	57.05
	neutral	224.65	53.84	221.43	24.76	228.35	66.67
	angry	215.16	44.05	219.46	34.59	202.57	57.05
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 7.95	
	p-value					$p = 0.003$	
Pitch range	happy	99.26	61.76	74.51	31.97	91.10	42.22
	neutral	56.45	34.59	47.91	17.82	56.99	25.65
	angry	70.79	37.77	59.40	27.13	81.29	32.93
	F ratio	F(2, 18) = 3.95		F(2, 18) = 9.94		F(2, 18) = 8.26	
	p-value	$p = 0.038$		$p < 0.001$		$p < 0.001$	

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
Excursion size	happy	4.24	1.24	3.73	0.79	4.52	0.86
	neutral	3.32	1.17	2.82	0.63	3.50	1.17
	angry	3.23	0.78	3.40	1.06	4.29	0.97
	F ratio	F(2, 18) = 3.95		F(2, 18) = 9.94		F(2, 18) = 8.26	
	p-value	p = 0.038		p < 0.001		p < 0.001	
Formant dispersion 3	happy	955.12	97.52	935.82	90.92	826.33	118.66
	neutral	982.35	93.20	956.00	109.46	853.14	127.01
	angry	962.72	101.41	947.29	81.55	866.35	97.92
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
Formant dispersion 5	happy	956.46	69.59	902.16	58.91	880.06	41.37
	neutral	982.50	57.21	904.99	72.86	883.07	52.64
	angry	941.35	68.76	914.64	57.44	863.40	33.06
	F ratio	F(2, 18) = 5.97		*n. s.*		*n. s.*	
	p-value	p = 0.010					

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
H1-H2	happy	4.67	8.85	5.36	4.90	4.20	3.66
	neutral	5.28	6.56	4.80	3.80	4.31	4.58
	angry	3.00	6.05	5.42	4.59	4.02	4.51
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 3.79	
	p-value					*p* = 0.042	
H1*-H2*	happy	7.63	12.54	5.80	9.67	2.94	8.61
	neutral	8.80	10.08	5.81	10.13	2.97	7.34
	angry	1.30	9.75	4.55	11.40	0.87	10.34
	F ratio	F(2, 18) = 3.79		*n. s.*		*n. s.*	
	p-value	*p* = 0.042					
H1-A1	happy	27.81	11.46	24.18	6.73	21.21	7.19
	neutral	33.57	8.63	28.42	5.41	26.76	5.71
	angry	26.59	9.37	25.71	6.84	18.04	4.93
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 5.58	
	p-value					*p* = 0.013	

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
H1-A3	happy	27.81	11.46	24.18	6.73	21.21	7.19
	neutral	33.57	8.63	28.42	5.41	26.76	5.71
	angry	26.59	9.37	25.71	6.84	18.04	4.93
	F ratio	F(2, 18) = 5.11		F(2, 18) = 4.16		F(2, 18) = 11.66	
	p-value	$p = 0.017$		$p = 0.033$		$p < 0.001$	
Centre of gravity (Hz)	happy	530.87	256.12	623.98	161.78	755.75	262.44
	neutral	423.06	188.07	559.12	139.57	608.87	190.90
	angry	592.34	327.28	587.46	178.41	866.50	296.36
	F ratio	F(2, 18) = 4.03		*n. s.*		F(2, 18) = 16.98	
	p-value	$p = 0.036$				$p < 0.001$	
Centre of gravity (ST)	happy	106.69	8.76	110.88	4.57	113.84	5.83
	neutral	103.17	7.69	109.02	4.45	110.19	5.56
	angry	108.24	9.31	109.59	5.70	116.23	5.75
	F ratio	F(2, 18) = 6.30		*n. s.*		F(2, 18) = 21.67	
	p-value	$p = 0.008$				$p < 0.001$	

Table 12. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by AA, TA and YA groups

		AA (*n*=10)		TA (*n*=10)		YA (*n*=10)	
	Emotion	Mean	SD	Mean	SD	Mean	SD
Energy below 500 Hz	happy	0.68	0.26	0.52	0.19	0.44	0.14
	neutral	0.76	0.21	0.56	0.18	0.55	0.15
	angry	0.63	0.25	0.53	0.21	0.36	0.15
	F ratio	F(2, 18) = 8.05		*n. s.*		F(2, 18) = 17.34	
	p-value	$p = 0.003$				$p < 0.001$	
Energy below 1,000 Hz	happy	0.88	0.11	0.86	0.08	0.77	0.19
	neutral	0.92	0.07	0.89	0.08	0.84	0.14
	angry	0.84	0.18	0.88	0.06	0.71	0.21
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 12.89	
	p-value					$p < 0.001$	

4.2.2 Expression of attitude by AA, TA and AA groups

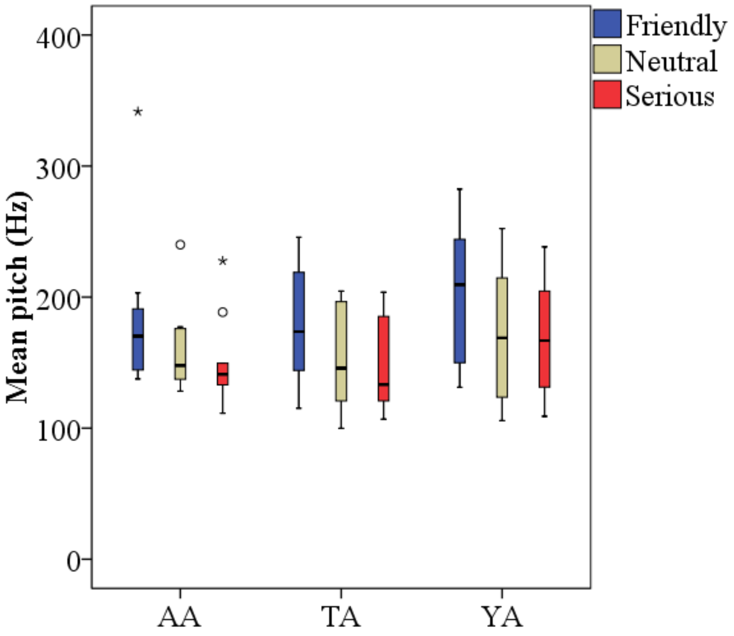

Figure 30: *Effects of mean pitch (Hz) on attitude expression by AA, TA and YA groups*

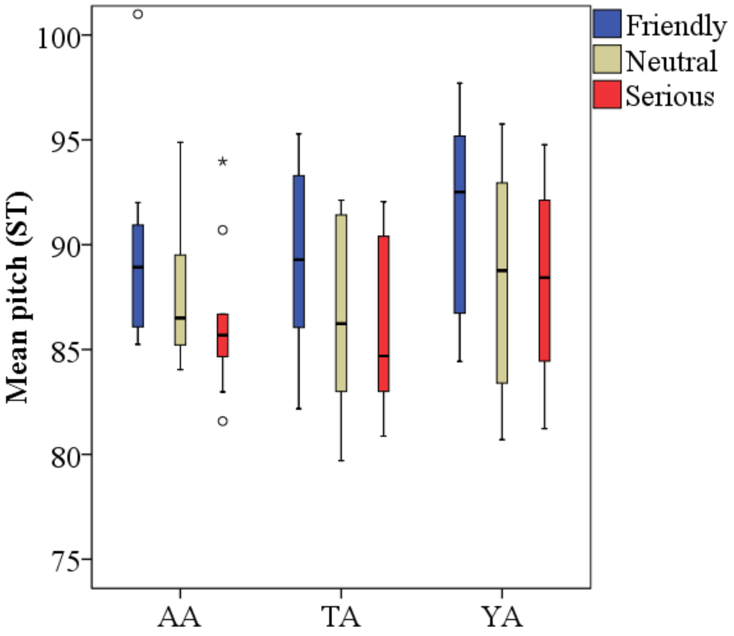

Figure 31: *Effects of mean pitch (ST) on attitude expression by AA, TA and YA groups*

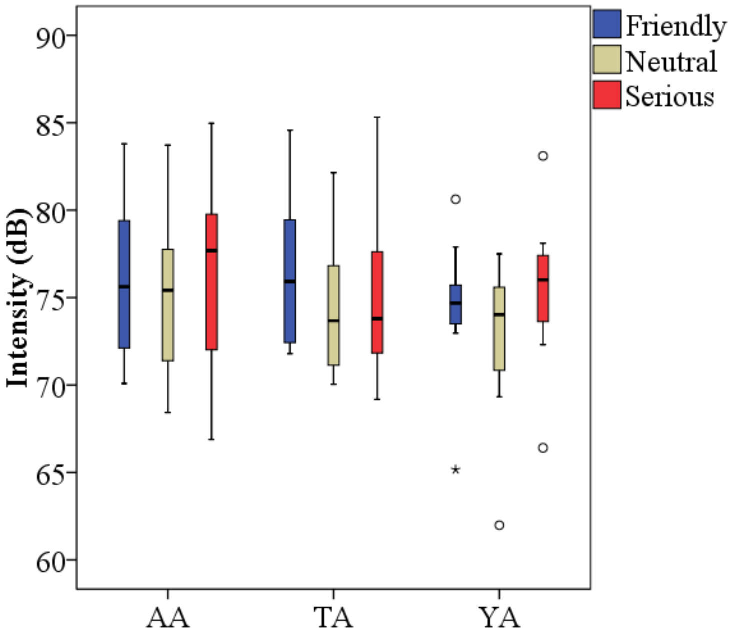

Figure 32: *Effect of intensity on attitude expression by AA, TA and YA groups*

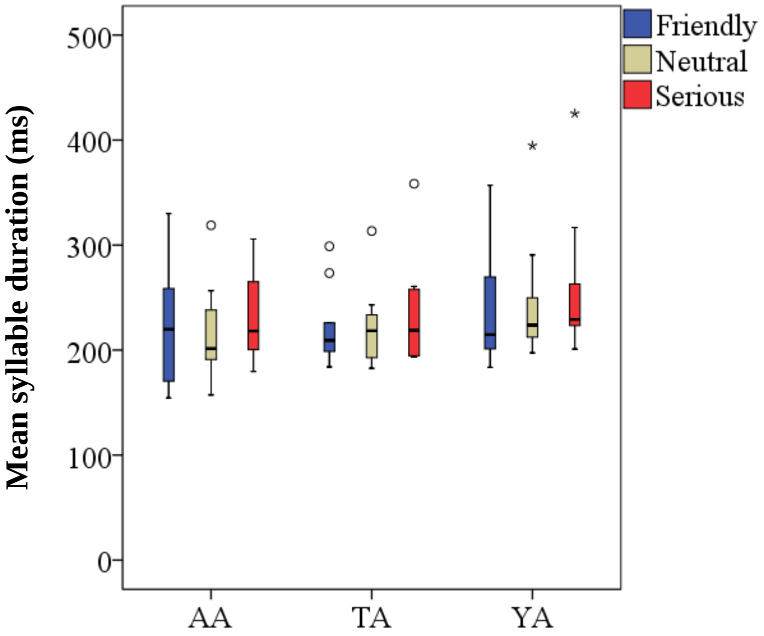

Figure 33: *Effect of mean syllable duration on attitude expression by AA, TA and YA groups*

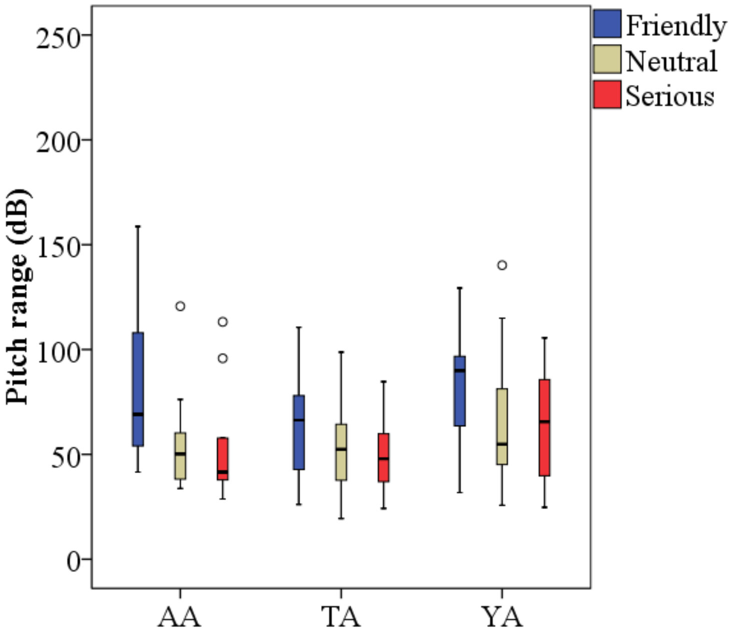

Figure 34: *Effect of pitch range on attitude expression by AA, TA and YA groups*

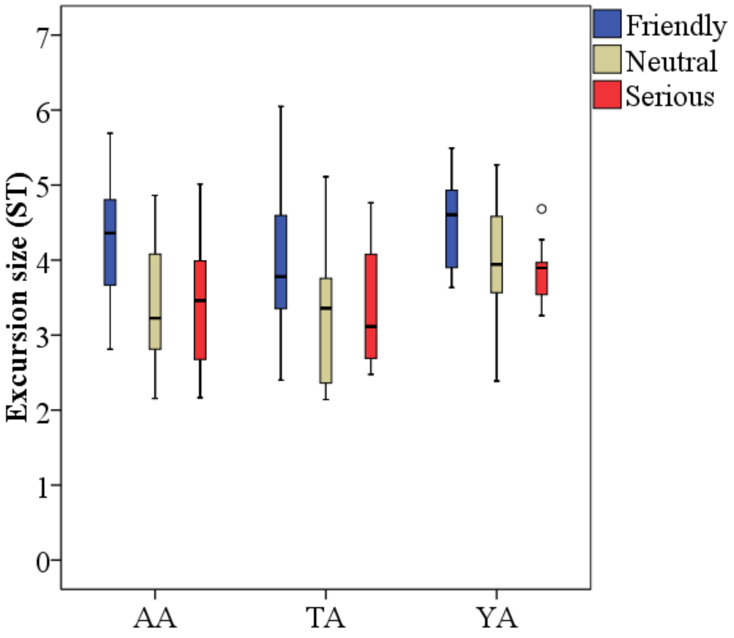

Figure 35: *Effect of excursion size on attitude expression by AA, TA and YA groups*

Figure 30 illustrates the relationship between mean pitch (Hz) and expression of attitude. All of the three groups expressed seriousness with lower mean pitch (Hz) than friendliness. As lower pitch is associated with larger speakers, thus the participants conveyed seriousness sounding as large speakers, and friendliness sounding as smaller speakers. ANOVA results indicate that all the three groups have a significant effect on mean pitch (Hz): the AA group ($F(2, 18) = 8.27$, $p = 0.003$), the TA group ($F(2, 18) = 30.93$, $p < 0.001$), and the YA group ($F(2, 18) = 40.65$, $p < 0.001$).

Figure 31 demonstrates the relationship between mean pitch (ST) and expression of attitude. Similarly, all of the three groups expressed seriousness with lower mean pitch (ST) than friendliness. ANOVA results indicate that all the three groups have a significant effect on mean pitch (ST): the AA group ($F(2, 18) = 11.24$, $p = 0.001$), the TA group ($F(2, 18) = 32.38$, $p < 0.001$), and the YA group ($F(2, 18) = 38.21$, $p < 0.001$).

Figure 32 shows the relationship between intensity and expression of attitude. The AA and the YA expressed seriousness with higher intensity than friendly, thus seriousness is more audible than friendliness for these two groups, while the TA performed the opposite. ANOVA results demonstrate that only the TA and YA groups have a significant effect on intensity: the TA group ($F(2, 18) = 4.31$, $p = 0.030$), and the YA group ($F(2, 18) = 16.73$, $p < 0.001$)

Figure 33 shows the relationship between mean syllable duration and expression of attitude. Similar to the results in emotion task, mean syllable duration is influenced by attitude only for the YA group ($F(2, 18) = 7.95$, $p = 0.003$).

Figure 34 describes the relationship between pitch range and expression of attitude. All the three groups conveyed seriousness with narrower pitch range than friendliness. This is similar to the results from the emotion task, where all the participants performed anger with narrower pitch range than happiness. ANOVA results show that only the AA and TA groups have a significant effect on pitch range: the AA group ($F(2, 18) = 4.94$, $p = 0.020$), and the TA group ($F(2, 18) = 8.04$, $p < 0.003$).

Figure 35 shows the relationship between excursion size and expression of attitude. Similar to pitch range, all the three groups conveyed seriousness with narrower excursion size than friendliness. ANOVA results indicate that only the AA and TA groups have a significant effect on excursion size: the AA group (F(2, 18) = 4.94, *p* = 0.020), and the TA group (F(2, 18) = 8.04, *p* < 0.003).

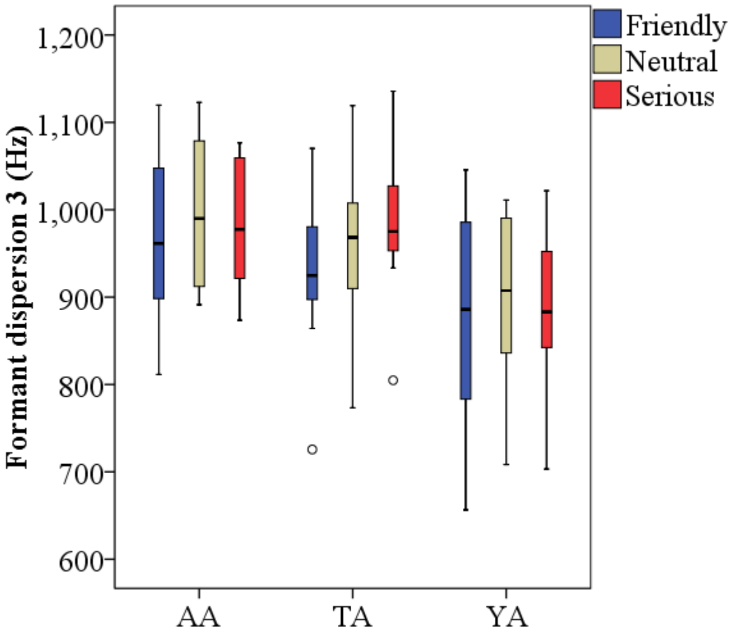

Figure 36: *Effects of formant dispersion 3 on attitude expression by AA, TA and YA groups*

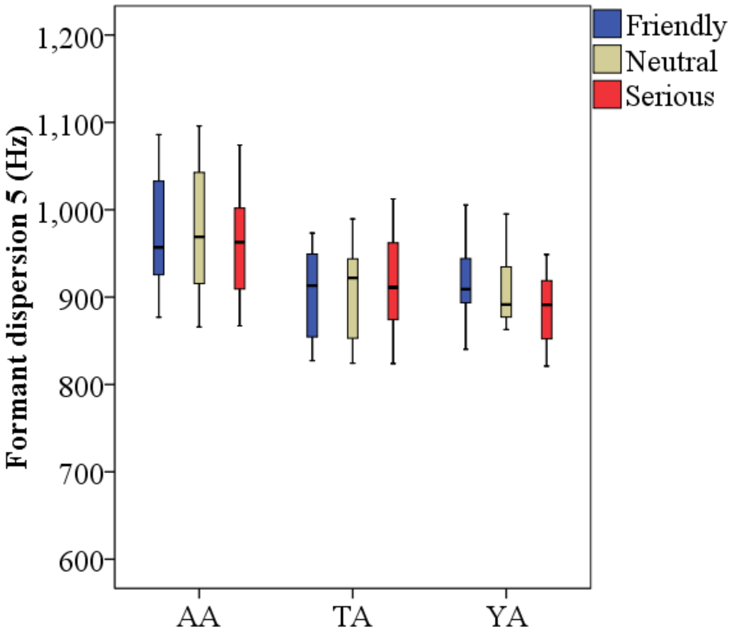

Figure 37: *Effects of formant dispersion 5 on attitude expression by AA, TA and YA groups*

Among the two categories BID-associated parameters, attitude has no significant effects on any VTL-related parameters in any of the three groups. Figures 37 and 38 show the relationship between formant dispersion 3, formant dispersion 5 and expression of attitude. ANOVA results confirmed that only formant dispersion 3 is affected by affect only for the AA group (F(2, 18) = 7.33, *p* = 0.005). The AA group expressed anger with higher formant dispersion 3 than happiness, and neutral emotion with the highest formant dispersion 3. As body size projection predicts that lower formant dispersion characterises longer vocal tract length, the AA group expressed anger and happiness different from these predictions.

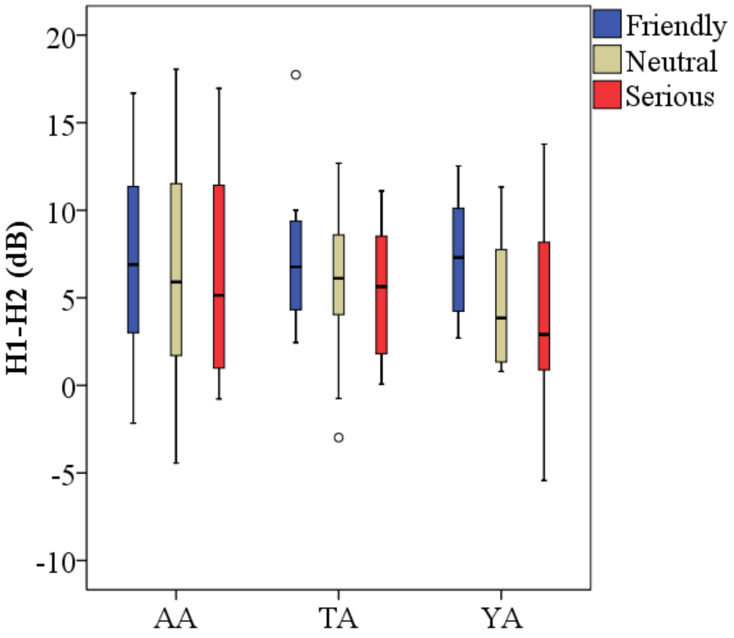

Figure 38: *Effects of H1-H2 on attitude expression by AA, TA and YA groups*

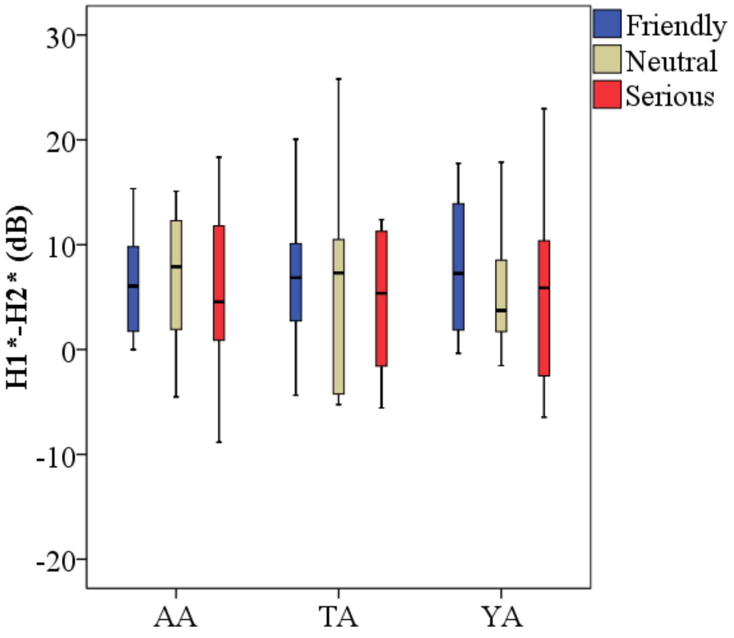

Figure 39: *Effects of H1*-H2* on attitude expression by AA, TA and YA groups*

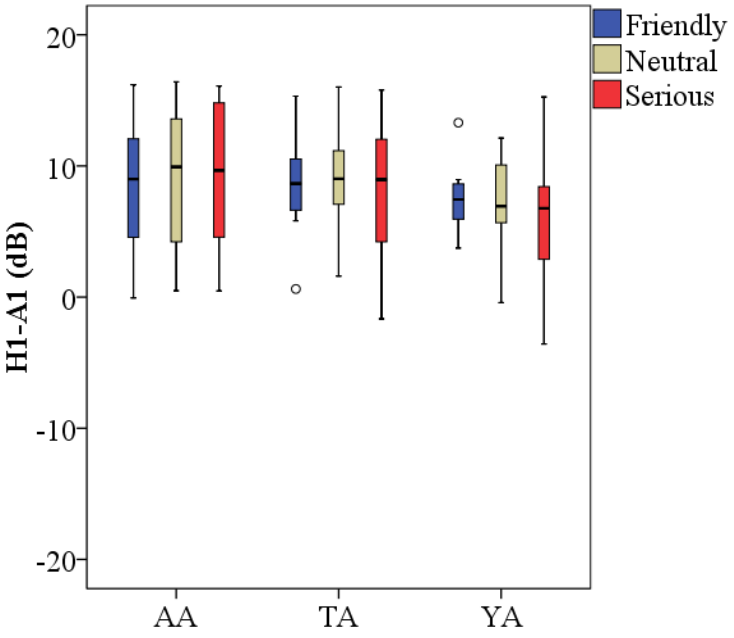

Figure 40: *Effects of H1-A1 on attitude expression by AA, TA and YA groups*

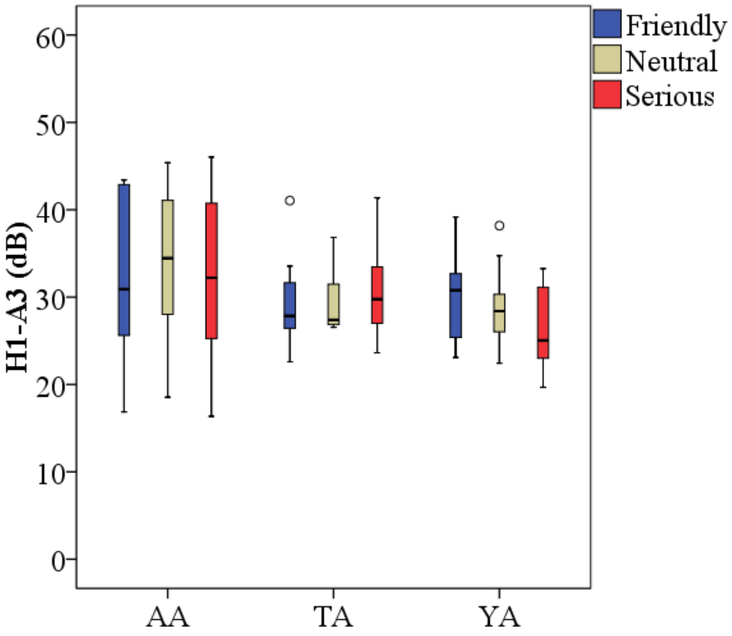

Figure 41: *Effects of H1-A3 on attitude expression by AA, TA and YA groups*

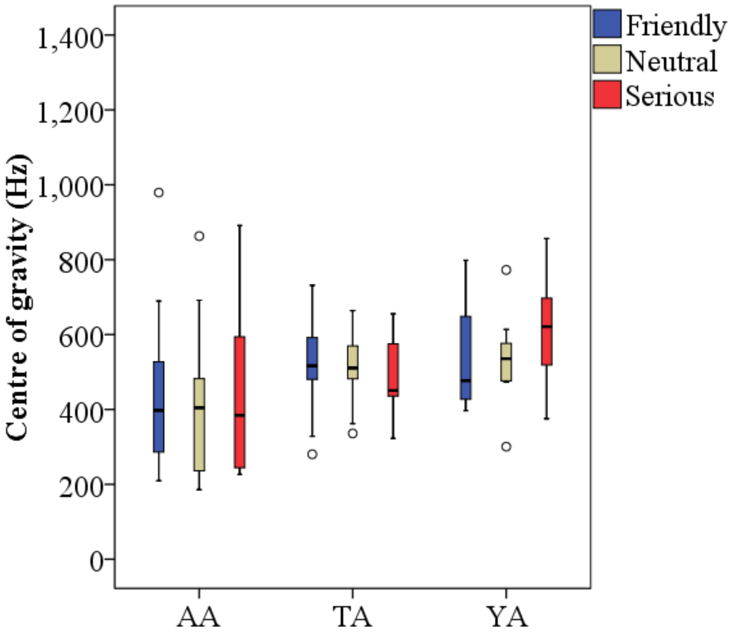

Figure 42: *Effect of centre of gravity (Hz) on attitude expression by AA, TA and YA groups*

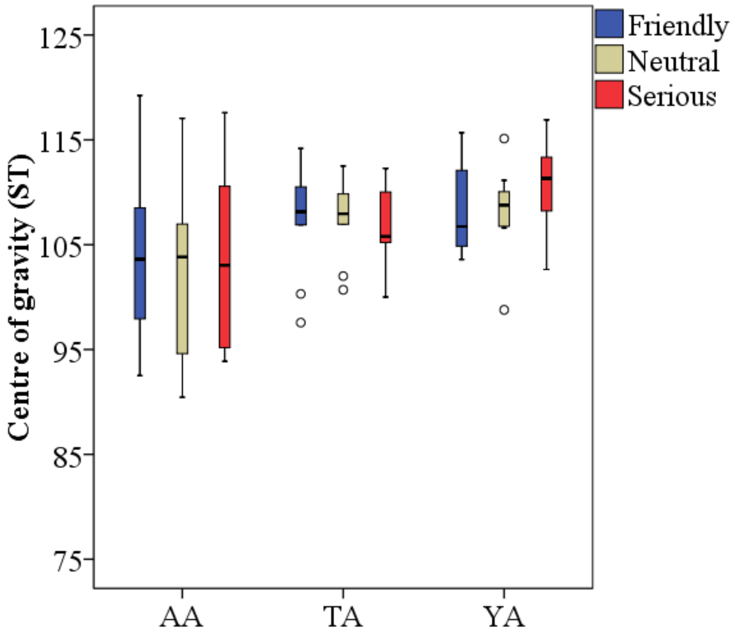

Figure 43: *Effect of centre of gravity (ST) on attitude expression by AA, TA and YA groups*

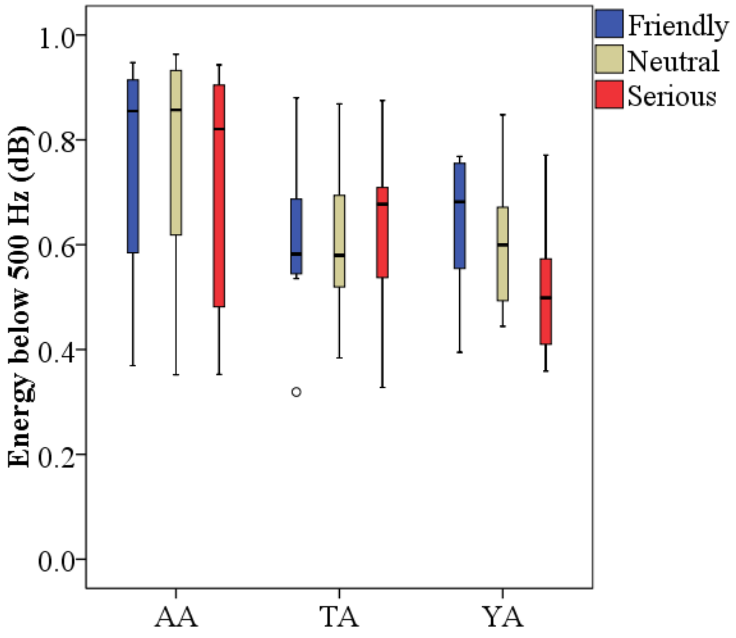

Figure 44: *Effect of energy below 500 Hz on attitude expression by AA, TA and YA groups*

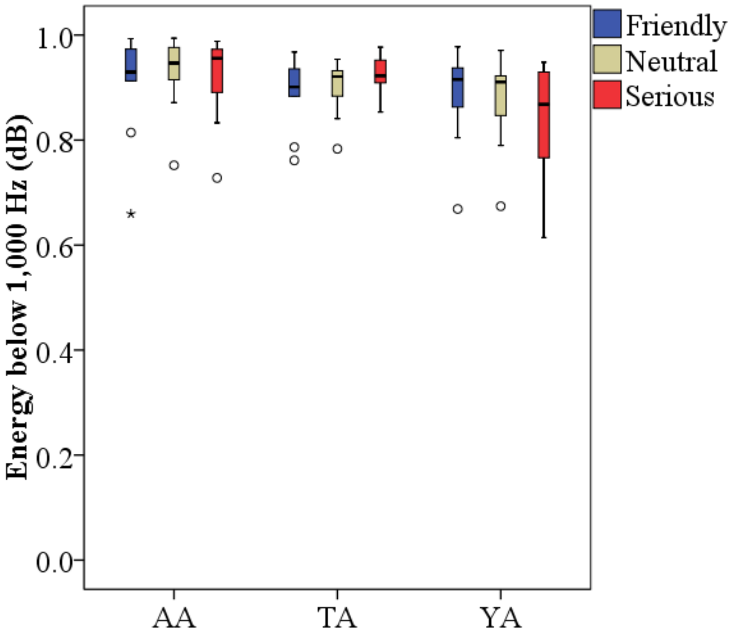

Figure 45: *Effect of energy below 1,000 Hz on attitude expression by AA, TA and YA groups*

Figures 39 and 40 show the relationship between H1*-H2*, H1-A1 and expression of attitude. ANOVA results indicate that none of these acoustic parameters are affected by attitude for any of the three groups.

Figure 38 and 41 depict the relationship between H1-H2, H1-A3 and expression of attitude. The AA and TA groups expressed seriousness with higher mean H1-H3 values than friendliness, whereas the YA group performed seriousness with lower H1-A3 and lower H1-H2 than friendliness. As lower H1-A3 and H1-H2 indicate more pressed voice quality, which is associated with larger speakers, only the YA group conveyed seriousness sounding as smaller speakers, and friendliness sounding as large speakers. ANOVA results confirm that only the YA group has a significant effect on H1-H2 (F(2, 18) = 5.23, p = 0.016) and H1-H3(F(2, 18) = 4.16, p = 0.033).

Figure 42 describes the relationship between CG (Hz) and expression of attitude. Both the AA and TA groups expressed seriousness with lower CG (Hz) than friendliness, yet the YA group produced the opposite. Since higher CG (Hz) are associated with more pressed voice quality, projecting a larger speaker, only the YA group conveyed seriousness sounding as a large person, and friendliness sounding as a smaller person. Further ANOVA results confirm that only the YA group (F(2, 18) = 7.47, *p* = 0.004) has a significant effect on CG (Hz).

Figure 43 shows the relationship between CG (ST) and expression of attitude. Only the YA group expressed seriousness with higher CG (ST) than friendliness, while the AA and Ta groups did the opposite. Further ANOVA results indicates that only the YA group (F(2, 18) = 6.03, *p* = 0.010) has a significant effect on CG (ST).

Figure 44 illustrates the relationship between energy below 500 Hz and expression of attitude. Both the AA and TA groups produced seriousness with higher energy below 500 Hz than friendliness, but the YA performed the opposite, by expressing seriousness with lower energy below 500 Hz than friendliness. As lower energy below 500 Hz indicates pressed voice quality, which is associated with larger speakers, thus only the YA expressed seriousness sounding as a large speaker, and friendliness sounding as a smaller speaker. ANOVA results confirmed that only the YA group (F(2, 18) = 7.21, *p* = 0.005) has a significant effect on energy below 500 Hz.

Figure 45 describes the relationship between energy below 1,000 Hz and expression of attitude. Similar to energy below 500, both the AA and TA groups produced seriousness with higher mean energy below 1,000 Hz values than friendliness, but the YA performed the opposite, by expressing seriousness with lower mean energy below 1,000 Hz values than friendliness. ANOVA results show that only the TA and YA groups have a significant effect on energy below 1,000 Hz: the TA group (F(2, 18) = 4.08, *p* = 0.034), and the YA group (F(2, 18) = 5.83, *p* = 0.011).

Within the acoustic parameters described above, no significant main group effects were found, suggesting the three groups performed the majority of acoustic parameters similarly. The main effect of attitudes are found in mean pitch (Hz) (F(2, 54)

= 51.90, p < 0.001), mean pitch (ST) (F(2, 54) = 64.20, p < 0.001), intensity (F(2, 54) = 8.55, p = 0.001), mean syllable duration (F(2, 54) = 5.19, p = 0.009), pitch range (F(2, 18) = 13.73, p < 0.001), excursion size (F(2, 54) = 11.68, p < 0.001), and H1-H2 (F(2, 54) = 8.95 p < 0.001). These results imply that the participants were producing these acoustic parameters differently for expressing each attitude – seriousness is expressed with lower mean pitch, higher intensity, narrower pitch range, narrower excursion size and decreased H1-H2 than friendliness. Significant interactions between group and attitude are found in H1-A3 (F(4, 54) = 2.61, p = 0.046), CG (Hz) (F(4, 54) = 3.41, p = 0.015), CG (ST) (F(4, 54) = 3.23, p = 0.019), signal below 500 Hz (F(4, 54) = 3.61, p = 0.011), and signal energy below 1,000 Hz (F(4, 54) = 3.56, p = 0.012), thus the three groups performed these three acoustic parameters significantly different across the three attitudes. The results of attitude expression by the AA, TA and YA groups are summarised in the Table 13.

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Mean pitch (Hz)	friendly	184.78	59.60	177.23	44.45	206.09	52.23
	neutral	158.22	33.85	151.65	38.82	170.83	51.43
	serious	149.65	34.48	146.45	35.25	167.34	46.75
	F ratio	F(2, 18) = 8.27		F(2, 18) = 30.93		F(2, 18) = 40.65	
	p-value	p = 0.003		p < 0.001		p < 0.001	

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Mean pitch (ST)	friendly	89.73	4.59	89.14	4.35	91.70	4.65
	neutral	87.36	3.33	86.42	4.44	88.26	5.35
	serious	86.34	3.64	85.90	4.01	88.01	4.97
	F ratio	F(2, 18) = 11.24		F(2, 18) = 32.38		F(2, 18) = 38.21	
	p-value	$p = 0.001$		$p < 0.001$		$p < 0.001$	
Intensity	friendly	76.25	4.66	76.45	4.27	74.44	4.07
	neutral	75.24	4.75	74.16	4.08	72.57	4.47
	serious	76.20	5.55	74.99	4.74	75.42	4.35
	F ratio	*n. s.*		F(2, 18) = 4.31		F(2, 18) = 16.73	
	p-value			$p = 0.030$		$p < 0.001$	
Mean syllable duration	friendly	213.59	47.72	220.67	39.00	244.15	59.28
	neutral	227.35	41.75	233.79	50.24	256.33	67.92
	serious	222.81	58.56	220.82	38.06	234.27	56.41
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 7.95	
	p-value					$p = 0.003$	

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Pitch range	friendly	81.56	40.23	65.43	26.34	80.48	30.28
	neutral	56.82	26.48	51.80	26.71	66.60	38.69
	serious	53.67	28.81	48.62	18.99	64.82	29.21
	F ratio	F(2, 18) = 4.94		F(2, 18) = 8.04		*n. s.*	
	p-value	$p = 0.020$		$p < 0.003$			
Excursion size	friendly	4.27	1.05	3.96	1.15	4.50	0.79
	neutral	3.38	0.87	3.27	0.99	3.93	0.96
	serious	3.41	1.08	3.36	0.90	3.85	0.65
	F ratio	F(2, 18) = 4.94		F(2, 18) = 8.04		*n. s.*	
	p-value	$p = 0.020$		$p < 0.003$			
Formant dispersion 3	friendly	974.45	107.12	927.79	103.92	872.68	136.05
	neutral	997.23	97.67	957.21	98.90	900.54	105.23
	serious	985.00	84.23	981.73	94.56	892.10	102.82
	F ratio	F(2, 18) = 7.33		*n. s.*		*n. s.*	
	p-value	$p = 0.005$					

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Formant dispersion 5	friendly	974.03	74.40	901.05	62.43	918.18	52.58
	neutral	973.38	81.36	905.41	67.20	909.21	46.39
	serious	959.61	66.00	917.16	66.61	887.70	47.21
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
H1-H2	friendly	7.29	6.09	7.62	4.52	7.62	3.93
	neutral	6.42	6.51	5.78	5.31	4.61	3.66
	serious	6.45	6.28	5.34	4.18	4.10	5.61
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 5.23	
	p-value					p = 0.016	
H1*-H2*	friendly	6.31	7.18	7.53	8.93	7.75	8.66
	neutral	6.92	8.89	7.14	10.96	5.74	8.92
	serious	6.14	10.18	4.40	8.03	5.33	11.05
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
H1-A1	friendly	8.64	5.75	8.87	4.44	7.47	3.08
	neutral	9.36	5.38	8.84	4.70	7.24	3.97
	serious	9.31	5.48	8.16	5.49	5.91	5.28
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
H1-A3	friendly	31.76	9.60	29.34	5.24	29.94	6.11
	neutral	33.55	9.17	29.20	4.58	28.95	4.75
	serious	32.63	9.39	30.39	5.78	26.40	4.63
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 4.16	
	p-value					p = 0.033	
Centre of gravity (Hz)	friendly	453.43	238.01	517.42	139.40	522.76	137.02
	neutral	424.95	217.25	506.70	106.94	535.79	125.10
	serious	442.74	222.88	474.72	114.24	621.35	158.96
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 7.47	
	p-value					p = 0.004	

Table 13. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by AA, TA and YA groups

		AA (n=10)		TA (n=10)		YA (n=10)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Centre of gravity (ST)	friendly	103.95	8.48	107.52	5.18	107.86	4.24
	neutral	102.83	8.59	107.42	3.87	108.32	4.31
	serious	103.58	8.43	106.21	4.31	110.81	4.55
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 6.03	
	p-value					*p* = 0.010	
Energy below 500 Hz	friendly	0.74	0.22	0.62	0.17	0.64	0.14
	neutral	0.75	0.24	0.61	0.15	0.61	0.13
	serious	0.73	0.22	0.64	0.18	0.52	0.15
	F ratio	*n. s.*		*n. s.*		F(2, 18) = 7.21	
	p-value					*p* = 0.005	
Energy below 1,000 Hz	friendly	0.91	0.10	0.89	0.07	0.88	0.09
	neutral	0.93	0.07	0.90	0.06	0.88	0.09
	serious	0.92	0.08	0.92	0.05	0.83	0.12
	F ratio	*n. s.*		F(2, 18) = 4.08		F(2, 18) = 5.83	
	p-value			*p* = 0.034		*p* = 0.011	

4.3 Perception of PD patients, healthy elderly and young adults

The responses of participants in the PD, HE and YA groups were automatically collected by Praat (Boersma & Weenink, 2013) as tables and saved by the experimenter. The responses obtained in the three experiments were coded as follows:

- Small as 0% large, medium as 50% large, and large as 100% large.
- Happy as 0% angry, neutral as 50%, and angry as 100% angry.
- Friendly as 0% serious, neutral as 50% serious, and serious as 100% serious.

The coded responses of each experiment were first analysed with four 3x4 mixed design ANOVAs, one for each acoustic parameters – voice quality, formant shift ratio, pitch shift and pitch range, using SPSS version 22.0 (IBM Corp., 2013), with the three groups (AA, TA and YA) as between-subjects factor and three manipulation levels as within-subjects factors. In order to find the main effects of each acoustic parameter on each group, the responses of each group in each experiment were also independently analysed with 12 independent one-way ANOVAs. A *p*-value < 0.050 was taken as significant.

4.3.1 Judgement of body size by PD, HE and YA groups

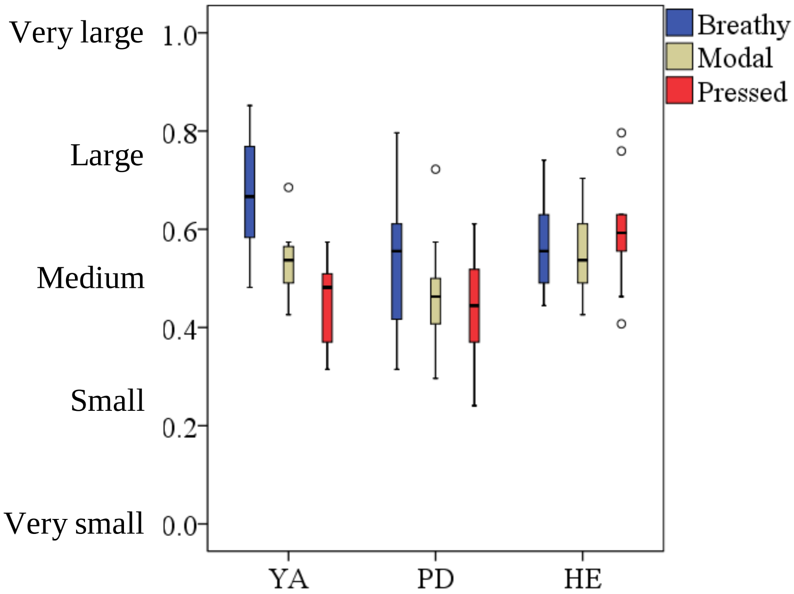

Figure 46: *Effects of voice quality on body size judgment by PD, HE and YA*

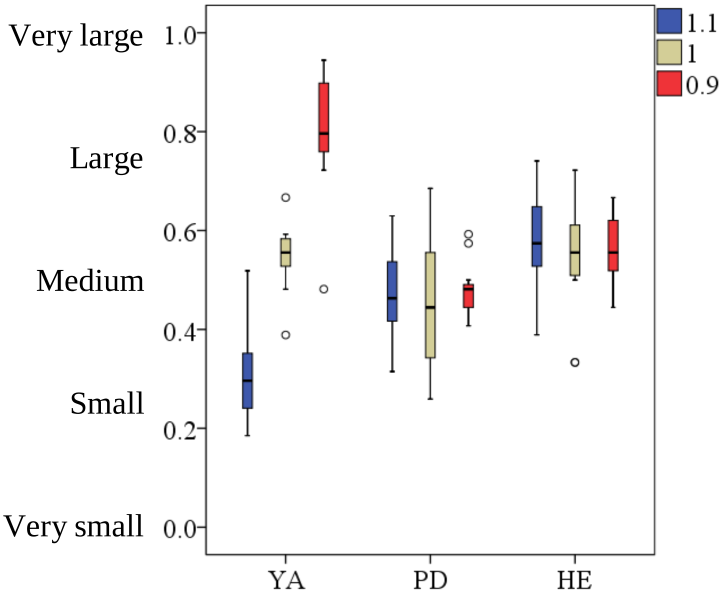

Figure 47: *Effects of formant shift ratio on body size judgment by PD, HE and YA groups*

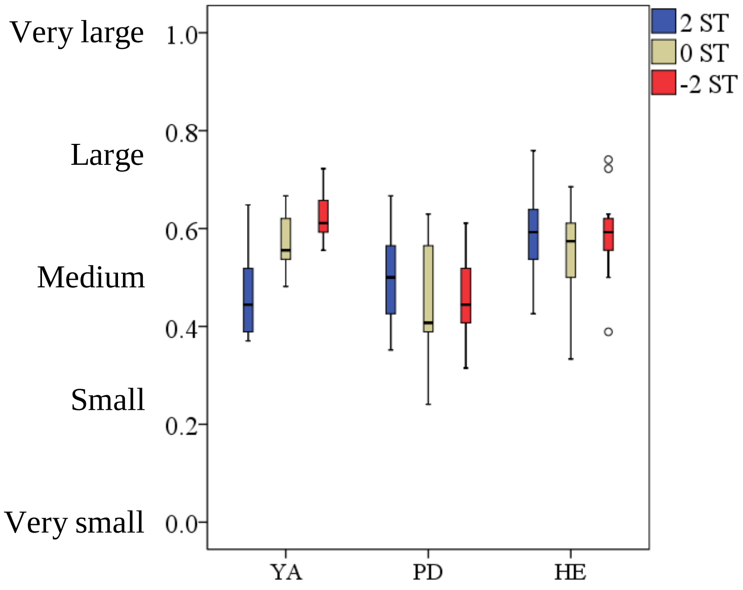

Figure 48: *Effects of pitch shift on body size judgment by PD, HE and YA groups*

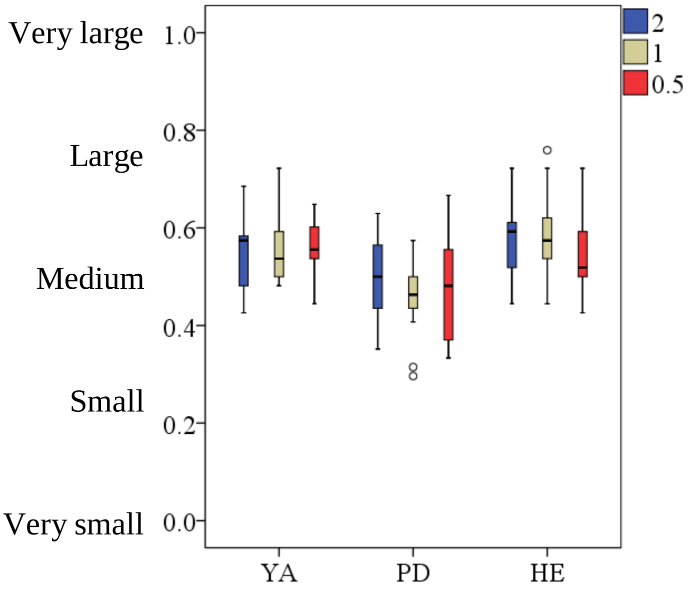

Figure 49: *Effects of pitch range on body size judgment by PD, HE and YA groups*

Figure 46 shows the relationship between voice quality and judgment of body size. Voice quality had a significant effect only for the YA group (F(2, 20) = 14.53, $p < 0.001$) in body size judgement. However, a more detailed examination revealed that they perceived breathy voice quality as from a larger speaker, and pressed voice quality as from a smaller one. Yet, this result contradicts the findings of Xu, Lee, et al. (2013).

Figure 47 displays the relationship between formant shift ratio and judgment of body size. Only the YA group rated utterances with a smaller formant shift ratio as having a larger body size, and a larger formant shift ratio as having a smaller size. The PD and HE groups did not make much differentiated judgment of body size based on this acoustic parameter. ANOVA results confirmed that formant shift ratio had a significant effect only for the YA group (F(2, 20) = 52.22, $p < 0.001$).

Figure 48 shows the relationship between pitch shift and judgment of body size. The YA groups judged downward pitch shift as from a larger speaker, and upward pitch shift as from a smaller one. However, the opposite applies to the PD and HE groups – they perceived upward pitch shift as sounding large, and downward pitch shift as sounding small. ANOVA results once again indicated this acoustic parameter has a significant effect only for the YA group (F(2, 20) = 21.11, $p < 0.001$). Pitch range is the only acoustic parameter with no significant effect among the three groups (Figure 49).

In general, there is a main group effect in body size judgment (F(2, 30) = 5.96, $p = 0.007$), demonstrating that the three groups performed this task differently. Considering the four manipulated acoustic parameters, the main effect of voice quality (F(2, 60) = 8.45, $p = 0.001$), formant shift ratio (F(2, 60) = 26.449, $p < 0.001$), and pitch shift (F(2, 60) = 3.19, $p < 0.048$) achieved statistical significance. Therefore, these three acoustic parameters were most imperative for the participants of the present study in judging the speaker's body size - the PD group perceived synthetic speech with breathy voice quality, larger formant shift ratio and upward pitch shift as from a larger person, the HE group judged stimuli with pressed voice quality, larger formant shift and downward pitch shift as from a larger speaker, while the YA group perceived the utterances with breathy voice quality, smaller formant shift ratio and downward pitch shift as from a larger person. Significant interactions of group and acoustic parameter have been found for voice quality (F(4, 60) = 5.24, $p = 0.001$) and formant shift ratio (F(4, 54) = 27.79, $p < 0.001$). These results indicate that the three groups responded differently to voice quality and formant shift ratio. The results of body size judgment by the PD, HE and YA groups are summarised in the Table 14.

Table 14. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in body size judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		size	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	small	0.68	0.12	0.52	0.14	0.56	0.07
	modal		0.53	0.07	0.47	0.11	0.56	0.09
	pressed	large	0.45	0.09	0.43	0.11	0.60	0.13
	F-ratio p-value		F(2, 20) = 14.53 $p < 0.001$		*n. s.*		*n. s.*	
Formant shift ratio	1.1	small	0.31	0.10	0.47	0.10	0.58	0.10
	1.0		0.55	0.07	0.46	0.13	0.55	0.13
	0.9	large	0.80	0.13	0.48	0.06	0.56	0.08
	F-ratio p-value		F(2, 20) = 52.22 $p < 0.001$		*n. s.*		*n. s.*	

Table 14. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in body size judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		size	Mean	SD	Mean	SD	Mean	SD
Pitch shift	+2 ST	small	0.47	0.09	0.50	0.10	0.58	0.10
	0 ST		0.58	0.06	0.46	0.12	0.55	0.10
	-2 ST	large	0.62	0.05	0.46	0.08	0.59	0.10
	F-ratio p-value		F(2, 20) = 21.11 $p < 0.001$		*n. s.*		*n. s.*	
Pitch range	2.0	small	0.55	0.08	0.50	0.09	0.57	0.09
	1.0		0.56	0.07	0.46	0.09	0.59	0.09
	0.5	large	0.56	0.06	0.47	0.11	0.56	0.09
	F-ratio p-value		*n. s.*		*n. s.*		*n. s.*	

4.3.2 Judgement of emotion by PD, HE and YA groups

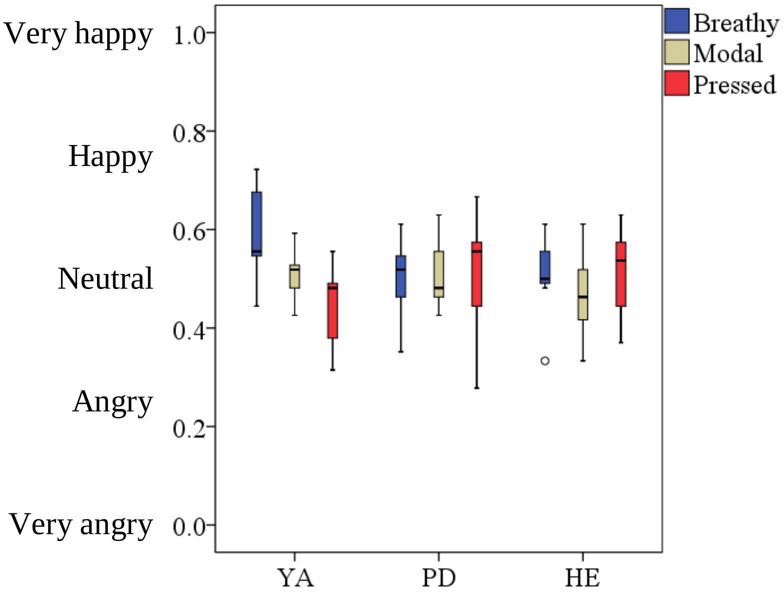

Figure 50: *Effects of voice quality on emotion judgment by PD, HE and YA groups*

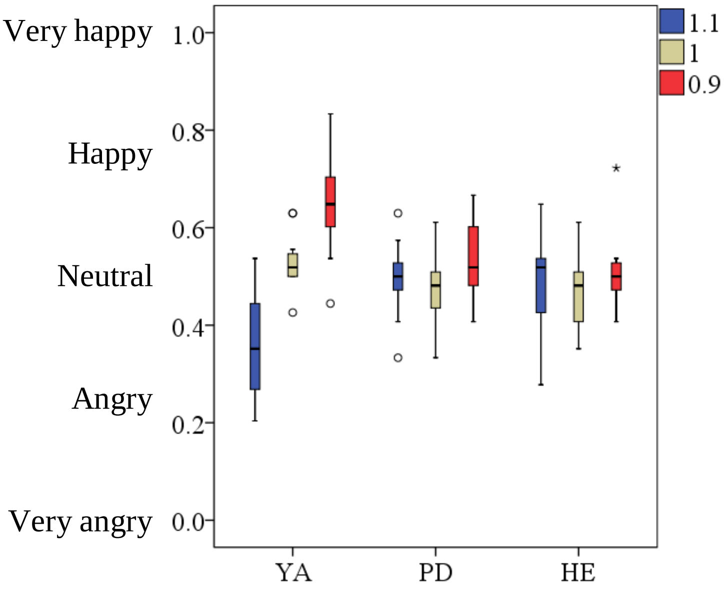

Figure 51: *Effects of formant shift ratio on emotion judgment by PD, HE and YA groups*

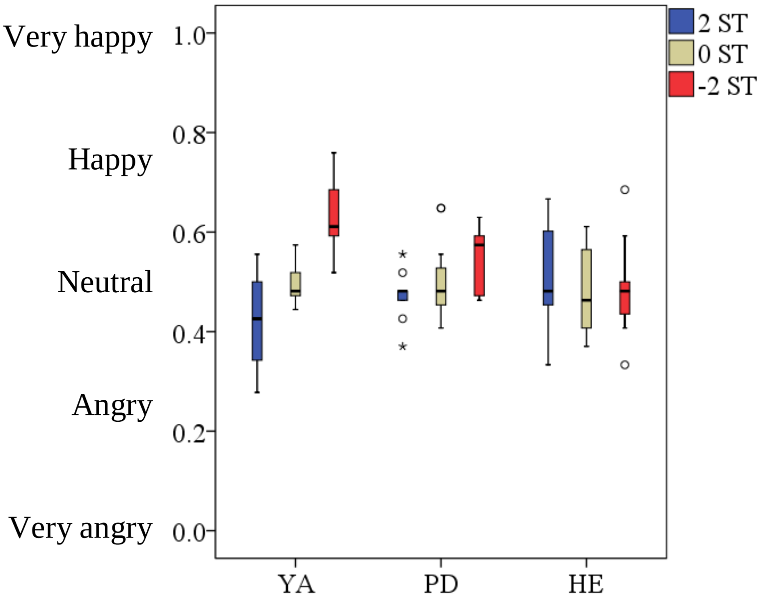

Figure 52: *Effect of pitch shift on emotion judgment by PD, HE and YA groups*

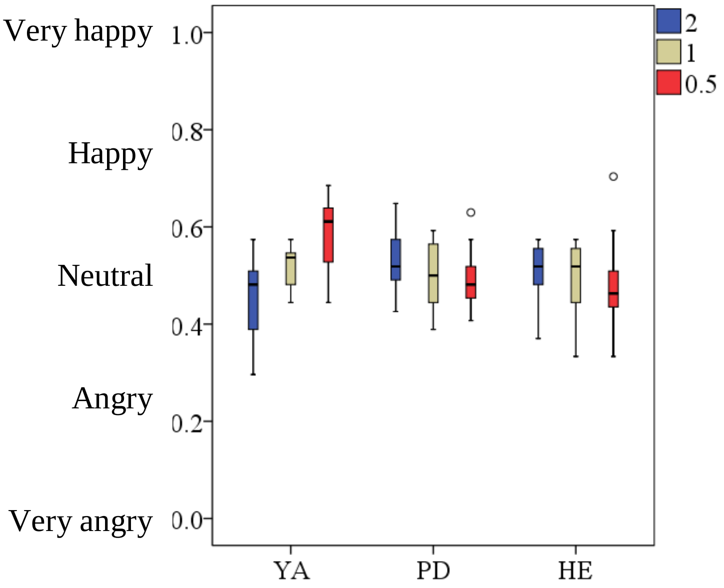

Figure 53: *Effects of pitch range on emotion judgment by PD, HE and YA groups*

Figure 50 displays the relationship between voice quality and judgment of emotion. Similar to the body size judgment, the effect of voice quality reached statistical significance only for the YA group (F(2, 20) = 9.10, *p* = 0.002), but they perceived pressed voice as angry-sounding and breathy happy-sounding, which contradicts the findings by Chuenwattanapranithi et al. (2008) and Xu, Lee, et al. (2013).

Figure 51 displays the relationship between formant shift ratio and judgment of emotion. Only the YA group judged smaller formant shift ratio as sounding angry, and greater formant shift ratio as happy. Similarly to the body size judgement, the PD and HE groups did not rate the speaker's emotion based on this acoustic parameter. ANOVA results confirmed that formant shift ratio had a significant effect only on the YA group (F(2, 20) = 20.50, $p < 0.001$).

Figure 52 shows the relationship between pitch shift and judgment of emotion. The PD and YA groups perceived downward pitch shift as angry, and upward pitch shift as happy. ANOVA results indicate pitch shift had a significant effect on the PD (F(2, 20) = 6.21, $p = 0.008$) and YA (F(2, 20) = 17.91, $p < 0.001$) groups.

Figure 53 demonstrates the relationship between pitch range and judgment of emotion. The YA group rated narrow pitch range as angry-sounding, while wide pitch range as happy-sounding. However, the PD and the HE groups made the opposite judgment. ANOVA results show that pitch range has a significant effect only for the YA group (F(2, 20) = 7.97, $p = 0.003$).

There was no main group effect for the emotion judgment, suggesting that in general the three groups performed likewise. The main effect of three acoustic parameters reached statistical significance – voice quality (F(2, 60) = 4.66, $p = 0.013$), formant shift ratio (F(2, 60) =15.29, $p < 0.001$), and pitch shift (F(2, 60) = 10.83, $p < 0.001$). These results demonstrated that the three groups were responsive to the changes in these acoustic parameters when making emotion judgments – the PD group perceived stimuli with smaller formant shift, downward pitch shift and wider pitch range as sounding angry, the HE group judged utterances with smaller formant shift ration, upward pitch shift and wider pitch range as from a angry-sounding speakers, which the YA group associated breathy voice quality, smaller formant shift ratio, downward pitch shift and narrower pitch range with anger. The interactions between group acoustic parameters were all significant – voice quality (F(4, 60) = 3.96, $p = 0.006$), between group and formant shift ratio (F(4, 60) = 8.93, $p < 0.001$), between group and pitch shift (F(4, 60) = 7.42, $p < 0.001$), and between group and pitch range (F(4, 60) = 5.87, $p < 0.001$). Therefore, the three groups responded differently to all the four manipulated acoustic parameters. The results of emotion judgment by the PD, HE and YA groups are summarised in the Table 15.

Table 15. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in emotion judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		emotion	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	happy	0.60	0.09	0.51	0.08	0.51	0.07
	modal	neutral	0.51	0.05	0.51	0.07	0.46	0.08
	pressed	angry	0.45	0.08	0.51	0.11	0.51	0.09
	F-ratio		F(2, 20) = 9.10		*n. s.*		*n. s.*	
	p-value		$p = 0.002$					
Formant shift ratio	1.1	happy	0.36	0.11	0.50	0.08	0.49	0.11
	1.0	neutral	0.53	0.59	0.47	0.08	0.47	0.08
	0.9	angry	0.65	0.11	0.54	0.08	0.51	0.08
	F-ratio		F(2, 20) = 20.50		*n. s.*		*n. s.*	
	p-value		$p < 0.001$					

Table 15. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in emotion judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		emotion	Mean	SD	Mean	SD	Mean	SD
Pitch shift	+2 ST	happy	0.42	0.10	0.47	0.05	0.52	0.10
	0 ST	neutral	0.50	0.04	0.51	0.08	0.49	0.09
	-2 ST	angry	0.63	0.08	0.55	0.07	0.48	0.09
	F-ratio		F(2, 20) = 17.91		*n. s.*		F(2, 20) = 6.21	
	p-value		*p* < 0.001				*p* = 0.008	
Pitch range	2.0	happy	0.45	0.08	0.53	0.07	0.51	0.06
	1.0	neutral	0.52	0.05	0.50	0.07	0.49	0.09
	0.5	angry	0.59	0.08	0.49	0.07	0.49	0.10
	F-ratio		F(2, 20) = 7.97		*n. s.*		*n. s.*	
	p-value		*p* = 0.003					

4.3.3 Judgement of attitude by PD, HE and YA groups

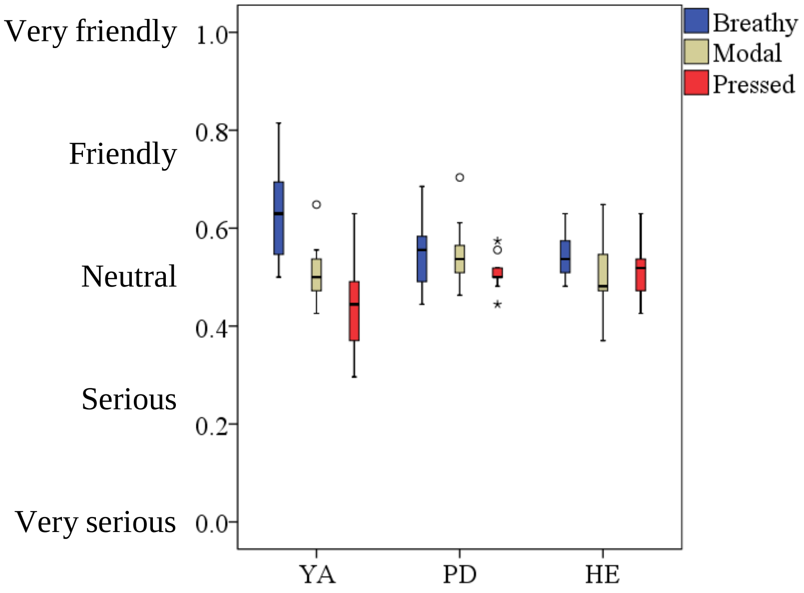

Figure 54*: Effects of voice quality on attitude judgment by PD, HE and YA groups*

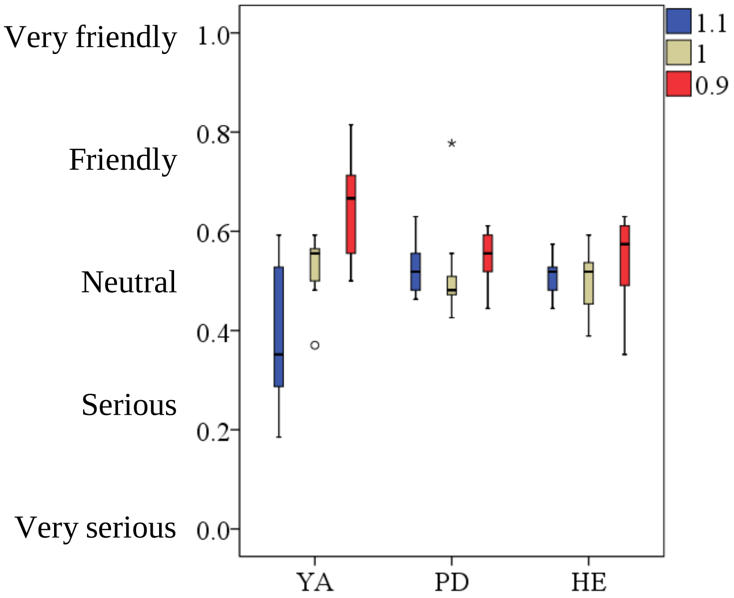

Figure 55*: Effects of formant shift ratio on attitude judgment by PD, HE and YA groups*

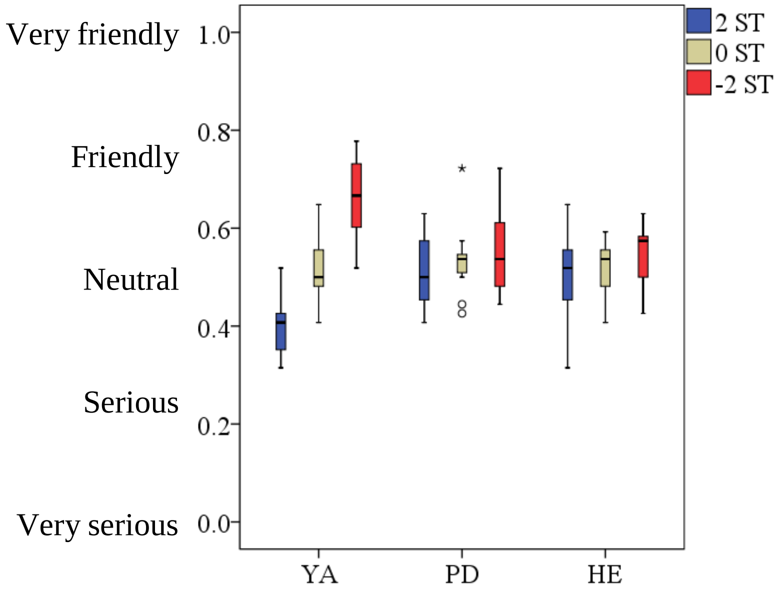

Figure 56: *Effects of pitch shift on attitude judgment by PD, HE and YA groups*

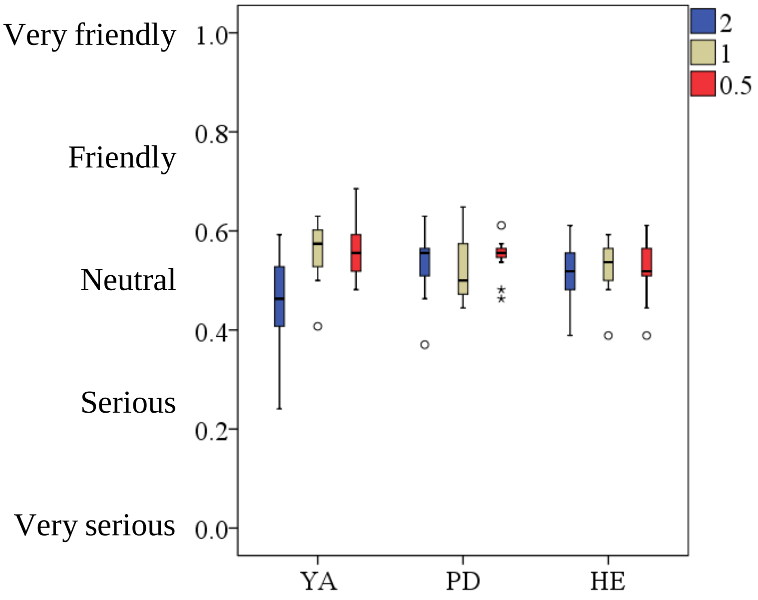

Figure 57: *Effects of pitch range on attitude judgment by PD, HE and YA groups*

Figure 54 illustrates the relationship between voice quality and judgment of attitude. As found in the emotion and the body size judgments, the effect of voice quality reached statistical significance only for the YA group (F(2, 20) = 9.88, p = 0.001). However, they perceived pressed voice quality as sounding serious and breathy as sounding friendly, thus contradicting of the findings in Chuenwattanapranithi et al. (2008) and Xu, Lee, et al. (2013).

Figure 55 shows the relationship between formant shift ratio and judgment of attitude. Similar to the results in emotion judgment task, the YA group rated smaller formant shift ratio as serious-sounding, and larger formant shift ratio as friendly-sounding person. ANOVA results confirm that this acoustic parameter has significant effect on the YA group (F(2, 20) = 11.75, $p < 0.001$). Although the PD and the HE groups also showed this trend, ANOVA results do not show that formant shift ratio had a significant effect for these two groups.

The relationship between pitch shift and judgment of attitude is shown in Figure 56. The YA group perceived downward pitch shift as sounding more serious, while upward pitch shift as friendlier. ANOVA results show that this acoustic parameter has significant effect on the YA group (F(2, 20) = 30.40, $p < 0.001$). Yet, ANOVA results show no significant effect of pitch shift on the PD and HE groups, both groups also judged changes in this acoustic parameter similar to the YA group.

Figure 57 demonstrates the relationship between pitch range and judgment of attitude. The YA group is the only one judged wider pitch range as from friendly speaker, and narrower pitch range as from friendly speaker. ANOVA results show that the effect of this acoustic parameter is significant only for the YA group (F(2, 20) = 5.35, $p = 0.014$).

Similar to the emotion judgment, there was no significant main group effect for the attitude judgment, indicating the three groups did not differ much in their performance. The main effect of three out of four manipulated acoustic parameters reached statistical significance: formant shift ratio (F(2, 54) = 43.07, $p < 0.001$), pitch shift (F(2, 54) = 43.74, $p < 0.001$), and pitch range (F(2, 54) = 9.84, $p < 0.001$).

Therefore, the participants of the current study were all sensitive to the changes of these parameters – they judged synthetic speech with breathy voice quality, smaller formant shift, downward pitch shift and narrower pitch range as from angry sounding speakers. Once again, only the interaction between group and voice quality has reached significance (F(4, 54) = 3.55, $p = 0.012$), indicating that the three groups responded differently to this parameter. The results of attitude judgment by the PD, HE and YA groups are summarised in the Table 16.

Table 16. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in attitude judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		attitude	Mean	SD	Mean	SD	Mean	SD
Voice quality	breathy	friendly	0.63	0.11	0.55	0.07	0.55	0.05
	modal	neutral	0.51	0.06	0.55	0.07	0.50	0.08
	pressed	serious	0.43	0.10	0.51	0.04	0.51	0.06
	F-ratio p-value		F(2, 20) = 9.88 $p = 0.001$		*n. s.*		*n. s.*	
Formant shift ratio	1.1	friendly	0.39	0.14	0.53	0.05	0.51	0.04
	1.0	neutral	0.53	0.07	0.51	0.09	0.50	0.06
	0.9	serious	0.65	0.10	0.55	0.05	0.54	0.10
	F-ratio p-value		F(2, 20) = 11.75 $p < 0.001$		*n. s.*		*n. s.*	

Table 16. Means, standard deviations, F ratio and p-value for voice quality, formant shift ratio, pitch shift, and pitch range in attitude judgment by PD, HE and YA groups

			YA *(n=11)*		PD *(n=11)*		HE *(n=11)*	
		attitude	Mean	SD	Mean	SD	Mean	SD
Pitch shift	+2 ST	friendly	0.40	0.06	0.51	0.08	0.50	0.09
	0 ST	neutral	0.52	0.07	0.53	0.08	0.52	0.06
	-2 ST	serious	0.66	0.09	0.55	0.09	0.54	0.06
	F-ratio p-value		F(2, 20) = 30.40 $p < 0.001$		*n. s.*		*n. s.*	
Pitch range	2.0	friendly	0.46	0.10	0.53	0.07	0.51	0.06
	1.0	neutral	0.56	0.06	0.52	0.07	0.52	0.06
	0.5	serious	0.56	0.07	0.55	0.04	0.52	0.06
	F-ratio p-value		F(2, 20) = 5.35 $p = 0.014$		*n. s.*		*n. s.*	

4.4 Production of emotion and attitude by PD patients, healthy elderly and young adults

All the speech samples from the participants were analysed with 16 independent 3x3 mixed design ANOVAs, using SPSS version 22.0 (IBM Corp., 2013), with group (PD, HE and YA) as a between-subject factor and different types of emotions (happy, neutral, angry) or attitudes (friendly, neutral, serious) as a within-subject factor for each acoustic parameter (mean pitch (Hz), mean pitch (ST), pitch range, excursion size, intensity, mean syllable duration, formant dispersion 3, formant dispersion 5, H1-H2, H1*-H2* H1-A1, H1-A3, centre of gravity (Hz), centre of gravity (ST), energy below 500 Hz and energy below 1000 Hz). The speech samples were also analysed with 48 independent one-way ANOVAs for individual group, aiming to find the main effects of individual emotion on each acoustic parameter per group. A p-value < 0.050 was taken as significant.

4.4.1 Expression of emotion by PD, HE and YA groups

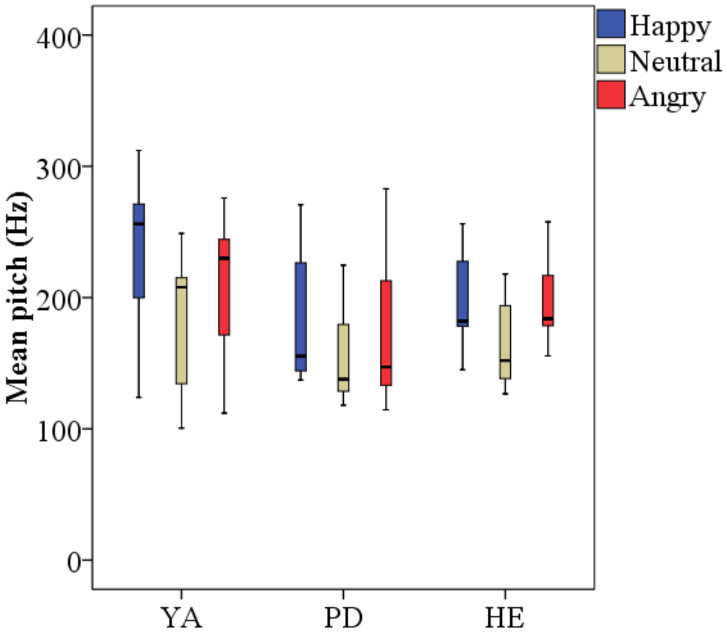

Figure 58: *Effects of mean pitch (Hz) on emotion expression by PD, HE and YA groups*

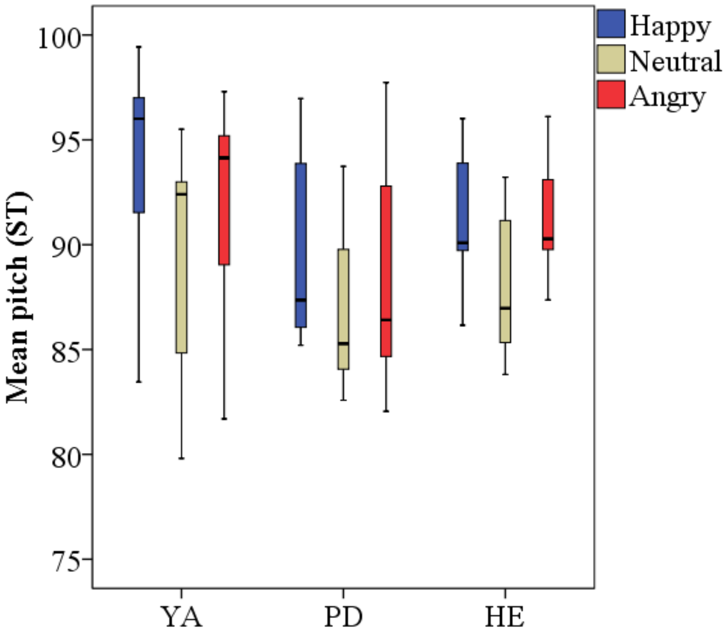

Figure 59: *Effects of mean pitch (ST) on emotion expression by PD, HE and YA groups*

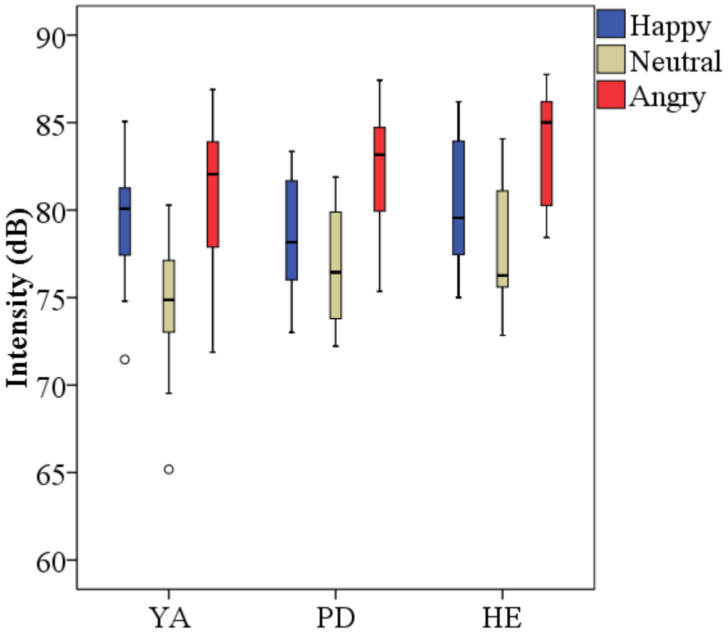

Figure 60: *Effect of intensity on emotion expression by PD, HE and YA groups*

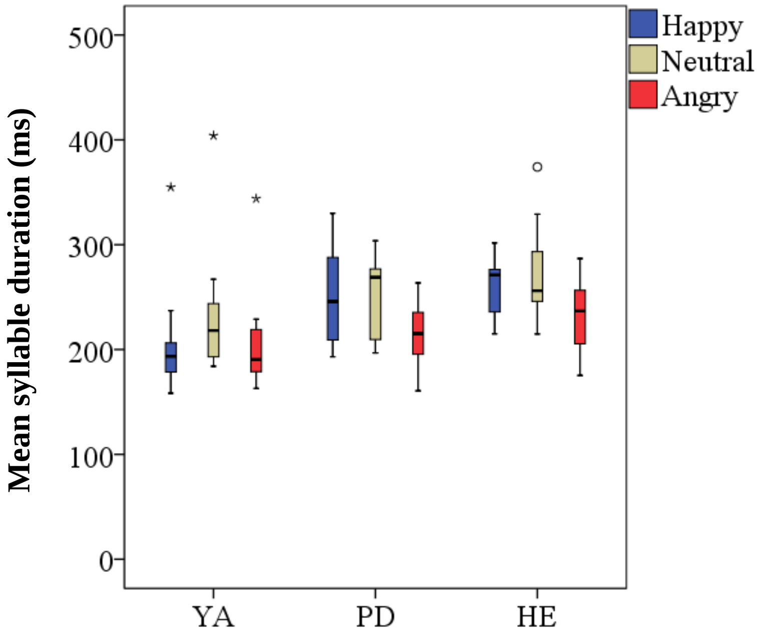

Figure 61: *Effect of mean syllable duration on emotion expression by PD, HE and YA groups*

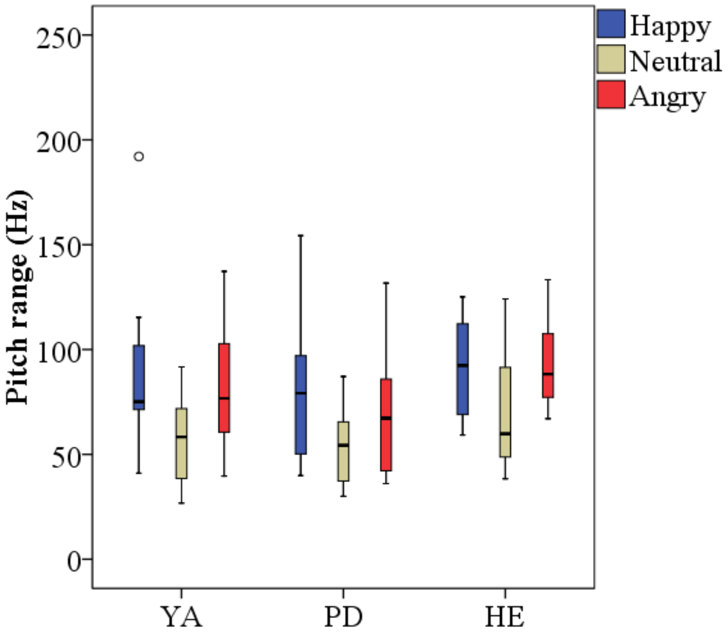

Figure 62: *Effect of pitch range on emotion expression by PD, HE and YA groups*

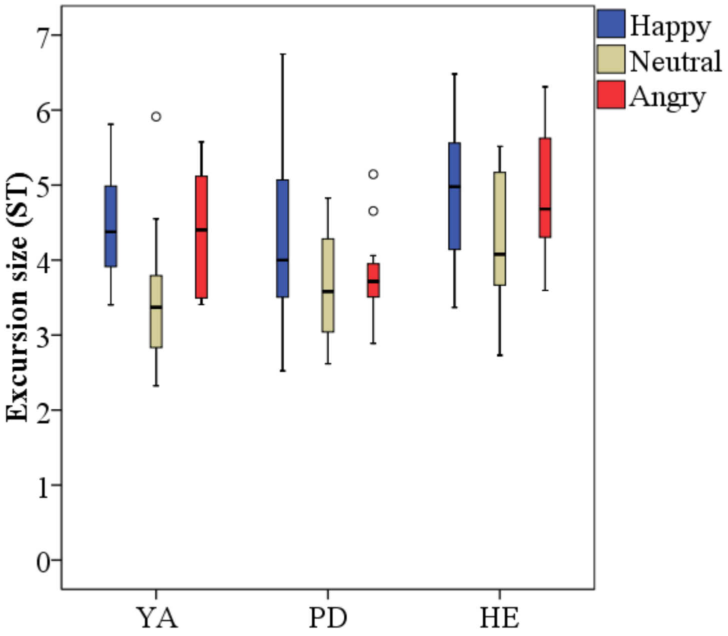

Figure 63: *Effect of excursion size on emotion expression by PD, HE and YA groups*

Figure 58 shows the relationship between mean pitch (Hz) and expression of emotion. The PD and YA groups produced anger with lower mean pitch value than happiness, while the HE group performed both with similar mean pitch (Hz). All the three groups expressed neutral emotion with the lowest mean pitch (Hz). Following the prediction of body-size projection, lower pitch is associated with a larger and angry-sounding person, thus the PD and YA groups performed as predicted when conveying anger with lower mean pitch than happiness. ANOVA results demonstrate that all the three groups have a significant effect on mean pitch (Hz): the PD group ($F(2, 20) = 8.40$, $p = 0.002$), the HE group ($F(2, 20) = 29.86$, $p < 0.001$), and the YA group ($F(2, 20) = 39.35$, $p < 0.001$).

Figure 59 demonstrates the relationship between mean pitch (ST) and expression of emotion. Similar to mean pitch (ST), the PD and YA groups produced anger with lower mean pitch value than happiness, while the HE group performed both with similar mean pitch (ST). All the three groups expressed neutral emotion with the lowest mean pitch (ST). ANOVA results indicates that all the three groups have a significant effect on mean pitch (ST): the PD group (F(2, 20) = 10.27, $p = 0.001$), the HE group (F(2, 20) = 28.44, $p < 0.001$), and the YA group (F(2, 20) = 34.13, $p < 0.001$).

Figure 60 depicts the relationship between intensity and expression of emotion. All the participants performed anger with higher intensity than happiness, and neutral emotion with the lowest intensity. This indicates that anger is more audible than happiness, and both of them are more audible than neutral emotion. ANOVA results show all the three groups have a significance effect on intensity: the PD group (F(2, 20) = 16.62, $p < 0.001$), the HE group (F(2, 20) = 82.02, $p < 0.001$), and the YA group (F(2, 20) = 55.93, $p < 0.001$).

Figure 61 describes the relationship between mean syllable duration and expression of emotion. All the three groups conveyed anger with shorter mean syllable duration than happiness. ANOVA results show the three groups have a significant effect on mean syllable duration: PD group (F(2, 20) = 8.28, $p = 0.002$), HE group (F(2, 20) = 6.89, $p = 0.005$), and YA group (F(2, 20) = 10.13, $p = 0.001$).

Figure 62 demonstrates the relationship between pitch range and expression of emotion. The PD and HE groups expressed anger with narrower pitch range than happiness, whereas the YA group performed the two emotions with similar pitch range. Although body-size projection does not provide particular predictions for pitch range interpretations, however, past research findings in emotion suggested that synthetic speech with wider pitch range was often perceived as from happier speakers (Murray & Arnott, 1993; Nwe et al., 2003). ANOVA results show that all the three groups have a significant effect on pitch range: the PD group (F(2, 20) = 8.81, $p = 0.002$), the HE group (F(2, 20) = 9.18, $p = 0.001$), and the YA group (F(2, 20) = 12.94, $p < 0.001$).

Figure 63 demonstrates the relationship between excursion size and expression of emotion. Similar to pitch range, the PD and HE groups conveyed anger with narrower excursion size than happiness. Yet, the YA group performed the two emotions with similar excursion size. ANOVA results show that all the three groups have a significant effect on excursion size: the PD group (F(2, 20) = 4.65, *p* = 0.022), the HE group (F(2, 20) = 6.35, *p* = 0.007), and the YA group (F(2, 20) = 8.82, *p* = 0.002).

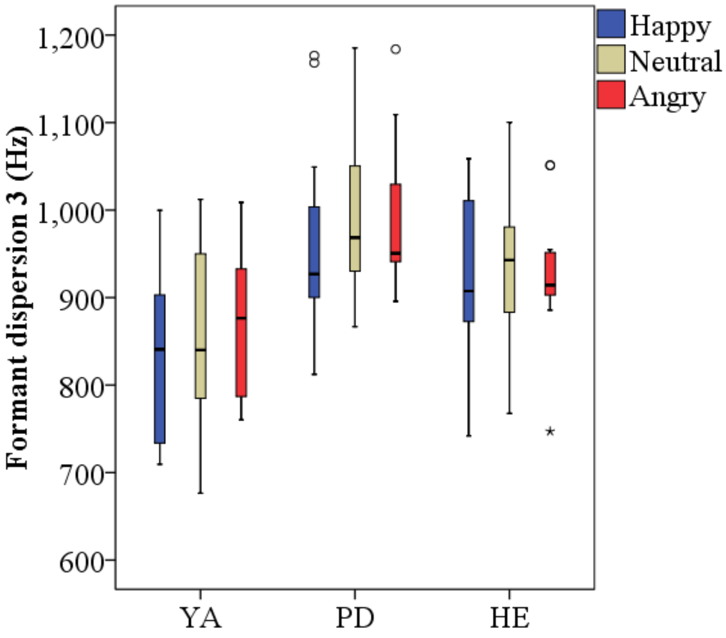

Figure 64: *Effects of formant dispersion 3 on emotion expression by PD, HE and YA groups*

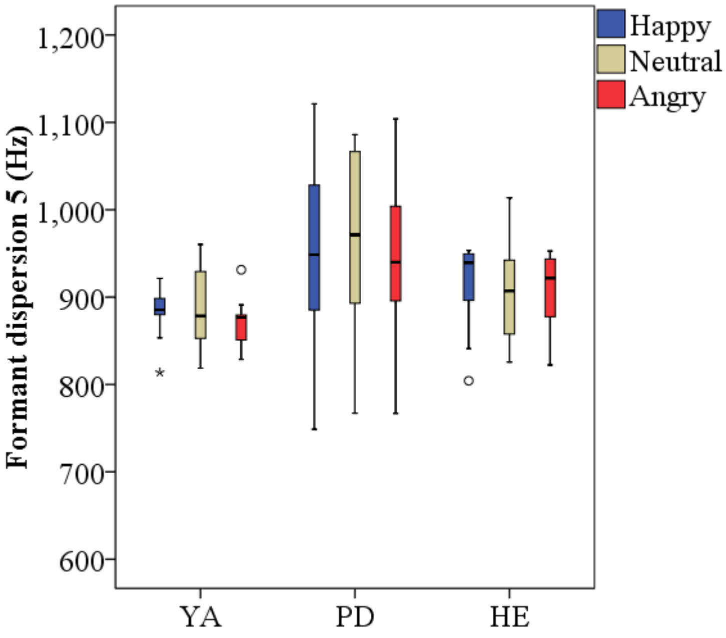

Figure 65: *Effects of formant dispersion 5 on emotion expression by PD, HE and YA groups*

Among the two categories BID-associated parameters, emotion has no significant effect on any VTL-related parameters in any of the three groups. Figures 63 and 64 illustrate the relationship between formant dispersion 3, formant dispersion 5 and expression of emotion. ANOVA results confirmed that formant dispersion is not affected by emotion for any of the groups.

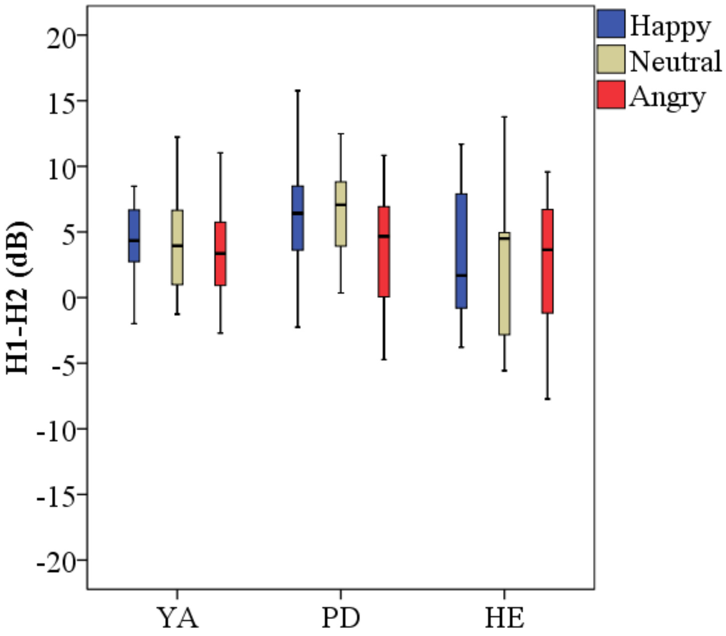

Figure 66: *Effects of H1-H2 on emotion expression by PD, HE and YA groups*

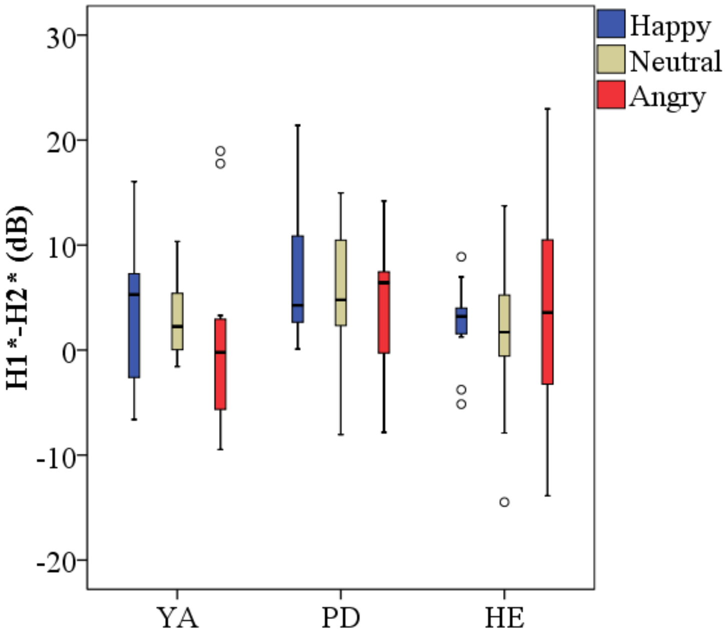

Figure 67: *Effects of H1*-H2* on emotion expression by PD, HE and YA groups*

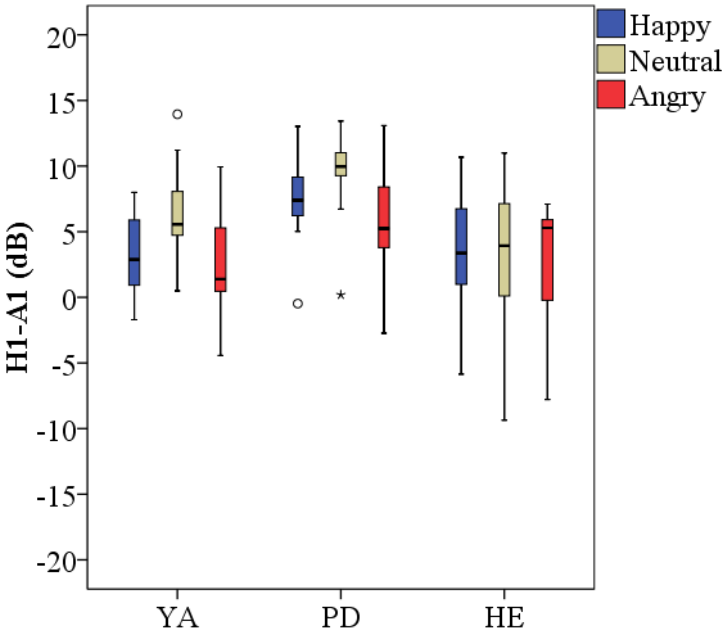

Figure 68: *Effects of H1-A1 on emotion expression by PD, HE and YA groups*

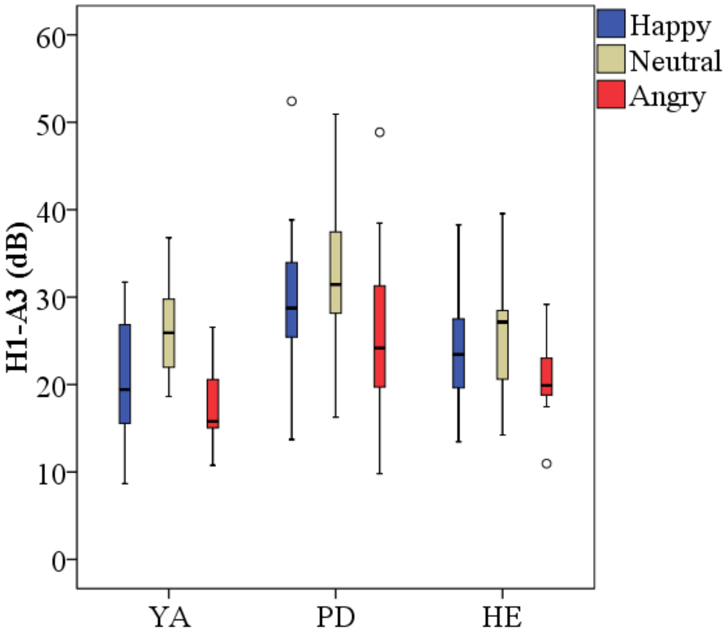

Figure 69: *Effects of H1-A3 on emotion expression by PD, HE and YA groups*

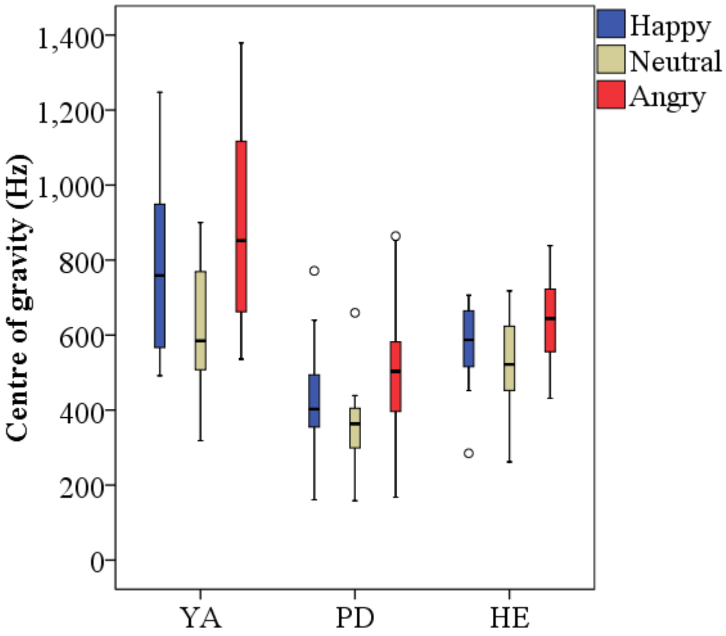

Figure 70: *Effect of centre of gravity (Hz) on emotion expression by PD, HE and YA groups*

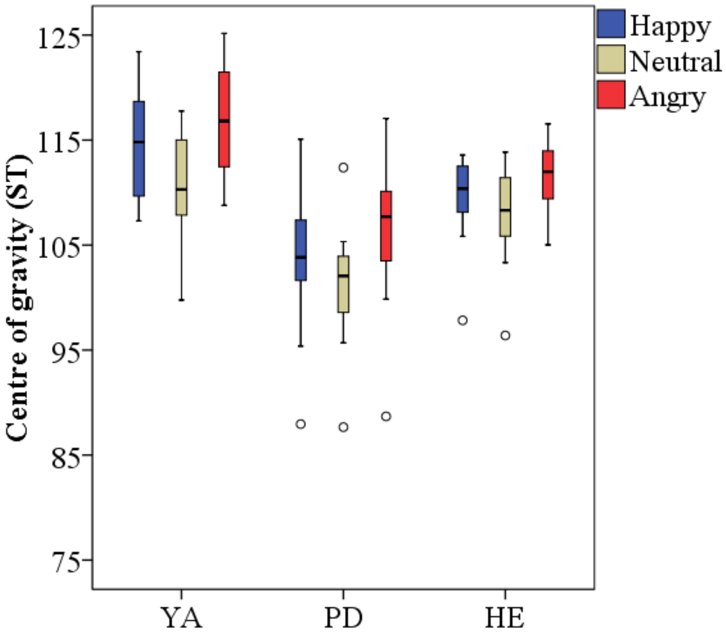

Figure 71: *Effect of centre of gravity (ST) on emotion expression by PD, HE and YA groups*

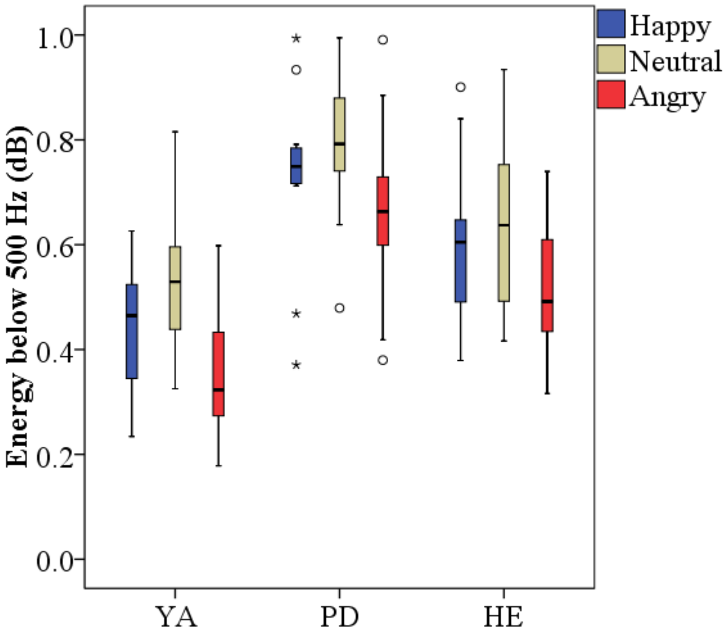

Figure 72: *Effect of energy below 500 Hz on emotion expression by PD, HE and YA groups*

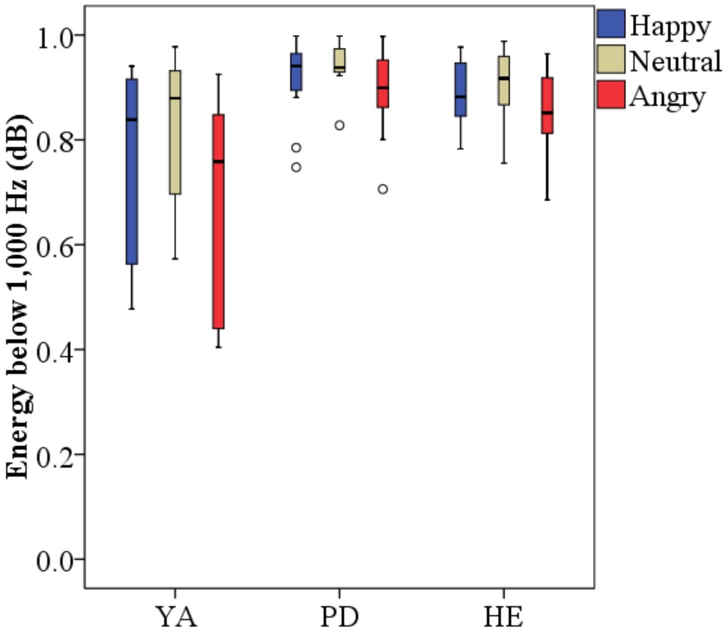

Figure 73: *Effect of energy below 1,000 Hz on emotion expression by PD, HE and YA groups*

Figure 66 shows the relationship between H1-H2 and expression of emotion. The PD and YA groups produced anger with lower mean H1-H2 than happiness, yet the HE group performed oppositely. As lower H1-H2 indicate more pressed voice quality, which is associated with larger speakers, thus only the PD and YA groups performed anger sounding as larger speakers, and happiness sounding as smaller speakers. ANOVA results show that only the PD group has a significant effect on H1-H2 (F(2, 20) = 4.48, *p* = 0.025).

Figure 67 illustrates the relationship between H1*-H2* and expression of emotion. ANOVA results indicates that none of the three groups has any significant effect on H1*-H2*.

Figure 68 shows the relationship between H1-A1 and expression of emotion. The PD and YA groups expressed anger with lower mean H1-A1than happiness, but the HE group produced the opposite trend. As lower H1-A1 indicates a more pressed voice quality, which is associated with larger speakers, thus only the PD and YA groups performed angry-sounding as larger speakers, and happy-sounding as smaller speakers. ANOVA results show that only the PD (F(2, 20) = 8.69, *p* = 0.002) and YA (F(2, 20) = 7.62, *p* = 0.003) groups have a significant effect on H1-A1.

Figure 69 shows the relationship between H1-A3 and expression of emotion. All the three groups produced anger with lower mean H1-A3 value than happiness. As lower H1-A3 values indicate more pressed voice quality, which is associated with larger body-size projection, thus all the participants performed anger sounding as larger speakers, and happiness sounding as smaller speakers. ANOVA results indicate all the three groups have a significant effect on H1-A3: the PD group (F(2, 20) = 9.58, *p* = 0.001), the HE group (F(2, 20) = 9.55, *p* = 0.001), and the YA group (F(2, 20) = 13.40, $p < 0.001$).

Figure 70 shows the relationship between CG (Hz) and expression of emotion. All the three groups expressed anger with larger CG (Hz) than happiness. As larger CG indicates more pressed voice quality, which is linked to larger body-size projection, thus all the participants performed anger sounding as a large person, and happiness sounding as a small person. ANOVA results show that all the three groups have a significant effect on CG (Hz): the PD group (F(2, 20) = 8.40, *p* = 0.002), the HE group (F(2, 20) = 29.86, *p* < 0.001), and the YA group (F(2, 20) = 39.35, *p* < 0.001).

Figure 71 shows the relationship between CG (ST) and expression of emotion. Similar to CG (Hz), all the three groups expressed anger with larger CG (ST) than happiness. ANOVA results show that all the three groups have a significant effect on CG (ST): the PD group (F(2, 20) = 10.04, *p* = 0.001), the HE group (F(2, 20) = 10.09, *p* = 0.001), and the YA group (F(2, 20) = 25.04, *p* < 0.001).

Figure 72 describes the relationship between energy below 500 Hz and expression of emotion. All the three groups produced anger with lower mean energy below 500 Hz value than happiness. Since lower energy below 500 Hz values indicate more pressed voice quality, which is associated with large speakers, all the participants conveyed anger sounding as large speakers, and happiness sounding as small speakers. ANOVA results demonstrate that all the three groups have a significant effect on energy below 500 Hz: the PD group (F(2, 20) = 6.80, *p* = 0.006), the HE group (F(2, 20) = 12.23, *p* < 0.001), and the YA group (F(2, 20) = 19.97, *p* < 0.001).

Figure 73 depicts the relationship between energy below 1,000 Hz and expression of emotion. All the three groups expressed anger with lower mean energy below 1,000 Hz than happiness. As lower energy below 1,000 Hz indicate more pressed voice quality, which is associated with larger body-size projection, therefore, all the participants performed anger sounding as a large person, and happiness sounding as a small person. ANOVA results confirm that all the three groups have significant effects on energy below 1,000 Hz: the PD group (F(2, 20) = 4.37, *p* = 0.027), the HE group (F(2, 20) = 5.30, *p* = 0.014), and the YA group (F(2, 20) = 16.36, *p* < 0.001).

Within acoustic parameters described above, the main group effects are found to be significant in H1-A3 (F(2, 30) = 4.13, *p* = 0.026), H1-A1 (F(2, 30) = 3.59, *p* = 0.040), CG (Hz) (F(2, 20) = 8.40, *p* = 0.002), CG (ST) (F(2, 30) = 9.44, *p* < 0.001), energy below 500Hz (F(2, 30) = 10.93, *p* < 0.001), and energy below 1,000 Hz (F(2, 30) = 6.11, *p* = 0.006), suggesting the three groups produced these acoustic parameters similarly. The main effect of emotions are found in mean pitch (Hz) (F(2, 60) = 62.82, *p* < 0.001), mean pitch (ST) (F(2, 60) = 63.91, *p* < 0.001), intensity (F(2, 60) = 100.14, *p* < 0.001), mean syllable duration (F(2, 60) = 18.79, *p* < 0.001), pitch range (F(2, 60) = 29.18, *p* < 0.001), excursion size (F(2, 60) = 16.78, *p* < 0.001), H1-A1 (F(2, 60) = 13.61, *p* < 0.001), H1-H3 (F(2, 60) = 30.85, *p* < 0.001), CG (Hz) (F(2, 60) = 35.15, *p* < 0.001), CG (ST) (F(2, 60) = 39.37, *p* < 0.001), signal energy below 500 Hz (F(2, 60) = 35.41, *p* < 0.001), and signal energy below 1,000 Hz (F(2, 60) = 25.26, *p* < 0.001). These results imply that the participants were producing these acoustic parameters differently for expressing each emotion – they conveyed anger with lower mean pitch, higher intensity, narrower pitch range and excursion size, decreased H1-A1 and H1-A3, and lower energy below 500 Hz and 1,000 Hz than happiness. Significant interactions between group and emotion are found in mean pitch (Hz) (F(4, 60) = 3.35, *p* = 0.015), mean pitch (ST) (F(4, 60) = 2.72, *p* = 0.038), mean syllable duration (F(4, 60) = 2.52, *p* = 0.050), H1-A1 (F(4, 60) = 2.62, *p* = 0.044), and the signal energy below 1,000 Hz (F(4, 60) = 3.68, *p* = 0.010), thus the three groups performed these three acoustic parameters significantly different across the three emotions. The results of emotion expression by the PD, HE and YA groups are summarised in the Table 17.

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
Mean pitch (Hz)	happy	234.17	59.68	185.07	53.73	198.52	38.69
	neutral	178.84	52.26	155.87	37.96	165.38	33.99
	angry	208.52	56.20	176.14	55.48	196.98	34.19
	F ratio	F(2, 20) = 39.35		F(2, 20) = 8.40		F(2, 20) = 29.86	
	p-value	$p < 0.001$		$p = 0.002$		$p < 0.001$	
Mean pitch (ST)	happy	93.84	5.08	89.77	4.76	91.30	3.31
	neutral	89.03	5.57	86.97	4.00	88.11	3.49
	angry	91.75	5.32	88.77	5.29	91.23	2.92
	F ratio	F(2, 20) = 34.13		F(2, 20) = 10.27		F(2, 20) = 28.44	
	p-value	$p < 0.001$		$p = 0.001$		$p < 0.001$	
Intensity	happy	79.06	3.82	78.38	3.59	80.43	4.06
	neutral	74.48	4.28	76.81	3.63	78.10	3.87
	angry	81.05	4.54	82.29	3.80	83.71	3.47
	F ratio	F(2, 20) = 55.93		F(2, 20) = 16.62		F(2, 20) = 82.02	
	p-value	$p < 0.001$		$p < 0.001$		$p < 0.001$	

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
Mean syllable duration	happy	205.71	55.11	252.00	53.05	258.05	32.83
	neutral	231.87	64.38	248.73	46.75	271.17	50.66
	angry	207.00	50.64	215.68	32.78	231.22	39.57
	F ratio	F(2, 20) = 10.13		F(2, 20) = 8.28		F(2, 20) = 6.89	
	p-value	$p = 0.001$		$p = 0.002$		$p = 0.005$	
Pitch range	happy	90.17	40.19	80.81	37.01	90.89	31.57
	neutral	57.11	24.49	53.53	20.38	70.13	29.08
	angry	81.86	31.34	67.28	32.01	92.87	25.52
	F ratio	F(2, 20) = 12.94		F(2, 20) = 8.81		F(2, 20) = 9.18	
	p-value	$p < 0.001$		$p = 0.002$		$p = 0.001$	
Excursion size	happy	4.47	0.83	4.39	1.46	4.90	1.21
	neutral	3.51	1.12	3.64	0.89	4.30	0.99
	angry	4.35	0.99	3.81	0.73	4.91	1.16
	F ratio	F(2, 20) = 8.82		F(2, 20) = 4.65		F(2, 20) = 6.35	
	p-value	$p = 0.002$		$p = 0.022$		$p = 0.007$	

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
Formant dispersion 3	happy	833.52	115.08	966.63	125.24	921.95	103.49
	neutral	859.48	122.32	1000.23	102.38	939.70	104.74
	angry	868.31	93.26	991.42	95.52	926.13	84.74
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
Formant dispersion 5	happy	883.81	41.51	954.49	114.39	914.50	59.91
	neutral	888.27	52.86	968.15	109.00	908.60	61.75
	angry	869.58	37.53	945.66	105.87	906.34	52.13
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
H1-H2	happy	4.21	3.49	6.09	4.83	3.44	5.94
	neutral	4.21	4.37	6.46	4.11	2.20	6.08
	angry	3.73	4.45	3.44	5.41	2.37	6.13
	F ratio	*n. s.*		$F(2, 20) = 4.48$		*n. s.*	
	p-value			$p = 0.025$			

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
H1*-H2*	happy	3.75	8.97	7.37	8.44	2.51	8.20
	neutral	3.03	7.03	5.49	7.68	1.51	10.15
	angry	1.03	10.38	3.47	8.40	4.01	12.90
	F ratio p-value	*n. s.*		*n. s.*		*n. s.*	
H1-A1	happy	3.11	3.52	7.40	3.56	3.42	5.01
	neutral	6.22	4.18	9.37	3.98	3.22	6.04
	angry	2.58	4.48	5.53	5.25	2.55	5.26
	F ratio	F(2, 20) = 7.62		F(2, 20) = 8.69		*n. s.*	
	p-value	*p* = 0.003		*p* = 0.002			
H1-A3	happy	20.69	7.03	30.77	9.92	24.32	7.62
	neutral	26.23	5.71	33.13	9.24	25.78	7.45
	angry	17.74	4.85	26.13	11.35	20.96	5.31
	F ratio	F(2, 20) = 13.40		F(2, 20) = 9.58		F(2, 20) = 9.55	
	p-value	*p* < 0.001		*p* = 0.001		*p* = 0.001	

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
Centre of gravity (Hz)	happy	773.79	256.26	430.07	172.75	567.96	130.74
	neutral	627.58	192.68	362.86	128.52	522.92	138.31
	angry	884.57	288.05	513.12	213.07	646.46	118.00
	F ratio	F(2, 20) = 39.35		F(2, 20) = 8.40		F(2, 20) = 29.86	
	p-value	$p < 0.001$		$p = 0.002$		$p < 0.001$	
Centre of gravity (ST)	happy	114.29	5.73	103.61	7.51	109.24	4.69
	neutral	110.72	5.58	101.04	6.31	107.72	5.13
	angry	116.63	5.63	106.49	8.08	111.74	3.30
	F ratio	F(2, 20) = 25.04		F(2, 20) = 10.04		F(2, 20) = 10.09	
	p-value	$p < 0.001$		$p = 0.001$		$p = 0.001$	
Energy below 500 Hz	happy	0.43	0.13	0.73	0.18	0.60	0.17
	neutral	0.54	0.15	0.79	0.14	0.64	0.18
	angry	0.36	0.14	0.67	0.18	0.52	0.13
	F ratio	F(2, 20) = 19.97		F(2, 20) = 6.80		F(2, 20) = 12.23	
	p-value	$p < 0.001$		$p = 0.006$		$p < 0.001$	

Table 17. Means, standard deviations, F ratios and p-values of acoustic parameters in emotion expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	emotion	Mean	SD	Mean	SD	Mean	SD
Energy below 1,000 Hz	happy	0.75	0.19	0.91	0.08	0.89	0.07
	neutral	0.82	0.15	0.94	0.05	0.90	0.07
	angry	0.68	0.22	0.89	0.09	0.85	0.08
	F ratio	F(2, 20) = 16.36		F(2, 20) = 4.37		F(2, 20) = 5.30	
	p-value	$p < 0.001$		$p = 0.027$		$p = 0.014$	

4.4.2 Expression of attitude by PD, HE and YA groups

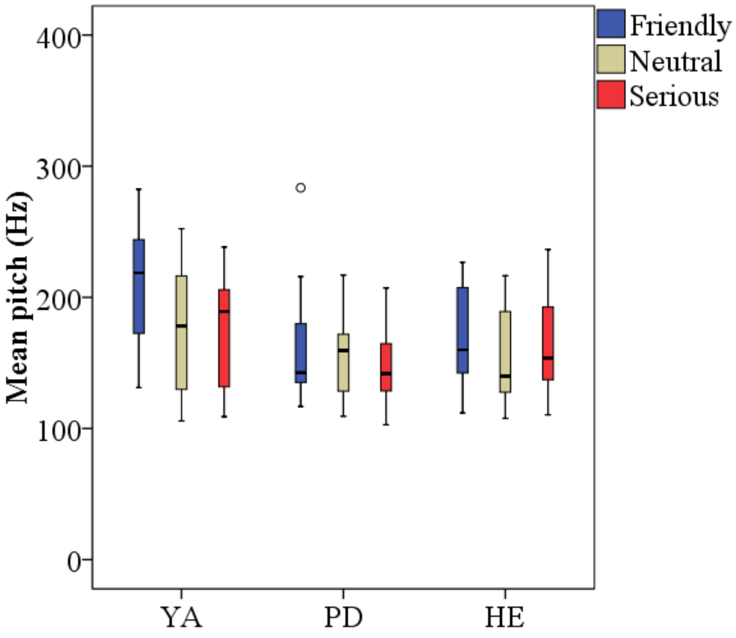

Figure 74: *Effects of mean pitch (Hz) on attitude expression by PD, HE and YA groups*

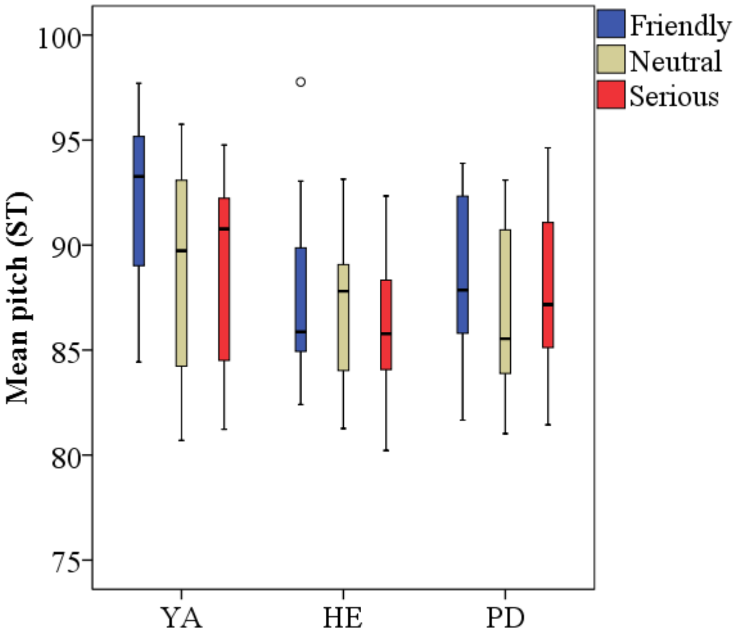

Figure 75: *Effects of mean pitch (ST) on attitude expression by PD, HE and YA groups*

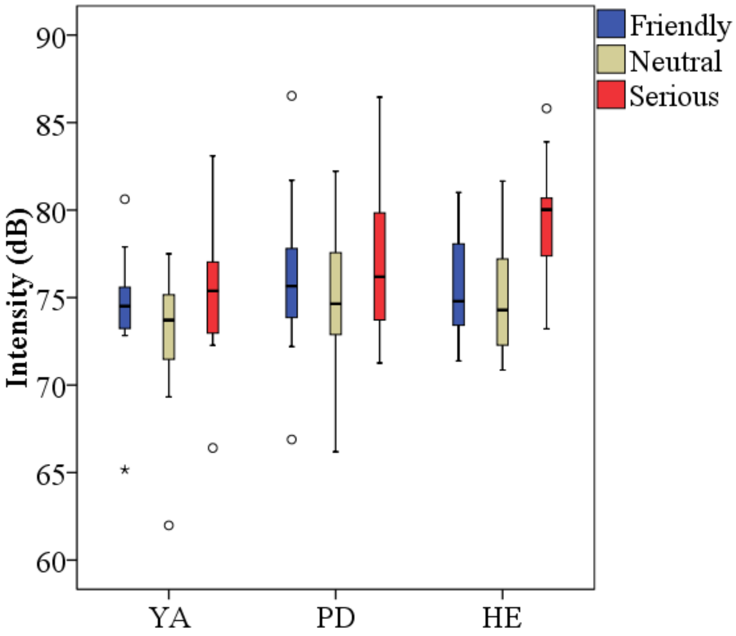

Figure 76: *Effect of intensity on attitude expression by PD, HE and YA groups*

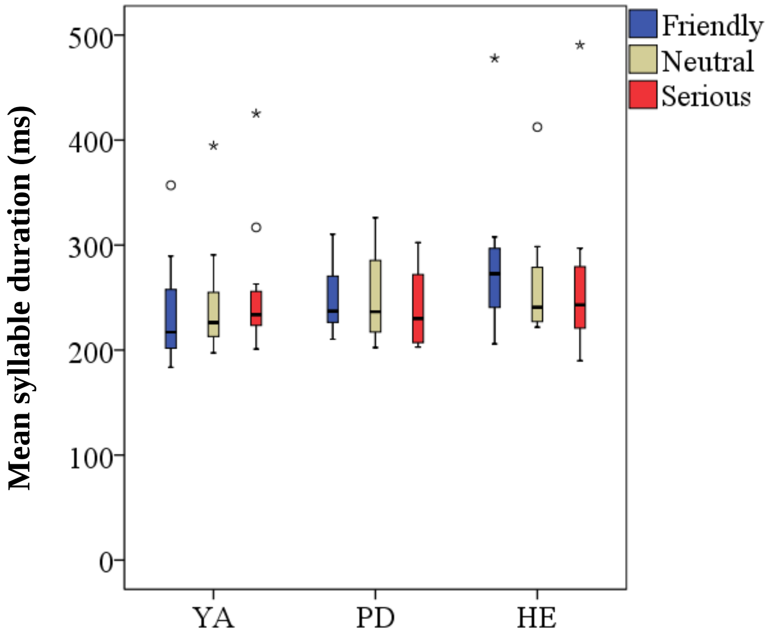

Figure 77: *Effect of mean syllable duration on attitude expression by PD, HE and YA groups*

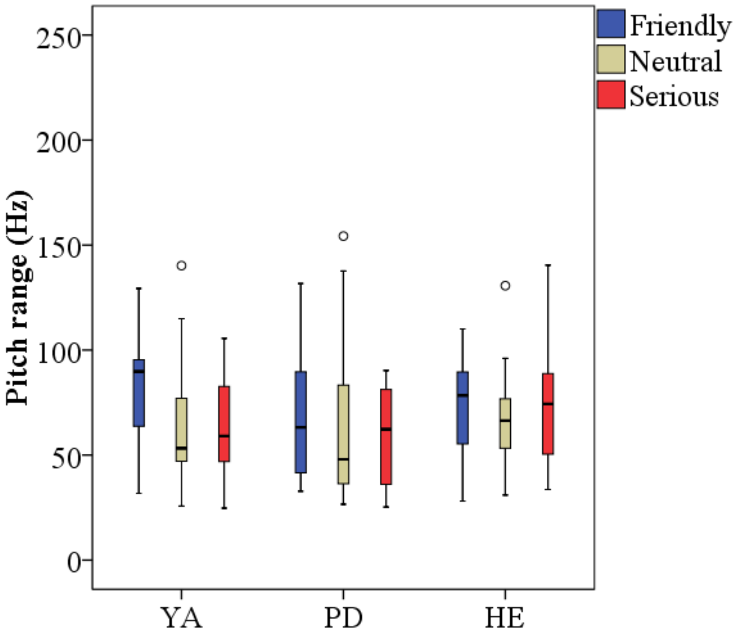

Figure 78: *Effect of pitch range on attitude expression by PD, HE and YA groups*

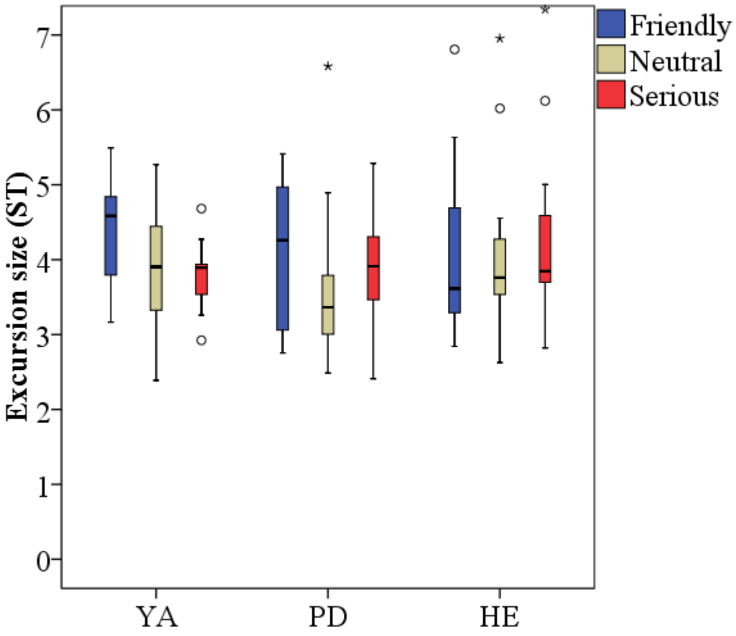

Figure 79: *Effect of excursion size on attitude expression by PD, HE and YA groups*

Among the traditional acoustic parameters, attitude has no significant main effects on any acoustic parameters for the PD group.

Figure 74 describes the relationship between mean pitch (Hz) and expression of attitude. The HE and YA groups performed serious attitude with lower mean pitch (Hz) than friendliness, while the PD group did the opposite. Lower mean pitch is associated with a larger body-size projection, thus the HE and YA groups expressed serious-sounding as larger speakers, and friendly-sounding as smaller speakers. ANOVA results confirm that the HE and the YA groups have a significant effect on mean pitch (Hz): the HE group (F(2, 20) = 11.73, $p < 0.001$), and the YA group (F(2, 20) = 46.47, $p < 0.001$).

Figure 75 shows the relationship between mean pitch (ST) and expression of attitude. Similar to mean pitch (Hz), the HE and YA groups conveyd seriousness with lower mean pitch (ST) than friendliness, but the PD group did the opposite. ANOVA results indicates that the HE and the YA groups have a significant effect on mean pitch (ST): the HE group (F(2, 20) = 17.18, $p < 0.001$), and the YA group (F(2, 20) = 40.84, $p < 0.001$).

Figure 76 demonstrates the relationship between intensity and expression of attitude. The HE and YA groups produced seriousness with higher mean intensity than friendliness, whereas the PD group conveyed both attitudes with similar mean intensity. Similar to the results in the emotion judgement, where anger is more audible than happiness, in the attitude task, seriousness is more audible than friendliness. ANOVA results demonstrate that the HE and YA groups have significant effects on intensity: the HE group (F(2, 20) = 54.60, $p < 0.001$) and the YA group (F(2, 20) = 3.46, $p = 0.050$).

Figure 77 shows the relationship between mean syllable duration and expression of attitude. ANOVA results reveal that mean syllable duration is influenced by attitude only for the YA group (F(2, 20) = 4.12, $p = 0.032$).

Figure 78 illustrates the relationship between pitch range and expression of attitude. Only the YA and the HE groups expressed seriousness with smaller pitch range than friendliness. Although body-size projection does not provide particular predictions for pitch range interpretations, past empirical results in emotion suggested that synthetic speech with wider pitch range was often perceived as from happier speakers (Murray & Arnott, 1993; Nwe et al., 2003). Yet, ANOVA results indicate that pitch range is influenced by attitude only for the YA group (F(2, 20) = 4.16, $p = 0.031$).

Figure 79 depicts the relationship between excursion size and expression of attitude. Similar to pitch range, only the YA and the HE groups performed seriousness with smaller pitch range than friendliness. However, ANOVA results reveal that excursion size is influenced by attitude only for the YA group (F(2, 20) = 3.97, $p = 0.035$).

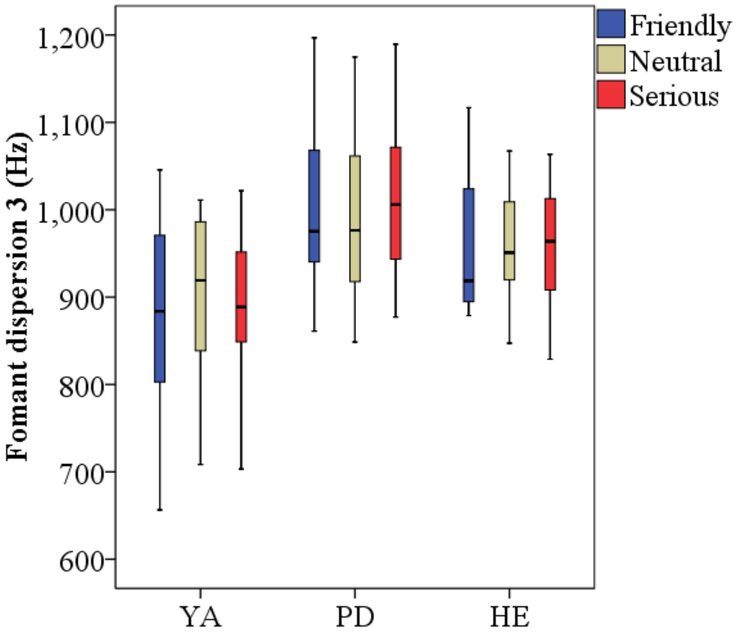

Figure 80: *Effects of formant dispersion 3 on attitude expression by PD, HE and YA groups*

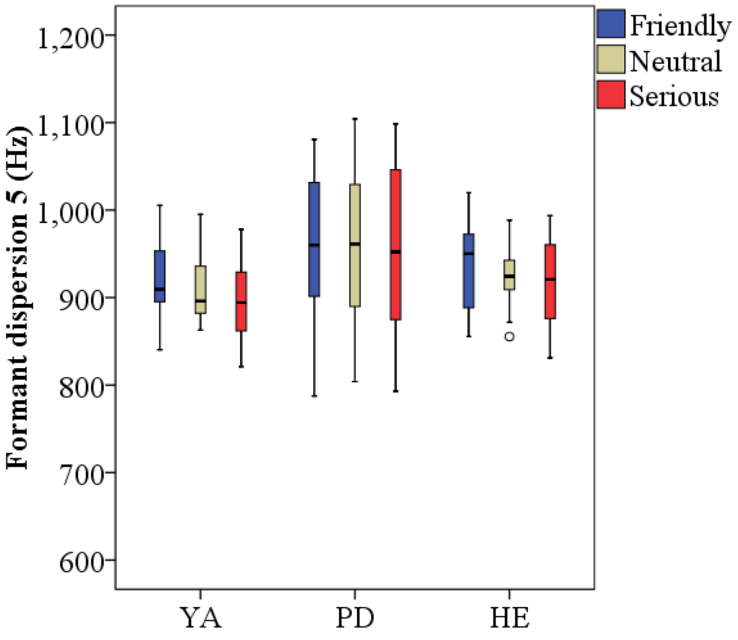

Figure 81: *Effects of formant dispersion 5 on attitude expression by PD, HE and YA groups*

Among the two categories BID-associated parameters, attitude has no significant effect on any VTL-related parameters in any of the three groups. Figures 80 and 81 illustrate the relationship between formant dispersion 3, formant dispersion 5 and expression of attitude. ANOVA results confirmed that formant dispersion is not affected by attitude for any of the groups.

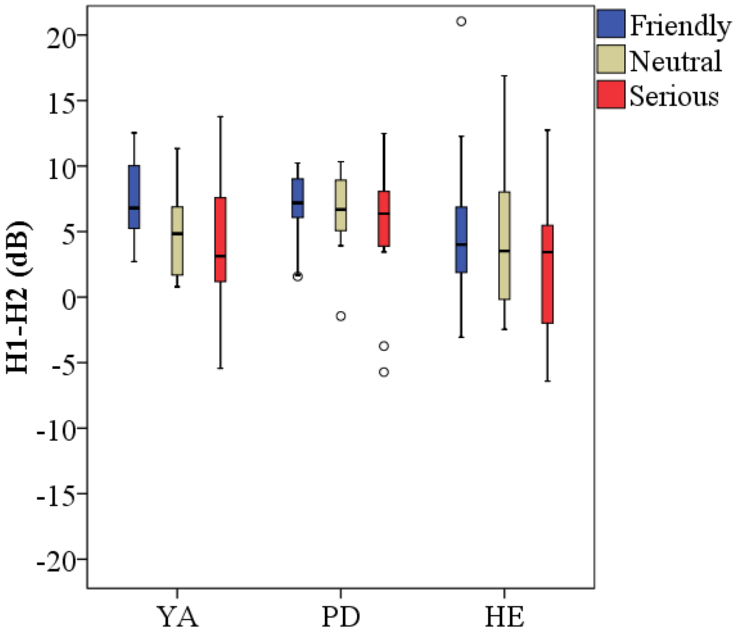

Figure 82: *Effects of H1-H2 on attitude expression by PD, HE and YA groups*

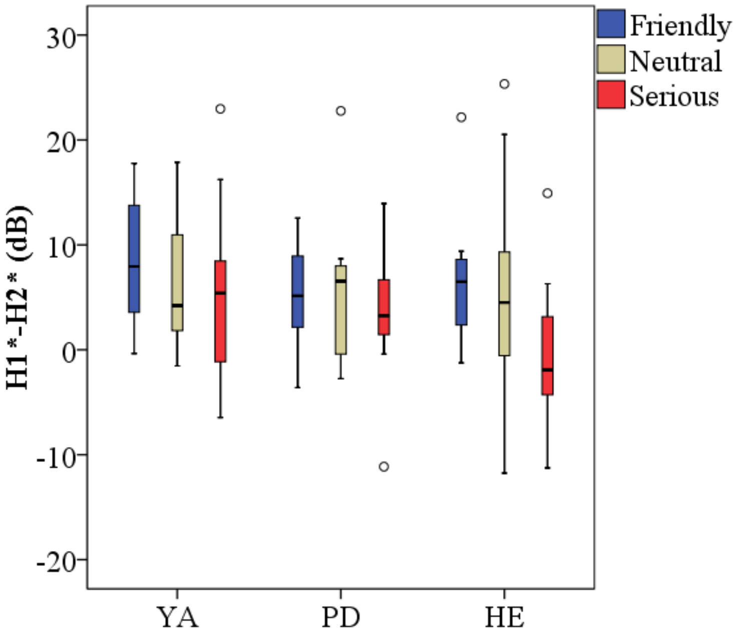

Figure 83: *Effects of H1*-H2* on attitude expression by PD, HE and YA groups*

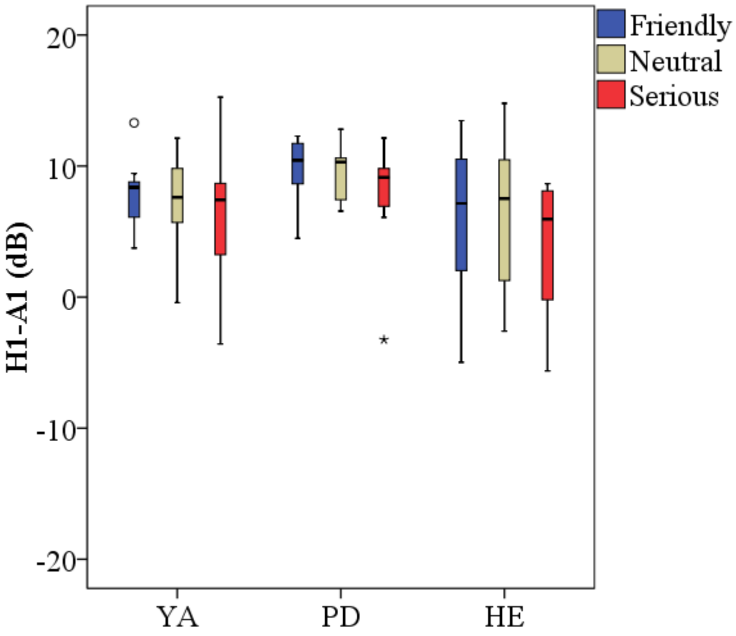

Figure 84: *Effects of H1-A1 on attitude expression by PD, HE and YA groups*

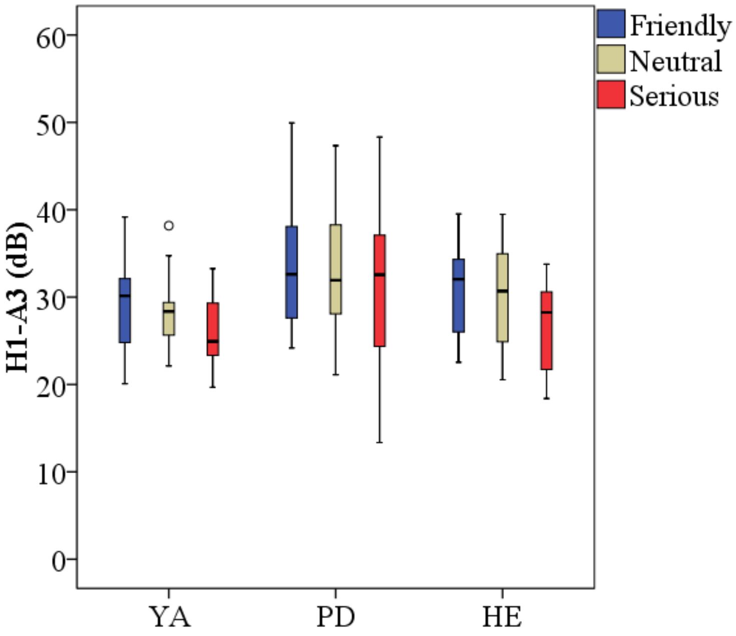

Figure 85: *Effects of H1-A3 on attitude expression by PD, HE and YA groups*

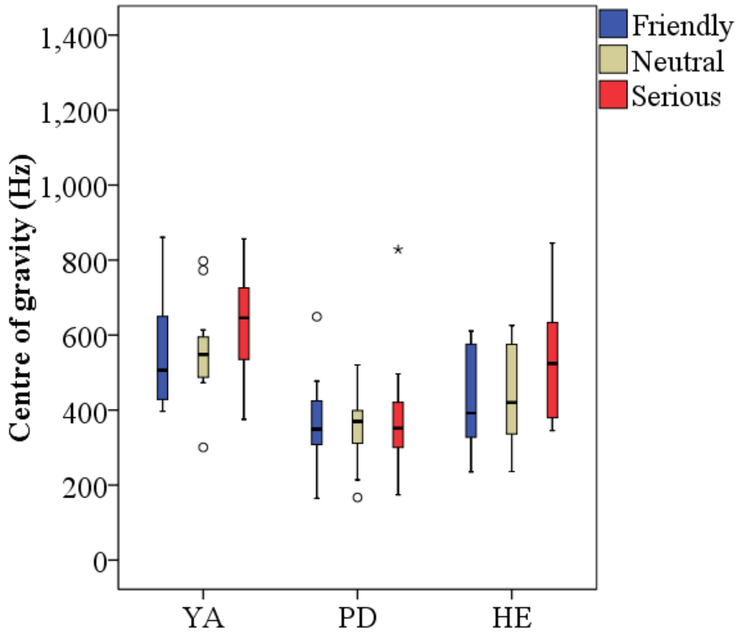

Figure 86: *Effect of centre of gravity (Hz) on attitude expression by PD, HE and YA groups*

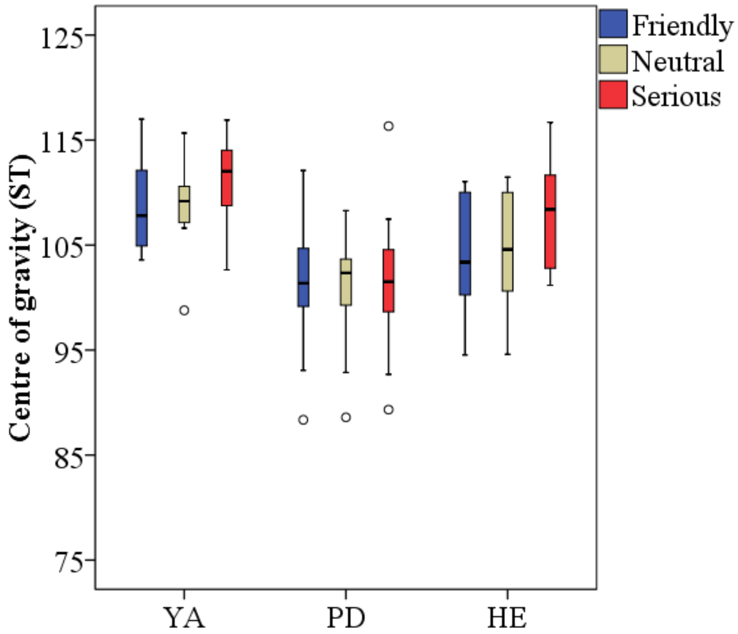

Figure 87: *Effect of centre of gravity (ST) on attitude expression by PD, HE and YA groups*

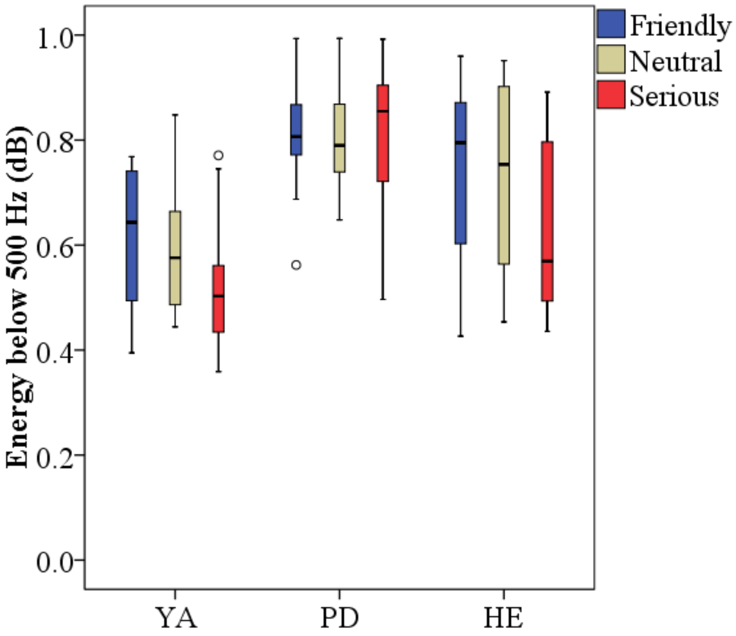

Figure 88: *Effect of energy below 500 Hz on attitude expression by PD, HE and YA groups*

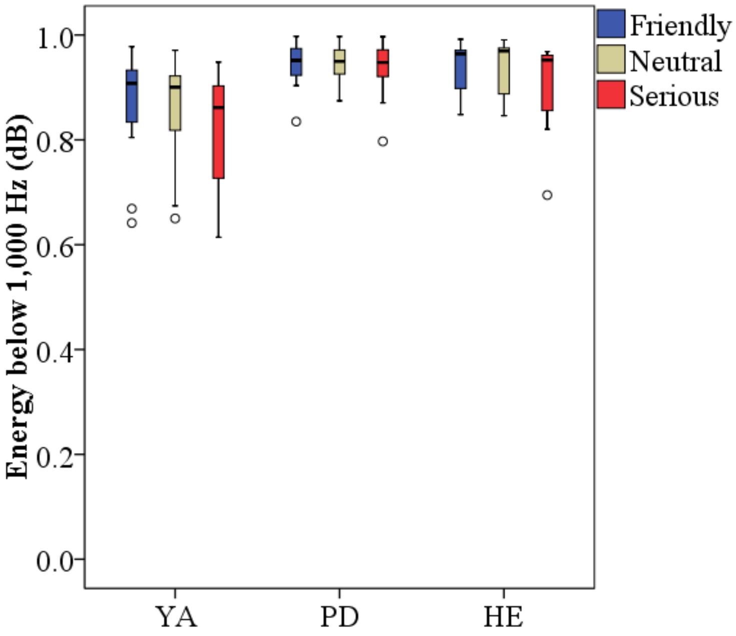

Figure 89: *Effect of energy below 1,000 Hz on attitude expression by PD, HE and YA groups*

Figure 82 depicts the relationship between H1-H2 and expression of attitude. All the three groups expressed seriousness with lower H1-H2 than friendliness. Lower H1-H2 indicates more pressed voice quality, which is associated with larger speakers. Thus all the participants conveyed seriousness as larger speakers, and friendliness as smaller speakers. ANOVA results show that the HE (F(2, 18) = 4.16, *p* = 0.033) and YA (F(2, 20) = 4.93, *p* = 0.018) groups have significant effects on H1-H2.

Figure 83, 84 and 85 depict the relationship between H1*-H2*, H1-A1, H1-A3 and expression of attitude. Similar to H1-H2, all the three groups produced seriousness with lower H1*-H2*, H1-A1 and H1-A3 than friendliness. As lower H1*-H2*, H1-A1 and H1-A3 indicate a more pressed voice quality, which is associated with larger speakers, thus all the participants conveyed serious-sounding as larger speakers, and friendly-sounding as smaller speakers. Yet, ANOVA results demonstrates that only the HE group has a significant effect on H1*-H2* (F(2, 20) = 4.18, *p* = 0.030), H1-A1 (F(2, 20) = 9.50, *p* = 0.001) and H1-A3 (F(2, 20) = 10.23, *p* = 0.001) .

Figure 86 shows the relationship between CG (Hz) and expression of attitude. The HE and YA groups conveyed seriousness with higher CG (Hz) than friendliness, yet the PD group expressed different attitudes without distinction. As higher CG indicate more pressed voice quality, which implies larger speakers, the HE and the YA groups produced seriousness as larger speakers, and friendliness as smaller ones. ANOVA results confirm that both the HE (F(2, 20) = 8.61, *p* = 0.002) and YA (F(2, 20) = 4.93, *p* = 0.018) groups have a significant effect on CG (Hz).

Figure 87 demonstrates the relationship between CG (ST) and expression of attitude. Similar to CG (Hz), only the HE and YA groups conveyed seriousness with higher CG (ST) than friendliness. ANOVA results reveal that both the HE and YA groups have a significant effect on CG (ST): the HE group (F(2, 20) = 10.84, *p* = 0.001) and the YA group (F(2, 20) = 4.55, *p* = 0.023).

Figure 88 describes the relationship between energy below 500 Hz and expression of attitude. The HE and YA group produced seriousness with lower energy below 500 Hz than friendliness, while the PD group performed oppositely. Lower energy below 500 Hz demonstrate more pressed voice quality, which is linked to larger speakers, therefore, the HE and YA groups expressed seriousness with larger body-size projection, and friendliness with smaller body-size projection. ANOVA results indicate that both the YA and HE groups have a significant effect on energy below 500 Hz: the HE group (F(2, 20) = 16.63, $p < 0.001$) and the YA group (F(2, 20) = 5.55, $p = 0.012$).

Figure 89 demonstrates the relationship between energy below 1,000 Hz and expression of attitude. Similar to energy below 500 Hz, the HE and YA groups expressed seriousness with lower energy below 1,000 Hz, whereas the PD group performed attitudes similarly. ANOVA results show that the HE and YA groups have a significant effect on energy below 1,000 Hz: the HE group (F(2, 20) = 4.97, $p = 0.018$) and the YA group (F(2, 20) = 5.01, $p = 0.017$).

Within acoustic parameters described above, main group effects were significant for CG (Hz) (F(2, 30) = 3.62, $p = 0.039$), CG (ST) (F(2, 30) = 7.08, $p = 0.003$), energy below 500 Hz (F(2, 30) = 7.80, $p = 0.002$), and energy below 1,000 Hz (F(2, 30) = 4.76, $p = 0.016$), suggesting the three groups performed these acoustic parameters significantly different. The main effect of attitudes are found in mean pitch (Hz) (F(2, 20) = 11.73, $p < 0.001$), mean pitch (ST) (F(2, 60) = 29.98, $p < 0.001$), intensity (F(2, 60) = 26.63, $p < 0.001$), H1-H2 (F(2, 60) = 8.31, $p = 0.001$), H1*-H2* (F(2, 60) = 3.89, $p = 0.026$), H1-A1 (F(2, 60) = 7.50, $p = 0.001$), H1-A3 (F(2, 60) = 8.66, $p < 0.001$), CG (Hz) (F(2, 60) = 7.14, $p = 0.002$), CG (ST) (F(2, 60) = 6.98, $p = 0.002$), signal energy below 500 Hz (F(2, 60) = 9.39, $p < 0.001$), and signal energy below 1,000 Hz (F(2, 60) = 7.33, $p = 0.001$). These results infer that the participants were producing these acoustic parameters differently for expressing each attitude – they expressed seriousness with lower mean pitch, higher intensity, narrower pitch range and excursion size, decreased H1-H2, H*-H2*, H1-A1 and H1-A3, higher CG and lower energy below 500

Hz than friendliness. Significant interactions between group and attitude are found in mean pitch (Hz) (F(4, 60) = 5.16, *p* = 0.001), mean pitch (ST) (F(4, 60) = 5.73, *p* = 0.001), and mean syllable duration (F(4, 60) = 3.32, *p* = 0.016), thus the three groups performed these acoustic parameters significantly differently across the three attitudes. The results of attitude expression by the PD, HE and YA groups are summarised in the Table 18.

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Mean pitch (Hz)	friendly	209.53	50.84	165.58	49.27	170.57	40.21
	neutral	175.27	50.97	153.96	35.65	155.73	38.94
	serious	171.33	46.28	145.98	32.30	165.80	42.21
	F ratio	F(2, 20) = 46.47		*n. s.*		F(2, 20) = 11.73	
	p-value	*p* < 0.001				*p* < 0.001	
Mean pitch (ST)	friendly	92.02	4.53	87.85	4.61	88.52	4.13
	neutral	88.73	5.30	86.78	3.96	86.91	4.27
	serious	88.43	4.92	85.89	3.82	87.97	4.38
	F ratio	F(2, 20) = 40.84		*n. s.*		F(2, 20) = 17.18	
	p-value	*p* < 0.001				*p* < 0.001	

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Intensity	friendly	74.30	3.90	76.14	5.22	75.58	3.29
	neutral	72.53	4.24	75.32	4.70	75.13	3.50
	serious	75.14	4.24	77.18	5.05	79.46	3.56
	F ratio	(F(2, 20) = 3.46		*n. s.*		F(2, 20) = 54.60	
	p-value	*p* = 0.050				*p* < 0.001	
Mean syllable duration	friendly	245.62	56.50	252.75	48.03	264.12	56.98
	neutral	255.64	64.60	242.07	42.14	263.43	83.14
	serious	235.31	53.81	249.58	37.71	282.73	74.58
	F ratio	F(2, 20) = 4.12		*n. s.*		*n. s.*	
	p-value	*p* = 0.032					
Pitch range	friendly	79.20	29.22	68.80	35.51	73.46	29.43
	neutral	65.39	36.96	68.48	50.96	68.75	31.18
	serious	63.84	28.08	59.09	25.45	75.09	33.01
	F ratio	F(2, 20) = 4.16		*n. s.*		*n. s.*	
	p-value	*p* = 0.031					

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Excursion size	friendly	4.38	0.86	4.10	1.25	4.14	1.33
	neutral	3.85	0.96	3.68	1.21	4.17	1.34
	serious	3.77	0.69	3.86	0.94	4.36	1.47
	F ratio	F(2, 20) = 3.97		*n. s.*		*n. s.*	
	p-value	*p* = 0.035					
Formant dispersion 3	friendly	873.68	129.31	1004.78	108.59	956.37	87.48
	neutral	902.24	101.00	990.79	105.29	957.70	79.00
	serious	895.49	98.56	1009.98	100.71	955.92	84.45
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						
Formant dispersion 5	friendly	922.23	51.79	957.18	98.84	935.19	58.73
	neutral	911.76	45.25	951.64	102.01	923.84	41.81
	serious	895.90	52.47	951.46	107.10	918.97	61.21
	F ratio	*n. s.*		*n. s.*		*n. s.*	
	p-value						

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
H1-H2	friendly	7.50	3.75	6.83	3.29	5.42	6.95
	neutral	4.73	3.50	6.42	3.77	4.61	6.29
	serious	4.18	5.35	4.97	5.57	2.37	6.28
	F ratio	F(2, 20) = 4.93		*n. s.*		F(2, 18) = 4.16	
	p-value	*p* = 0.018				*p* = 0.033	
H1*-H2*	friendly	8.28	8.41	5.51	6.65	6.48	8.75
	neutral	6.76	9.28	5.67	8.31	5.32	11.69
	serious	5.17	10.68	3.58	9.16	-0.55	9.62
	F ratio	*n. s.*		F(2, 20) = 4.18		*n. s.*	
	p-value			*p* = 0.030			
H1-A1	friendly	7.65	3.02	9.62	3.22	6.05	5.73
	neutral	7.45	3.84	9.46	2.37	6.00	6.26
	serious	6.18	5.10	7.87	4.31	3.58	5.89
	F ratio	*n. s.*		F(2, 20) = 9.50		*n. s.*	
	p-value			*p* = 0.001			

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
H1-A3	friendly	29.05	6.52	33.82	8.55	30.77	6.34
	neutral	28.33	4.96	33.30	8.35	30.27	6.42
	serious	26.15	4.53	31.42	10.45	26.16	6.15
	F ratio	*n. s.*		F(2, 20) = 10.23		*n. s.*	
	p-value			*p* = 0.001			
Centre of gravity (Hz)	friendly	553.50	165.67	365.96	132.72	441.01	141.99
	neutral	559.57	142.51	356.27	109.61	443.16	141.83
	serious	633.49	157.61	382.74	176.66	529.49	167.39
	F ratio	F(2, 20) = 4.93		*n. s.*		F(2, 20) = 8.61	
	p-value	*p* = 0.018				*p* = 0.002	
Centre of gravity (ST)	friendly	108.69	4.89	101.11	6.55	104.54	5.82
	neutral	108.99	4.65	100.85	6.00	104.59	6.01
	serious	111.17	4.50	101.50	7.33	107.78	5.45
	F ratio	F(2, 20) = 4.55		*n. s.*		F(2, 20) = 10.84	
	p-value	*p* = 0.023				*p* = 0.001	

Table 18. Means, standard deviations, F ratios and p-values of acoustic parameters in attitude expression by PD, HE and YA groups

		YA (*n=11*)		PD (*n=11*)		HE (*n=11*)	
	attitude	Mean	SD	Mean	SD	Mean	SD
Energy below 500 Hz	friendly	0.62	0.15	0.80	0.12	0.73	0.18
	neutral	0.60	0.13	0.81	0.10	0.73	0.18
	serious	0.52	0.14	0.81	0.14	0.63	0.17
	F ratio	F(2, 20) = 5.55		*n. s.*		F(2, 20) = 16.63	
	p-value	$p = 0.012$				$p < 0.001$	
Energy below 1,000 Hz	friendly	0.86	0.11	0.94	0.05	0.93	0.05
	neutral	0.86	0.11	0.94	0.04	0.93	0.06
	serious	0.82	0.13	0.94	0.06	0.90	0.09
	F ratio	F(2, 20) = 5.01		*n. s.*		F(2, 20) = 4.97	
	p-value	$p = 0.017$				$p = 0.018$	

Chapter 5 Discussion

The present study aimed to explore perception and production of emotional and attitudinal prosody of adolescents with ASD and PD patients, by comparing their performance to the performance of typically developing adolescents, healthy elderly and young adults. All the participants judged synthetic utterances according to body size (small, medium or large), emotions (happy, neutral, and angry) and attitudes (friendly, neutral and serious), and produced utterances in TM with different emotions and attitude. Acoustic parameters in synthetic utterances were manipulated following the principles of BID theory. Two types of acoustic parameters – traditional and BID-associated, were extracted from participants' speech samples for statistical analysis.

5.1 Perception of emotional and attitudinal prosody in ASD

The results of the present study showed that, overall, adolescents with ASD, typically developing adolescents and young adults made similar perceptual judgments in perceptual tasks concerning body-size projection, emotion and attitude. Yet, adolescents with ASD were less accurate in perceiving body size, emotion and attitude than their typically developing peers and young adults when relying on the changes of acoustic parameters. Moreover, the performance of the same tasks by typically developing adolescents was inferior when compared to young adults. The sensitivity of adolescents with ASD was especially reduced for voice quality manipulations. Thus, to judge emotion through prosody, they seemed to rely mainly on formant shift ratio and pitch shift, and on pitch range to a lesser extent. In speech, fundamental frequency is associated to the perception of vocal pitch, while formant frequency, which reflects the vocal tract resonance properties, is associated to the perception of timber (Pisanski & Rendall, 2011). Past research has shown that some individuals with ASD, especially those with more severe language difficulties, were more sensitive to pitch variation from pure tone than their typical peers. Nevertheless, this strength is not transferred to speech prosody (Jiang, Liu, Wan, & Jiang, 2015; O'Connor, 2012).

The reduced sensitivity of adolescents with ASD to acoustic changes across the three perceptual tasks suggest that their difficulties with the perception of emotional prosody are present even in a context-free situation, such as in a lab testing session. Thus, it is not surprising they have challenges in real life where the social interactions require accurate real-time use of social-cognitive skills, such as display rules, which are culture-dependent unwritten codes that govern the manner in which emotions should be expressed (Ekman & Friesen, 1969; Malatesta & Haviland, 1982). ASD is a pervasive condition affecting all aspects of social-communicative skills in different degrees of severity (Globerson et al., 2015; Heaton et al., 2012; Philip et al., 2010). Some aspects can be more or less disturbed than others within the same skill, while other aspects can be fully intact. Thus the behaviour in ASD may vary drastically from one person to another. This may be the reason why empirical results are so diverse across studies, which adds to the challenges in treating prosodic disorder in clinical settings (see Peppé (2009) for a review).

Adolescents with ASD, typically developing adolescents and young adults were all sensitive to changes in formant shift ratio and pitch shift in the three experiments, but only young adults showed sensitivity to voice quality manipulations. Even their responses were in the opposite direction of previous findings, where pressed voice is associated with a larger and angrier person, and breathy voice with a smaller and happier person (Chuenwattanapranithi et al., 2008; Xu, Kelly, et al., 2013; Xu, Lee, et al., 2013).

In our experiment, young adults perceived breathy voice as from a large, angry or serious person, and pressed as from a small, happy or friendly person. Upon listening to the experimental stimuli again ourselves after the experiments, we noticed that the breathy version sounded rather effortful. It is possible that this effortfullness may be associated with angrier and/or more serious speakers, which is consistent with the basic BID hypothesis. The reason for the effortful effect of breathy voice may lie in the sophisticated nature of VocalTractLab, the articulatory synthesizer used to create the stimuli (Birkholz, Jackèl, & Kröger, 2006). This synthesiser simulates speech production based on aerodynamics. As a result, the breathy version of our base sentence had greater airflow through the vocal tract. And because the base sentence contains both high vowels and glides that involve narrow vocal tract constrictions, the increased airflow due to breathiness through these constrictions generated increased turbulent noise. The turbulent noise may have created the impression of increased roughness of voice. Because rough voice is associated with aggression and anger (Morton, 1977), sensitive listeners like the young adults in this study may have heard the breathy version of the stimuli to be more angry, serious and sounding like a larger person. Another possible explanation is that anger can be expressed in two ways – cold anger and hot anger (Banse & Scherer, 1996; Sinaceur, Van Kleef, Neale, Adam, & Haag, 2011). Cold anger usually is associated with more control over feelings and actions, whereas hot anger is characterised by impulsiveness and aggression (Sinaceur et al., 2011). As aggression is acoustically characterised by rough voice (Morton, 1977), and both breathy quality and rough quality have noise components in their acoustics, some respondents may have interpreted stimuli with breathy voice quality as sounding more aggressive, and thus associated these stimuli with anger and seriousness. But the above speculation is highly tentative, and so this particular issue needs to be further explored in future research.

Attitude has been included along with emotion in the experiment, as past research has found that the two are heavily overlapped (Mitchell & Ross, 2013).

Speaker intention can be inferred from attitudinal prosody, and successful interpretation can facilitate and enhance social interactions (Mitchell & Ross, 2013). In the present study, the perception of adolescents with ASD in attitudinal prosody was very similar to their perception of emotional prosody. In both tasks, they were less sensitive to the acoustic manipulations and used fewer acoustic parameters to judge speaker's attitude when compared to typically developing adolescents and young adults. Since their performance was already poorer in a context-free situation (in the lab), it is predictable that adolescents with ASD will have greater difficulty in real-life social events, where the auditory (prosody) cues, visual (facial expression) cues and linguistic content are delivered all at once, demanding immediate responses.

Interestingly, typically developing adolescents were less sensitive to the acoustic changes than young adults, suggesting that the ability to perceive emotional prosody is still developing in adolescence towards early adulthood, which is in line with the suggestion by Wells, Peppé, & Goulandris (2004). Moreover, if perceptual sensitivity to acoustic changes is related to different stages in prosodic development, then the lower sensitivity of adolescents with ASD in this study could be interpreted as a delay rather than a total absence in their developmental trajectory. Further research on the perception of emotional prosody by children and adults with ASD may provide more insights.

Table 19. Perceptual judgment according to the predictions from the body size projection by AA, TA and YA groups

		AA	TA	YA
Body size	Voice quality	✓(*)	✗(**)	✗
	Formant shift ratio	✓	✓	✓
	Pitch shift	✓	✓	✓
	Pitch range	✗	✗	✗
Emotion	Voice quality	✓	✗	✗
	Formant shift ratio	✓	✓	✓
	Pitch shift	✓	✓	✓
	Pitch range	✓	✓	✓
Attitude	Voice quality	✓	✗	✗
	Formant shift ratio	✓	✓	✓
	Pitch shift	✓	✓	✓
	Pitch range	✓	✓	✓
	Total ✓	11	8	8
	Total ✗	1	4	4

(*) ✓ = as predicted by the body size projection

(**) ✗ = different from the predictions of the body size projection

5.2 Production of emotional and attitudinal prosody in ASD

In the present study, adolescents with ASD, typically developing adolescents and young adults manipulated most of the acoustic parameters to convey emotions and attitudes through prosody. The performance of adolescents with ASD was comparable to typically developing adolescents and young adults in traditional acoustic parameters – mean pitch, intensity and pitch range. The participants of the three groups expressed anger and seriousness with lower mean pitch and narrower pitch range, which are both associated with larger or angry-sounding speakers. In general, anger and seriousness are produced with more intensity, suggesting that these two affects are more audible than happiness and friendliness. Thus, from current results, there is no evidence supporting monotonous/flat speech (as found in Bonneh et al., 2011; Nakai et al., 2014), greater variations in mean pitch and pitch range (as found in Bonneh et al., 2011; Green and Tobin, 2009; Nakai et al., 2014), or imbalanced use of acoustic parameters when conveying prosody (as found in Diehl et al., 2009; Hubbard & Trauner, 2007). Adolescents with ASD in the present study demonstrated their ability to use a wide variety of traditional acoustic parameters when expressing different emotions and attitudes through prosody.

In contrast, some discrepancies in performance were observed in the voice quality-related acoustic parameters. Only the performance of adolescents with ASD in emotional prosody was comparable to young adults, as both groups conveyed anger with more pressed voice quality than happiness. Since pressed voice quality is associated with larger or angry-sounding speakers, the participants of these two groups performed according to the predictions of body-size projection. Thus, the results of the present study suggest that adolescents with ASD are able to perform emotional prosody when requested in a context-free condition. Surprisingly, typically developing adolescents performed emotions poorer than adolescents with ASD. Typical adolescents conveyed anger with breathier voice quality than happiness, which contradicts the hypothesis of body-size projections. During the recording sessions, some adolescents mentioned their embarrassment to perform in front of an unfamiliar adult (the examiner). They also felt challenged to express emotions and attitudes out of context, as some pointed out: “I cannot say this angrily as I do not feel angry now”. In a study of spontaneous affective expression, Capps and colleagues (1993) found high-functioning children with ASD showed more appropriate facial expressions when watching empathy videos, while their peers displayed neutral facial expression most of the time. The display of neutral expression of typical children in Capps et al. (1993) and the embarrassment of typical adolescents in the present study may be interpreted as evidence for awareness of expressing emotion according to the social context.

However, the performance proficiency of adolescents with ASD in emotional prosody was not repeated in attitudinal prosody. Only the young adults conveyed seriousness with more pressed voice quality than friendliness, while all adolescents in the current study produced breathier voice quality for expressing seriousness. Since results from past research suggested that both happiness and friendliness are associated with smaller body-size projection (Noble & Xu, 2011), only the performance of young adults complied with the predictions from body-size projection. Two explanations are possible for this difference in performance. First, age may be a factor of influence, since only young adults were able to express emotions and attitudes according to predictions from body-size projection. Second, adolescents in this study might have had more difficulties in expressing attitudes than emotions, since attitudinal expressions are more subtle and more culturally controlled than emotional ones (Mitchell & Ross, 2013; Scherer, 2003). A comment from a typical adolescent participant illustrates this difference: “I think emotions are easier to act out than attitudes, because you need the right situation for saying this (stimulus) with the right attitude”. Therefore, the inability of adolescents with ASD and typically developing adolescents in performing attitude may be interpreted as developmental course of prosodic skills, with attitudinal prosody maturing later than emotional ones. Further study may investigate whether adolescents master attitudinal prosody skills later in age than emotional prosody skills.

No group seemed to use VTL-related acoustic parameters to express emotions and attitudes in prosody. In a previous study investigating prosodic perception among adolescents with ASD, typically developing adolescents and young adults, manipulations in formant shift ratio have proven of great importance for judging different emotions and attitudes (Hsu & Xu, 2014). Yet, they did not manipulate formant dispersions to produce prosody with different emotions and attitudes. A possible explanation is that, in fact, it is rather challenging to deliberately control the full vocal apparatus in order to achieve acoustic effects of a certain emotion or attitude. However, when speakers feel the emotion, they are able to intuitively produce enough acoustic variations to convey the feelings.

Table 20. Emotion and attitude production according to the predictions from the body size projection by AA, TA and YA groups

		AA	TA	YA
Emotion	Mean pitch (Hz)	✓	✓	✓
	Mean pitch (ST)	✓	✓	✓
	Pitch range	✓	✓	✓
	Excursion size	✓	✓	✓

Table 20. Emotion and attitude production according to the predictions from the body size projection by AA, TA and YA groups

		AA	TA	YA
	Formant dispersion 3	✗	✗	✗
	Formant dispersion 5	✗	✗	✗
	H1-H2	✓	✗	✓
	H1*-H2*	✓	✓	✓
	H1-A1	✓	✗	✓
	H1-A3	✓	✗	✓
	Centre of gravity (Hz)	✓	✗	✓
	Centre of gravity (ST)	✓	✗	✓
	Energy below 500 Hz	✓	✗	✓
	Energy below 1,000 Hz	✓	✗	✓
	Total ✓	12	5	12
	Total ✗	2	9	2
Attitude	Mean pitch (Hz)	✓	✓	✓
	Mean pitch (ST)	✓	✓	✓
	Pitch range	✓	✓	✓

Table 20. Emotion and attitude production according to the predictions from the body size projection by AA, TA and YA groups

	AA	TA	YA
Excursion size	✓	✓	✓
Formant dispersion 3	✗	✗	✗
Formant dispersion 5	✓	✗	✓
H1-H2	✓	✓	✓
H1*-H2*	✓	✓	✓
H1-A1	✗	✓	✓
H1-A3	✗	✗	✓
Centre of gravity (Hz)	✗	✗	✓
Centre of gravity (ST)	✗	✗	✓
Energy below 500 Hz	✓	✗	✓
Energy below 1,000 Hz	✗	✗	✓
Total ✓	8	7	13
Total ✗	6	7	1

5.3 Perception of emotional and attitudinal prosody in PD

In general, PD patients and healthy elderly made similar perceptual judgments in all body-size projection, emotion and attitude tasks, and both groups were less sensitive to the acoustic manipulations than young adults. The current findings concerning PD patients and healthy elderly were different from those found in other studies in English and German, which found a lower performance of PD patients when compared to healthy elderly, especially in emotions with negative valence, such as anger, sadness and fear (Breitenstein et al., 2001; Dara et al., 2008; Mitchell & Bouças, 2009). Yet, in the present study, the greatest difference was found when comparing the performance of the two elderly groups to the young adults. Similar to the findings in the perceptual study of adolescents with ASD, young adults demonstrated greater sensitivity to the changes in formant shift ratio, pitch shift and pitch range when judging body size, emotion and attitude. In addition to the findings in PD patients and healthy elderly, it implies that perception of emotional and attitudinal prosody may be influenced by both development and ageing.

Past research exploring prosody perception often found that older people were usually less accurate than young people in emotion identification tasks. For instance, Lima et al. (2014) showed that older adults were less accurate than younger adults when identifying vocal emotions, but this difference in performance was not associated with the emotional valence. The findings in Dupuis and Pichora-Fuller (2015) confirmed the tendency of older people to make more mistakes than younger people when judging emotion from semantically neutral sentences. Furthermore, Paulmann, Pell and Kotz (2008) demonstrated that adults with mean age of 42 years also presented similar deficits in perception of speech emotion, thus they concluded that the decline in vocal emotional perception may have started as early as middle age.

One may argue that auditory acuity plays a major role in speech perception of an ageing population, as older people often present some age-related hearing loss. Yet, in the current study, although all older participants had some hearing loss at higher frequencies as indicated in the results from the pure-tone audiometry, different hearing thresholds between old and young participants should not be the main cause of the difference in the perceptual performance, as all the participants were allowed to adjust the volume of the recordings according to their comfort hearing level. Moreover, Dupuis and Pichora-Fuller (2015) found no significant correlations between the vocal emotion identification accuracy and measures of auditory acuity in older adults. Therefore, the peripheral hearing loss may have a limited effect on their perception performance. Future studies have to be performed in order to clarify the effects of loudness on perception of emotional prosody.

Current results showed no difference in performance between PD patients and healthy elderly in judging body size, emotion and attitude in formant shift ratio, pitch shift and pitch range. However, Mitchell & Kingston (2014) found strong correlation between pitch perception abilities and emotional prosody discrimination performance in older adults. They concluded that pitch discrimination is sensitive to predict older people’s emotional perception ability. A possible explanation for the results in the present study is that the level of manipulation of acoustic parameters, which has been shown to be effective only for young people (Chuenwattanapranithi et al., 2008; Hsu & Xu, 2014; Xu, Kelly, et al., 2013; Xu, Lee, et al., 2013), may not be sufficient for older adults. Thus, further research is needed to explore whether perception by older adults improve with larger changes of acoustic parameters.

Table 21. Perceptual judgment according to the predictions from the body size projection by PD, HE and YA groups

		PD	HE	YA
Body size	Voice quality	✗	✓	✗
	Formant shift ratio	✓	✗	✓
	Pitch shift	✗	✓	✓
	Pitch range	✗	✗	✓
Emotion	Voice quality	✗	✗	✗
	Formant shift ratio	✓	✓	✓
	Pitch shift	✓	✗	✓
	Pitch range	✗	✗	✓
Attitude	Voice quality	✗	✗	✗
	Formant shift ratio	✓	✓	✓
	Pitch shift	✓	✓	✓
	Pitch range	✓	✓	✓
	Total ✓	8	8	9
	Total ✗	6	6	3

5.4 Production of emotional and attitudinal prosody in PD

PD patients, healthy elderly and young adults manipulated most of the acoustic parameters to express emotions and attitudes in prosody. The performance of emotional prosody by PD patients was similar to healthy elderly and young adults in all traditional acoustic parameters – mean pitch, intensity, mean syllable duration and pitch range. PD patients even outperformed healthy elderly in mean pitch, and young adults in pitch range. For instance, PD patients conveyed anger with lower mean pitch and narrower pitch range than happiness, while the two control groups did the opposite. As lower mean pitch and narrower pitch range are associated with larger or angry-sounding speakers, thus only PD patients performed these emotions according to the predictions of body-size projection theory.

Interestingly, in the present study, PD patients were able to adjust intensity and mean syllable duration to convey different emotions in the same way as the other two control groups, although the significance levels of the effects of these acoustic parameters for the PD groups are smaller. These findings suggest PD patients are able to control intensity and mean syllable duration for conveying different emotions, which disagree with the general clinical description of PD speech as invariant in loudness and slow in speech rate (Duffy, 2005).

In attitudinal prosody, PD patients, healthy elderly and young adults manipulated fewer traditional acoustic parameters than in emotional prosody – only mean pitch and intensity showed to be significantly different across different attitudes. Healthy elderly and young adults produced seriousness with lower mean pitch and higher intensity than friendliness, which are associated with larger speakers. Yet the performance of PD patients was exactly the inverse, contradicting the principles of body-size projection.

PD patients also performed voice-quality related acoustic parameters differently in the two prosodic modalities. In emotional prosody, PD patients were as proficient as healthy elderly and young adults in expressing different emotions. All participants of the three groups produced anger with lower H1-H3 and energy below 500 Hz and 1,000 Hz, and with higher CG than happiness. The changes of these acoustic parameters are associated with larger or angry-sounding speakers, as shown in past research (Chuenwattanapranithi et al., 2008; Hsu & Xu, 2014; Xu, Kelly, et al., 2013; Xu & Kelly, 2010; Xu, Lee, et al., 2013). Therefore, in the current study, PD patients were as competent as healthy elderly and young adults in manipulating traditional and voice quality-related acoustic parameters to express different emotions in a social context-free environment, such as the experimental tasks in this study.

The largest differences between testing groups were once again observed in the attitudinal prosody. PD patients did not show the same capability in changing voice quality-related acoustic parameters as in emotional prosody. Healthy elderly and young adults conveyed seriousness with lower H1-H2 and energy below 500 Hz and 1,000 Hz, and with higher CG than friendly. These manipulations are linked to larger sounding speakers. PD patients only managed to use H1-H2 similarly, yet they manipulated all other acoustic parameters mentioned above in the opposite manner. PD patients expressed seriousness with breathier voice quality than friendliness. Therefore, they performed seriousness with smaller body-size projection, and friendliness with larger projection, a finding opposite to that of Noble & Xu (2011).

Two issues can be raised from the difference in performance of emotional and attitudinal prosody. First, this difference may have its origins in motor speech symptoms. PD patients are commonly reported to present fewer variations in pitch and loudness and slow articulation in speech as a result of their motor symptoms. Yet in the current study, the same utterances were said in both emotion and attitude tasks. Therefore, if PD patients were really affected by the motor speech symptoms, they would have showed similar acoustic patterns for both tasks. In addition, Möbes and colleagues (2008) found that PD patients were not different from healthy speakers in imitating prosody, but performed significantly worse in the emotional expression task. Therefore, the researcher suggested that alterations of emotional processing, in addition to motor impairment, are the underlying cause of PD patients’ deficits in emotional prosody.

Second, the difference in cognitive skills between PD patients and healthy elderly may also account for this discrepancy in emotion and attitude task outcomes. In the present study, all PD patients passed the MoCA test, but their test scores were significantly lower than those of healthy elderly. It seems likely that cognitive decline is the major cause of this performance discrepancy, because attitude may require more cognitive skills than emotion, by being more subtle and socially driven than emotion (Mitchell & Ross, 2013; Scherer, 2003). However, in some past research studies in vocal emotion perception, no significant correlation has been found between cognitive test scores and outcomes of emotion tasks in PD patients when compared to healthy elderly (Dara et al., 2008; Dujardin et al., 2004; Garrido-Vásquez et al., 2013; Lima et al., 2014). Since attitude remains largely unexplored among populations with neurological conditions, such as PD (Monetta, Cheang, & Pell, 2008), further research is needed to explore vocal expression of attitude among atypical populations.

Finally, similar to the outcomes in expressive prosody tasks by adolescents with ASD, typically developing adolescents and young adults, PD patients, healthy elderly and young adults also did not manipulate VTL-related acoustic parameters to express different emotions and attitudes in prosody. Altogether, these results provide new insights into the complex mechanisms of vocal control in prosody production.

Table 22. Emotion and attitude production according to the predictions from the body size projection by PD, HE and YA groups

		PD	HE	YA
Emotion	Mean pitch (Hz)	✓	✓	✓
	Mean pitch (ST)	✓	✓	✓
	Pitch range	✓	✗	✓
	Excursion size	✓	✗	✓
	Formant dispersion 3	✗	✗	✗
	Formant dispersion 5	✓	✓	✓
	H1-H2	✓	✓	✓
	H1*-H2*	✓	✗	✓
	H1-A1	✓	✓	✓
	H1-A3	✓	✓	✓
	Centre of gravity (Hz)	✓	✓	✓
	Centre of gravity (ST)	✓	✓	✓
	Energy below 500 Hz	✓	✓	✓
	Energy below 1,000 Hz	✓	✓	✓

Table 22. Emotion and attitude production according to the predictions from the body size projection by PD, HE and YA groups

		PD	HE	YA
	Total ✓	13	10	13
	Total ✗	1	4	1
Attitude	Mean pitch (Hz)	✓	✓	✓
	Mean pitch (ST)	✓	✓	✓
	Pitch range	✓	✗	✓
	Excursion size	✓	✗	✓
	Formant dispersion 3	✗	✓	✗
	Formant dispersion 5	✓	✓	✓
	H1-H2	✓	✓	✓
	H1*-H2*	✓	✓	✓
	H1-A1	✓	✓	✓
	H1-A3	✓	✓	✓
	Centre of gravity (Hz)	✓	✓	✓
	Centre of gravity (ST)	✓	✓	✓
	Energy below 500 Hz	✓	✗	✓

Table 22. Emotion and attitude production according to the predictions from the body size projection by PD, HE and YA groups

		PD	HE	YA
	Energy below 1,000 Hz	✗	✓	✓
	Total ✓	12	11	13
	Total ✗	2	3	1

Chapter 6 Conclusion

The present study compared the perception and production of emotional and attitudinal prosody in five different groups – high-functioning adolescents with ASD, typically developing adolescents, young adults, PD patients and healthy elderly.

Compared to typically developing peers and young adults, adolescents with ASD are less sensitive to changes in acoustic parameters when judging speaker's body size, emotion and attitude through speech prosody, while typically developing adolescents are less sensitive to the same acoustic changes than young adults. Moreover, young adults are the only group who have shown sensitivity to changes in all acoustic parameters. When expressing emotions through prosody, the performance of adolescents with ASD is comparable to young adults, and they even outperform typically developing adolescents. However, only young adults are able to convey different attitudes in prosody.

PD patients and healthy elderly are not responsive to acoustic changes when perceiving speaker's body size, emotion and attitude from prosody, while young adults are aware of all the changes. When producing emotions in prosody, PD patients are as capable as healthy elderly and young adults to convey various emotions. Yet, PD patients are not able to express attitude with different prosody, whereas healthy elderly and young adults are.

These findings lead us to conclude that: 1) individuals with ASD have a reduced rather than a total lack of ability to perceive emotional prosody, while PD patients are not able to judge emotions from prosody, thus both groups already present disturbance in the perception of emotional prosody in the social context-free situation (i.e. during the recording session in the speech lab); 2) both individuals with ASD and PD patients are able to express emotion through prosody, but they are not capable to convey attitude in speech; 3) the ability to process emotional prosody may be still in development from adolescence to early adulthood, and gradually decline in late adulthood; 4) atypical populations such as ASD and PD have more difficulties in expressing attitudinal prosody than emotional prosody.

This research paradigm provides a more comprehensive description of perception and production of emotional and attitudinal prosody in ASD and PD populations, given that the results derived from the traditional acoustic parameters can be compared to past research outcomes, as the BID-associated acoustic parameters allow a closer examination of the suitability of BID theory, particularly of body size projection, in understanding prosody in normal and disordered speech.

Reference

Adolphs, R., Schul, R., & Tranel, D. (1998). Intact recognition of facial emotion in Parkinson's disease. *Neuropsychology*, *12*(2), 253–258. doi:10.1037//0894-4105.12.2.253

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *The Quarterly Journal of Experimental Psychology, 65*(5), 825–839. doi:10.1080/17470218.2012.676055

Ariatti, A., Benuzzi, F., & Nichelli, P. (2008). Recognition of emotions from visual and prosodic cues in Parkinson's disease. *Neurological Sciences*, *29*(4), 219–227. doi:10.1007/s10072-008-0971-9

Ashwin, C., Chapman, E., Colle, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience, 1*(3-4), 349–363. doi:10.1080/17470910601040772

Asperger, H. (1944). "Autistic psychopathy" in childhood. In U. Frith (Ed.), *Autism and Asperger Syndrome* (pp. 37–92). Cambridge: Cambridge University Press.

Assogna, F., Pontieri, F. E., Caltagirone, C., & Spalletta, G. (2008). The recognition of facial emotion expressions in Parkinson's disease. *European Neuropsychopharmacology, 18*(11), 835–848. doi:10.1016/j.euroneuro.2008.07.004

Baker, K. K., Ramig, L. O., Sapir, S., Luschei, E. S., & Smith, M. E. (2001). Control of vocal loudness in young and old adults. *Journal of Speech, Language, and Hearing Research, 44*(2), 297–305. doi:10.1044/1092-4388(2001/024)

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. doi:10.1037/0022-3514.70.3.614

Baron-Cohen, S. (1990). Autism: A Specific Cognitive Disorder of & “Mind-Blindness.” *International Review of Psychiatry*, *2*(1), 81–90. doi:10.3109/09540269009028274

Baron-Cohen, S. (1991). Do people with autism understand what causes emotion? *Child Development*, *62*(2), 385–395. doi:10.1111/j.1467-8624.1991.tb01539.x

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation*, *23*, 169–184. doi:http://dx.doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ) - Adolescent version. *Journal of Autism and Developmental Disorders*, *36*(3), 343–350. doi:10.1007/s10803-006-0073-6

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., & Brayne, C. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *British Journal of Psychiatry*, *194*(6), 500–509. doi:10.1192/bjp.bp.108.059345

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*, 311–331. doi:10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. doi:10.1023/A:1005653411471

Begeer, S., Banerjee, R., Rieffe, C., Terwogt, M. M., Potharst, E., Stegge, H., & Koot, H. M. (2011). The understanding and self-reported use of emotional display rules in children with autism spectrum disorders. *Cognition and Emotion*, *25*(5), 947–956. doi:10.1080/02699931.2010.516924

Begeer, S., Koot, H. M., Rieffe, C., Meerum Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review*, *28*(3), 342–369. doi:10.1016/j.dr.2007.09.001

Birkholz, P., Jackèl, D., & Kröger, B. J. (2006). Construction and control of a three-dimensional vocal tract model. In *ICASSP 2006*.

Birkholz, P., Kröger, B. J., & Neuschaefer-Rube, C. (2011). Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis. In *INTERSPEECH 2011* (pp. 2681–2684). Firenze, Italy.

Blonder, L. X., & Slevin, J. T. (2011). Emotional dysfunction in Parkinson's disease. *Behavioural Neurology*, *24*(3), 201–217. doi:10.3233/BEN-2011-0329

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer. Retrieved from http://www.praat.org/

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*, 1–7. doi:10.3389/fnhum.2010.00237

Braak, H., Del Tredici, K., Bratzke, H., Hamm-Clement, J., Sandmann-Keil, D., & Rüb, U. (2002). Staging of the intracerebral inclusion body pathology associated with idiopathic Parkinson's disease (preclinical and clinical stages). *Journal of Neurology*, *249*(supplement III), III/1–5. doi:10.1007/s00415-002-1301-4

Breitenstein, C., van Lancker, D., Daum, I., & Waters, C. H. (2001). Impaired perception of vocal emotions in Parkinson’s disease: Influence of speech time processing and executive functioning. *Brain and Cognition*, *45*(2), 277–314. doi:10.1006/brcg.2000.1246

Brennand, R., Schepman, A., & Rodway, P. (2011). Vocal emotion perception in pseudo-sentences by secondary-school children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *5*(4), 1567–1573. doi:10.1016/j.rasd.2011.03.002

British Society of Audiology. (2011). Recommended procedure: Pre-tone air-conduction and bone-conduction threshold audiometry with and without masking. Reading, UK: Author.

Brooks, P. J., & Ploog, B. O. (2013). Attention to emotional tone of voice in speech perception in children with autism. *Research in Autism Spectrum Disorders*, *7*(7), 845–857. doi:10.1016/j.rasd.2013.03.003

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence. Third Edition - A Language-free Measure of Cognitive Ability*. Austin, Texas: Pro-ed.

Buxton, S. L., MacDonald, L., & Tippett, L. J. (2013). Impaired recognition of prosody and subtle emotional facial expressions in Parkinson’s disease. *Behavioral Neuroscience*, *127*(2), 193–203. doi:10.1037/a0032013

Capps, L., Kasari, C., Yirmiya, N., & Sigman, M. (1993). Parental perception of emotional expressiveness in children with autism. *Journal of Consulting and Clinical Psychology*, *61*(3), 475–484. doi:10.1037/0022-006X.61.3.475

Centers for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly*

Report. Surveillance Summaries, *63*(2), 1–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24670961

Centers for Disease Control and Prevention. (2015). Autism Spectrum Disorders - Data and Statistics. Retrieved from http://www.cdc.gov/ncbddd/autism/data.html

Chao, Y. R. (1968). *Linguistic Inquiries*. Taipei, Taiwan: Taiwan Commercial Press.

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(3), 619–627. doi:10.1017/S0033291710000991

Cheang, H. S., & Pell, M. D. (2007). An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. *Journal of Neurolinguistics*, *20*(3), 221–241. doi:10.1016/j.jneuroling.2006.07.001

Chen, Y., Xu, Y., & Guion-Anderson, S. (2014). Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica*, *71*, 249–270. doi:10.1159/000371891

Chen, Y.-R., & Liu, H. M. (2010). The verbal intonation of expressive emotion in elementary school children with high-function autism. *Bulletin of Special Education*, *35*(2), 55–79.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger syndrome. *Research in Autism Spectrum Disorders, 3*, 502–516. doi:10.1016/j.rasd.2008.10.004

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code: A perceptual investigation.

Phonetica, *65*, 210–230. doi:10.1159/000192793

Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language Acquisition* (pp. 33–48). Cambridge: Cambridge University Press.

Crystal, D. (2009). Persevering with prosody. *International Journal of Speech-Language Pathology*, *11*(4), 257. doi:10.1080/17549500902858753

Cubells, J. F. (2013). Prevalence of autism spectrum dsorders in China. *Shanghai Archives of Psychiatry*, *25*(3), 176–177. doi:10.3969/j.issn.1002-

Da Fonseca, D., Santos, A., Bastard-Rosset, D., Rondan, C., Poinso, F., & Deruelle, C. (2009). Can children with autistic spectrum disorders extract emotions out of contextual cues? *Research in Autism Spectrum Disorders*, *3*, 50–56. doi:10.1016/j.rasd.2008.04.001

Dara, C., Monetta, L., & Pell, M. D. (2008). Vocal emotion processing in Parkinson’s disease: Reduced sensitivity to negative emotions. *Brain Research*, *1188*(1), 100–111. doi:10.1016/j.brainres.2007.10.034

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, *12*(2), 246–269. doi:10.1044/jshr.1202.246

Darwin, C. R. (1872). *The Expression of Emotion in Man and Animals*. London: John Murray.

den Brok, M. G. H. E., van Dalen, J. W., van Gool, W. A., van Charante, E. P. M., de Bie, R. M. A., & Richard, E. (2015). Apathy in Parkinson’s disease: A systematic review and meta-analysis. *Movement Disorders*, *30*(6), 759–769. doi:10.1002/mds.26208

DePaolis, R. A., Vihman, M. M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics, 36*(2), 406–422. doi:10.1016/j.wocn.2008.01.003

Diehl, J. J., & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics, 34*(1), 135–161. doi:10.1017/S0142716411000646

Diehl, J. J., Watson, D. G., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*. doi:10.1017/S0142716409090201

Duanmu, S. (2004). Tone and non-tone languages: An alternative to language typology and parameters. *Language and Linguistics*, *5*(4), 891–924.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management*. St. Louis, MO: Mosby-Year Book.

Dujardin, K., Blairy, S., Defebvre, L., Duhem, S., Noël, Y., Hess, U., & Destée, A. (2004). Deficits in decoding emotional facial expressions in Parkinson's disease. *Neuropsychologia*, *42*(2), 239–250. doi:10.1016/S0028-3932(03)00154-4

Dujardin, K., Sockeel, P., Devos, D., Delliaux, M., Krystkowiak, P., Destée, A., & Defebvre, L. (2007). Characteristics of apathy in Parkinson's disease. *Movement Disorders*, *22*(6), 778–784. doi:10.1002/mds.21316

Dunn, L. M., & Dunn, L. M. (1981). *The Peabody Picture Vocabulary Test - Revised*. San Antonio: PsychCorp.

Dupuis, K., & Pichora-Fuller, M. K. (2015). Aging affects identification of vocal emotions in semantically neutral sentences. *Journal of Speech, Language, and Hearing Research*, *58,* 1061–1076. doi:10.1044/2015

Dworkin, J. P. (1991). *Motor Speech Diorders - A treatment guide*. St. Louis, MO: Mosby.

Edelson, M. G. (2006). Are the majority of children with autism mentally retarded? A systematic evaluation of the data. *Focus on Autism and Other Developmental Disabilities, 21*(2), 66–83.

Edwards, M., Quinn, N., & Bhatia, K. (2008). *Parkinson's disease and other movement disorders*. New York: Oxford University Press. doi:10.1016/j.amjopharm.2010.08.005

Eigsti, I.-M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*, *5*(2), 681–691. doi:10.1016/j.rasd.2010.09.001

Ekman, P., & Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA, CA: Consulting Psychologists Press.

Fahn J. Jankovic, S., & Hallett, M. (2007). Clinical overview and phenomenology of movement disorders. In S. Fahn & J. Jankovic (Eds.), *Principles and Practice of Movement Disorders*. Philadelphia, PA: Elsevier, Churchill Livingstone.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). A Flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America, 102*(2), 1213–1222. doi:10.1121/1.421048

Fox, A. (2000). *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford: Oxford University Press.

Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, *32*(6), 969–979. doi:10.1016/S0896-6273(01)00552-9

Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language*, *14*(1), 1–22. doi:10.1111/1468-0017.00100

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., Strecker, K., Schwarz, J., & Kotz, S. A. (2013). An ERP study of vocal emotion processing in asymmetric Parkinson’s disease. *Social Cognitive and Affective Neuroscience*, *8*(8), 918–927. doi:10.1093/scan/nss094

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, *8*(2), 153–163. doi:10.1002/aur.1432

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The “Reading the Mind in the Voice” test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*, 1096–1106. doi:10.1007/s10803-006-0252-5

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult ... but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315. doi:10.1080/17549500903003060

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *53*(3), 778–793. doi:10.1044/1092-4388(2009/08-0127)

Grossman, R. B., & Tager-Flusberg, H. (2012). “Who said that?” Matching of low- and high-intensity emotional prosody to facial expressions by adolescents with ASD. *Journal of Autism and Developmental Disorders*, *42*, 2546–2557.

doi:10.1007/s10803-012-1511-2

Gussenhoven, C. (2001a). Suprasegmentals. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and the Behavioural Sciences* (pp. 15294–15298). Oxford, UK.

Gussenhoven, C. (2001b). Suprasegmentals. *International Encyclopedia of the Social and the Behavioural Sciences*, 15294–15298. doi:doi:10.1016/B0-08-043076-7/02980-6

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511616983

Gussenhoven, C. (2015). Suprasegmentals. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 714–721). Oxford, UK: Elsevier.

Halliday, G., Lees, A., & Stern, M. B. (2011). Milestones in Parkinson's disease-Clinical and pathologic features. *Movement Disorders*, *26*(6), 1015–1021. doi:10.1002/mds.23669

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*, 290–322. doi:10.1007/s11065-010-9138-6

Hartelius, L., & Svensson, P. (1994). Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: A survey. *Folia Phoniatrica et Logopaedica*, *46*(1), 9–17. doi:10.1159/000266286

Heaton, P., Reichenbacher, L., Sauter, D. A., Allen, R., Scott, S., & Hill, E. (2012). Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. *Psychological Medicine*, *42*(11), 2453–2459. doi:10.1017/S0033291712000621

Heikkinen, J., Jansson-Verkasalo, E., Toivanen, J., Suominen, K., Väyrynen, E., Moilanen, I., & Seppänen, T. (2010). Perception of basic emotions from speech prosody in adolescents with Asperger's syndrome. *Logopedics, Phoniatrics, Vocology, 35*, 113–120. doi:10.3109/14015430903311184

Hirtz, D., Thurman, D. J., Gwinn-Hardy, K., Mohamed, M., Chaudhuri, A. R., & Zalutsky, R. (2007). How common are the "common" neurologic disorders? *Neurology, 68*(5), 326–337. doi:10.1212/01.wnl.0000252807.38124.a3

Hoehn, M. M., & Yahr, M. D. (1967). Parkinsonism: Onset, progression, and mortality. *Neurology, 17*(5), 427–442.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research, 38*(6), 1212–1223.

Holroyd, S., & Baron-Cohen, S. (1993). Brief report: How far can people with autism go in developing a theory of mind? *Journal of Autism and Developmental Disorders, 23*(2), 379–385.

Hsu, C., & Xu, Y. (2014). Can adolescents with autism perceive emotional prosody? In *INTERSPEECH 2014* (pp. 1924–1928). Singapore.

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research, 36*, 159–173. doi:10.1007/s10936-006-9037-4

IBM Corp. (2013). IBM SPSS Statistics for Windows. Armonk, NY, NY: IBM Corp.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 10–13). doi:10.1109/ICASSP.2004.1326074

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America*, *121*(4), 2283–2295. doi:10.1121/1.2697522

Jaywant, A., & Pell, M. D. (2010). Listener impressions of speakers with Parkinson's disease. *Journal of the International Neuropsychological Society*, *16*(1), 49–57. doi:10.1017/S1355617709990919

Jazen, E. J. (2003). *Understanding the Nature of Autism: A guide to the autism spectrum disorders*. San Antonio: Therapy Skill Builders.

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of melodic contour and intonation in autism spectrum disorder: Evidence from Mandarin speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075. doi:10.1007/s10803-015-2370-4

Kan, Y., Kawamura, M., Hasegawa, Y., Mochizuki, S., & Nakamura, K. (2002). Recognition of emotion from facial, prosodic and written verbal stimuli in Parkinson's disease. *Cortex, 38*(4), 623–630. doi:10.1016/S0010-9452(08)70026-1

Kandiah, N., Zhang, A., Cenina, A. R., Au, W. L., Nadkarni, N., & Tan, L. C. S. (2014). Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease. *Parkinsonism and Related Disorders*, *20*(11), 1145–1148. doi:10.1016/j.parkreldis.2014.08.002

Kanner, L. (1943). Autistic disturbances of affective contact. *The Nervous Child, 2,* 217–250. doi:10.1105/tpc.11.5.949

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*(2), 353–365. doi:10.1121/1.388089

Kristen, S., Rossmann, F., & Sodian, B. (2014). Theory of own mind and autobiographical memory in adults with ASD. *Research in Autism Spectrum Disorders*, *8*(7), 827–837. doi:10.1016/j.rasd.2014.03.009

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Langston, J. W. (2006). The Parkinson's complex: Parkinsonism is just the tip of the iceberg. *Annals of Neurology*, *59*(4), 591–596. doi:10.1002/ana.20834

Le Sourn-Bissaoui, S., Aguert, M., Girard, P., Chevreuil, C., & Laval, V. (2013). Emotional speech comprehension in children and adolescents with autism spectrum disorders. *Journal of Communication Disorders*, *46*(4), 309–320. doi:10.1016/j.jcomdis.2013.03.002

Lehéricy, S., Sharman, M. A., Santos, C. L. dos, Paquin, R., & Gallea, C. (2012). Magnetic resonance imaging of the substantia nigra in Parkinson's disease. *Movement Disorders*, *27*(7), 822–830. doi:10.1002/mds.25015

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*. doi:10.1017/S0305000900001598

Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145–160. doi:10.1037/a0034287

Lin, J. D., Lin, L. P., & Wu, J. L. (2009). Administrative Prevalence of autism spectrum disorders based on national disability registers in Taiwan. *Research in Autism Spectrum Disorders*, *3*(1), 269–274. doi:10.1016/j.rasd.2008.07.002

Linville, S. E. (2004). The Aging Voice. *The ASHA Leader, October*, 11–13.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*(2-4), 70–87. doi:10.1159/000090090

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. *Journal of Speech and Hearing Disorders, 43*(1), 47–57. doi:10.1080/713789127

Ma, J. K. Y., Whitehill, T. L., & Cheung, K. S.-K. (2010). Dysprosody and stimulus effects in Cantonese speakers with Parkinson's disease. *International Journal of Language and Communication Disorders, 45*(6), 645–655. doi:10.3109/13682820903434813

Manstead, T. (2005). The social dimension of emotion. *Psychologist, 18*(8), 484–487. doi:10.1515/ling.1976.14.177.7

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders, 38*, 325–350. doi:10.1080/1368282031000154204

McCann, J., Peppé, S., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language and Communication Disorders*, *42*(6), 682–702. doi:10.1080/13682820601170102

Midi, I., Dogan, M., Koseoglu, M., Can, G., Sehitoglu, M. A., & Gunal, D. I. (2008). Voice abnormalities and their relation with motor dysfunction in Parkinson's disease. *Acta Neurologica Scandinavica, 117*(1), 26–34. doi:10.1111/j.1600-0404.2007.00965.x

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028–1039. doi:10.1111/j.1469-8986.2010.01031.x

Mitchell, R. L. C., & Bouças, S. B. (2009). Decoding emotional prosody in Parkinson's disease and its potential neuropsychological basis. *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 553–564. doi:10.1080/13803390802360534

Mitchell, R. L. C., & Kingston, R. A. (2014). Age-related decline in emotional prosody discrimination. *Experimental Psychology*, *61*(3), 215–223. doi:10.1027/1618-3169/a000241

Mitchell, R. L. C., & Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, *70*, 1–10. doi:10.1016/j.neuropsychologia.2015.02.018

Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience and Biobehavioral Reviews*, *37*(3), 471–479. doi:10.1016/j.neubiorev.2013.01.027

Möbes, J., Joppich, G., Stiebritz, F., Dengler, R., & Schröder, C. (2008). Emotional speech in Parkinson's disease. *Movement Disorders*, *23*(6), 824–829. doi:10.1002/mds.21940

Monetta, L., Cheang, H. S., & Pell, M. D. (2008). Understanding speaker attitudes from prosody by adults with Parkinson's disease. *Journal of Neuropsychology*, *2*(Pt 2), 415–430. doi:10.1348/174866407X216675

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855. doi:10.1086/283219

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, *4*(4), 345–357. doi:10.1177/1754073912445818

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097–1108. doi:10.1121/1.405558

Mwangi, S., Spiegl, W., Haderlein, T., & Maier, A. (2009). Effects of vocal aging on fundamental frequency and formants. In *NAG/DAGA 2009* (pp. 1761–1764). Rotterdam.

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. doi:10.1007/s10803-011-1264-3

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain and Development*, *36*(6), 516–522. doi:10.1016/j.braindev.2013.07.006

Narme, P., Bonnet, A. M., Dubois, B., & Chaby, L. (2011). Understanding facial emotion perception in Parkinson’s disease: The role of configural processing. *Neuropsychologia*, *49*(12), 3295–3302. doi:10.1016/j.neuropsychologia.2011.08.002

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. doi:10.1111/j.1532-5415.2005.53221.x

Noble, L., & Xu, Y. (2011). Friendly speech and happy speech – Are they the same? In *ICPhS XVII* (pp. 1502–1505). Hong Kong. Retrieved from http://discovery.ucl.ac.uk/1322248/

Nuffelen, G. Van. (2011). Speech prosody in dysarthria. In V. Stojanovik & J. Setter (Eds.), *Speech Prosody in Atypical Populations* (pp. 147–167). Albury, UK: J & R Press.

Nwe, T. L., Foo, S. W., & De Silva, L. C. (2003). Speech emotion recognition using hidden Markov models. *Speech Communication*, *41*(4), 603–623. doi:10.1016/S0167-6393(03)00099-2

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews*, *36*(2), 836–854. doi:10.1016/j.neubiorev.2011.11.008

Oguru, M., Tachibana, H., Toda, K., Okuda, B., & Oka, N. (2009). Apathy and depression in Parkinson disease. *Journal of Geriatric Psychiatry and Neurology*, *23*(1), 35–41. doi:10.1177/0891988709351834

Ohala, J. J. (1978). The production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). New York: Academic Press.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. doi:10.1159/000261706

Ohala, J. J., & Gilbert, J. B. (1978). Listeners' ability to identify languages by their prosody. *Report of the Phonology Laboratory Berkeley, Cal.*, *2*, 126–132.

Olufemi, T. D. (2012). Theories of attitudes. In C. D. Logan & M. I. Hodges (Eds.), *Psychology of Attitudes* (pp. 61–78). Nova Science Publishers, Inc.

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing "communication" and "apprehension". *Journal of Child Language*, *33*(1), 163–178. doi:10.1017/S0305000905007300

Parkinson, B. (1996). Emotions are social. *British Journal of Psychology, 87*, 663–683. doi:10.1631/jzus.2005.AS0128

Parkinson, J. (1817). *An Essay on the Shaking Palsy*. London, UK, UK: Sherwood, Neely and Jones. doi:10.1176/appi.neuropsych.14.2.223

Patel, R., & Brayton, J. T. (2009). Identifying prosodic contrasts in utterances produced by 4-, 7-, and 11-year-old children. *Journal of Speech, Language, and Hearing Research*, *52*(3), 790–801. doi:10.1044/1092-4388(2008/07-0137)

Patel, R., & Grigos, M. I. (2006). Acoustic characterization of the question-statement contrast in 4, 7 and 11 year-old children. *Speech Communication*, *48*(10), 1308–1318. doi:10.1016/j.specom.2006.06.007

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. doi:10.1007/s10803-004-1999-1

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. doi:10.1016/j.rasd.2007.04.001

Paul, R., Shriberg, L. D., McSweeny, J. L., Cicchetti, D., Klin, A., & Volkmar, F. R. (2005). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 861–869. doi:10.1007/s10803-005-0031-8

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, *104*(3), 262–269. doi:10.1016/j.bandl.2007.03.002

Pell, M. D., & Leonard, C. L. (2005). Facial expression decoding in early Parkinson's disease. *Cognitive Brain Research, 23*(2-3), 327–340. doi:10.1016/j.cogbrainres.2004.11.004

Penner, H., Miller, N., Hertrich, I., Ackermann, H., & Schumm, F. (2001). Dysprosody in Parkinson's disease: An investigation of intonation patterns. *Clinical Linguistics and Phonetics, 15*(7), 551–567. doi:10.1080/02699200110078140

Peppé, S. (2009). Why is prosody in speech-language pathology so difficult? *International Journal of Speech-Language Pathology, 11*(4), 258–271. doi:10.1080/17549500902906339

Peppé, S., Cleland, J., Gibbon, F., O'Hare, A., & Castilla, P. M. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics, 24*(1), 41–53. doi:10.1016/j.jneuroling.2010.07.005

Peppé, S., Martínez-Castilla, P., Coene, M., Hesling, I., Moen, I., & Gibbon, F. (2010). Assessing prosodic skills in five European languages: Cross-linguistic differences in typical and atypical populations. *International Journal of Speech-Language Pathology, 12*(1), 1–7. doi:10.3109/17549500903093731

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics, 17*(4-5), 345–354. doi:10.1080/0269920031000079994

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics, 38*(10), 1776–1791. doi:10.1016/j.pragma.2005.07.004

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 50,* 1015–1028.

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., … Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine, 40,* 1919–1929. doi:10.1017/S0033291709992364

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen & J. Morgan (Eds.), *Intentions in Communication* (pp. 271–311). Cambridge: MIT Press.

Pisanski, K., & Rendall, D. (2011). The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness. *The Journal of the Acoustical Society of America, 129*(4), 2201–2212. doi:10.1121/1.3552866

Ploog, B. O., Brooks, P. J., Scharf, A., & Aum, S. (2014). Perception of the prosody and content of sentences in an unfamiliar language in children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 8*(7), 775–787. doi:10.1016/j.rasd.2014.03.014

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4,* 515–526.

Pringsheim, T., Jette, N., Frolkis, A., & Steeves, T. D. L. (2014). The prevalence of Parkinson's disease: A systematic review and meta-analysis. *Movement Disorders, 29*(13), 1583–1590. doi:10.1002/mds.25945

Rigaldie, K., Nespoulous, J. L., & Vigouroux, N. (2006). Dysprosody in Parkinson's disease: Musical scale production and intonation patterns analysis. In *Speech Prososdy 2006* (pp. 4–7). Dresden, Germany.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32*(3), 189–194.

Samson, A. C., Phillips, J. M., Parker, K. J., Shah, S., Gross, J. J., & Hardan, A. Y. (2014). Emotion dysregulation and the core features of autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(7), 1766–1772. doi:10.1007/s10803-013-2022-5

Sapir, S., Ramig, L. O., & Fox, C. (2008). Speech and swallowing disorders in Parkinson disease. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 205–210. doi:10.1097/MOO.0b013e3282febd3a

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to Emotion* (pp. 293–317). Hillsdale, N.J., N.J.: Lawrence Erlbaum. Retrieved from http://books.google.com/books?hl=en&lr=&id=k0mhAwAAQBAJ&oi=fnd&pg=PA293&dq=On+the+Nature+and+Function+of+Emotion:+A+Component+Process+Approach&ots=kW_hOQiATU&sig=LDJmJ32m3_xCYdgZ4lVOvtu0A2M

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227–256. doi:10.1016/S0167-6393(02)00084-5

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76–92. doi:10.1177/0022022101032001009

Selkirk, E. (2005). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550–569). Oxford: Blackwell Publishing.

Shapiro, B. K., Menon, D. U., & Accardo, P. J. (2008). Clinical overview of the autism spectrum. In B. K. Shapiro & P. J. Accardo (Eds.), *Autism Frontiers: Clinical Issues and Innovations* (pp. 1–20). Baltimor: Paul H. Brookes Publishing.

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., … Singh, N. C. (2010). Sounds of melody - Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. doi:10.1016/j.neulet.2010.04.066

Shen, X. S. (1990). *The prosody of Mandarin Chinese*. Berkeley: University of California Press.

Shiota, M. N., & Kalat, J. W. (2012). *Emotion* (2nd ed.). Wadsworth.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. doi:10.1044/1092-4388(2001/087)

Shröder, C., Möbes, J., Schütze, M., Szymanowski, F., Nager, W., Bangert, M., … Dengler, R. (2006). Perception of emotional speech in Parkinson’s disease. *Movement Disorders*, *21*(10), 1774–1778. doi:10.1002/mds.21038

Sinaceur, M., Van Kleef, G. A., Neale, M. A., Adam, H., & Haag, C. (2011). Hot or cold: Is communicating anger or threats more effective in negotiation? *Journal of Applied Psychology*, *96*(5), 1018–1032. doi:10.1037/a0023896

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, *77*(2), 281–296. doi:10.1111/j.1467-8624.2006.00870.x

Spek, A. A., Scholte, E. M., & van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(3), 280–289. doi:10.1007/s10803-009-0860-y

Stark, R. E. (1979). Prespeech segmental feature development. In P. Fletcher & M.

Garman (Eds.), *Language Acquisition* (2nd ed., pp. 147–173). Cambridge: Cambridge University Press.

Stewart, T., Winfield, L., Hunt, A., Bressman, S. B., Fahn, S., Blitzer, I., & Brin, F. (1995). Speech dysfunction in early Parkinson's disease. *Movement Disorders*, *10*(5), 562–565.

Sun, X., Allison, C., Matthews, F. E., Sharp, S. J., Auyeung, B., Baron-Cohen, S., & Brayne, C. (2013). Prevalence of autism in mainland China, Hong Kong and Taiwan: A systematic review and meta-analysis. *Molecular Autism*, *4*(7), 1–13. doi:10.1186/2040-2392-4-7

Surendran, D., & Levow, G.-A. (2008). Can voice quality help Mandarin tone recognition? In *ICASSP 2008* (pp. 58–61). Las Vegas.

Tauber, S. K., James, L. E., & Noble, P. M. (2010). The effects of age on using prosody to convey meaning and on judging communicative effectiveness. *Psychology and Aging*, *25*(3), 702–707. doi:10.1037/a0019266

Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42*(5), 324–333. doi:10.1016/j.jcomdis.2009.03.001

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405. doi:10.1177/1754073911410747

Traunmüller, H., & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. *Consulté Le*, *12*(2).

Vaissière, J. (2008). Perception of Intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 236–263). Oxford: Blackwell Publishing.

doi:10.1002/9780470757024.ch10

Voss, T., & Hegeman Richard, I. (2011). Depression in Parkinson’s disease. In C. W. Olanow, F. Stocchi, & A. E. Lang (Eds.), *Parkinson’s Disease: Non-Motor and Non-Dopaminergic Features* (pp. 183–192). Oxford, UK: Blackwell Publishing Ltd.

Wang, J.-E., & Tsao, F.-M. (2015). Emotional prosody perception and its association with pragmatic language in school-aged children with high-function autism. *Research in Developmental Disabilities*, *37*(1), 162–170. doi:10.1016/j.ridd.2014.11.013

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, *31*(4), 749–778. doi:10.1017/S030500090400652X

Whitehill, T. L., Ma, J. K. Y., & Lee, A. S. Y. (2003). Perceptual characteristics of Cantonese hypokinetic dysarthria. *Clinical Linguistics and Phonetics*, *17*(4-5), 265–271. doi:10.1080/0269920031000080082

William, R., Gregory, A., & Frederick, A. (2009). Lewy Body Pathology in Normal Elderly Subjects, *68*(7), 816–822.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, *48*, 1065–1080.

Woodbury-Smith, M., Klin, A., & Volkmar, F. R. (2005). Asperger’s syndrome: A comparison of clinical diagnoses and those made according to the ICD-10 and DSM-IV. *Journal of Autism and Developmental Disorders*, *35*(2), 235–240. doi:10.1007/s10803-004-2002-x

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*, 2891–2902. doi:10.1007/s10803-013-1836-5

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y. (2006). Speech prosody as articulated communicative functions. In *Speech Prosody* (pp. SPS5–4–218). Retrieved from http://discovery.ucl.ac.uk/98103/

Xu, Y. (2014). ProsodyPro_BID.praat. London.

Xu, Y., Chen, S.-W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, *29*(2012), 131–147. doi:10.1515/tlr-2012-0006

Xu, Y., & Kelly, A. (2010). Perception of anger and happiness from resynthesized speech with size-related manipulations. In *Speech Prosody 2010* (pp. 1–4). Chicago. Retrieved from http://discovery.ucl.ac.uk/98127/

Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. In S. Hancil & D. Hirst (Eds.), *Prosody and Iconicity* (pp. 33–60). John Benjamins Publishing Company.

Xu, Y., Lee, A., Wu, W. L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLoS ONE*, *8*(4), e62397–e62397. doi:10.1371/journal.pone.0062397

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Appendix A

Table 23. British ISO Standard for reference zero for the calibration of audiometric equipment (in dB) used for the TDH 39 headphones

	ISO 389-1:2000	Right side	Left side
250 Hz	57.0	57.9	58.1
500 Hz	43.5	43.2	43.6
1,000 Hz	37.5	39.3	38.3
2,000 Hz	39.0	40.2	40.6
4,000 Hz	42.0	42.1	41.8
8,000 Hz	45.5	47.5	46.7

Appendix B

Table 24. Specifications of phonetic parameters of VocalTractLab for the synthetic utterance

MC	我		與	阿	姨	有			約	
pinyin	ǔo		yǔ	ā	yí	yǒu			yūe	
IPA	uo˨˩˦		y˨˩˦	a˥	i˧˥	iou˨˩˦			ye˥	
VocalTractLab phenetic symbol	w	O	ue:	A	i:	i:	o:	u:	ue:	ae:
Horizontal hyoid position	0.98	0.01	0.53	0.00	1.00	1.00	1.00	0.98	0.53	0.14
Vertical hyoid position	-5.33	-5.20	-4.45	-4.34	-5.40	-5.40	-5.79	-5.33	-4.45	-3.95
Horizontal jaw position	-0.15	-0.08	-0.20	-0.13	-0.27	-0.27	-0.14	-0.15	-0.20	-0.26
Vertical jaw position	-1.26	-1.38	-1.36	-1.29	-1.36	-1.36	-1.29	-1.26	-1.36	-1.34
Jaw opening angle	-0.09	-0.11	-0.13	-0.11	-0.08	-0.08	-0.11	-0.99	-0.13	-0.09
Lip protrusion	0.97	0.39	0.70	0.00	-0.44	-0.44	0.90	0.97	0.70	0.12

Table 24. Specifications of phonetic parameters of VocalTractLab for the synthetic utterance

MC	我		與	阿	姨	有			約	
pinyin	ǔo		yǔ	ā	yí	yǒu			yūe	
IPA	uo˨˩˦		y˨˩˦	a˥	i˧˥	iou˨˩˦			ye˥	
VocalTractLab phenetic symbol	w	O	ue:	A	i:	i:	o:	u:	ue:	ae:
Vertical lip distance	0.14	0.40	0.20	0.71	0.99	0.99	0.63	0.14	0.20	0.84
Velum position	0.00	0.05	0.03	0.22	0.00	0.00	0.00	0.00	0.03	0.01
Tongue centre X	-0.93	-0.48	2.23	0.00	1.58	1.58	-0.65	0.77	2.23	1.27
Tongue centre Y	0.81	-1.48	-0.75	-1.31	-0.80	-0.80	-1.29	-0.73	-0.75	-1.41
Tongue centre radius (X)	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80
Tongue centre radius (Y)	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80
Tongue tip X	3.11	3.68	4.57	3.96	4.27	4.27	2.14	3.11	4.57	4.25

Table 24. Specifications of phonetic parameters of VocalTractLab for the synthetic utterance

MC	我		與	阿	姨	有			約	
pinyin	ǔo		yǔ	ā	yí	yǒu			yūe	
IPA	uo˨˩˦		y˨˩˦	a˥	i˧˥	iou˨˩˦			ye˥	
VocalTractLab phenetic symbol	w	O	ue:	A	i:	i:	o:	u:	ue:	ae:
Tongue tip Y	-0.50	-2.16	-1.22	-1.65	-1.17	-1.17	-2.00	-0.50	-1.22	-1.04
Tongue blade X	2.85	1.61	3.00	1.92	3.60	3.60	1.78	2.85	3.00	2.61
Tongue blade Y	-0.19	-1.15	2.60	-0.82	2.00	2.00	-2.00	-1.00	2.60	0.77
Tongue root X	-2.40	-2.91	-0.68	-2.96	0.50	0.50	-1.50	-1.37	-0.68	-1.31
Tongue root Y	-2.97	-2.74	-1.13	-2.73	-2.74	-2.74	-2.98	-2.97	-1.13	-1.53
Tongue side elevation 1	0.99	0.30	1.00	0.61	0.36	0.36	0.18	0.99	1.00	0.98
Tongue side elevation 2	0.00	0.53	0.39	0.05	0.22	0.22	-0.48	0.00	0.39	0.47

Table 24. Specifications of phonetic parameters of VocalTractLab for the synthetic utterance

MC	我		與	阿	姨	有			約	
pinyin	ǔo		yǔ	ā	yí	yǒu			yūe	
IPA	uo˨˩˦		y˨˩˦	a˥	i˧˥	iou˨˩˦			ye˥	
VocalTractLab phenetic symbol	w	O	ue:	A	i:	i:	o:	u:	ue:	ae:
Tongue side elevation 3	0.06	0.26	0.01	0.43	0.15	0.15	-0.31	0.06	0.01	0.46
Tongue side elevation 4	-0.28	0.07	-0.05	-0.06	-0.01	-0.01	-0.40	-0.28	-0.05	0.16
Duration (ms)	0.12	0.20	0.32	0.20	0.30	0.07	0.12	0.09	0.23	0.28
F0 (ST)	33	35	33	36	35	29	29	29	30	30
Lung pressure (dPa) breathy	850	850	850	800	800	700	700	700	700	100
Lung pressure (dPa) modal	1000	1000	1000	900	700	700	700	700	700	100
Lung pressure (dPa) pressed	1500	1500	1500	1300	800	700	700	700	700	100

Appendix C

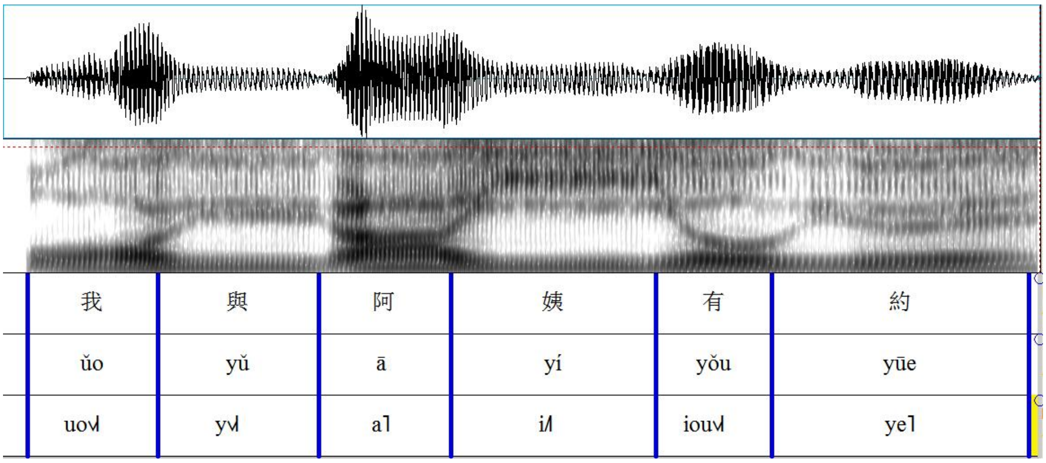

Figure 90: *Spectrogram of stimulus utterance with breathy voice quality*

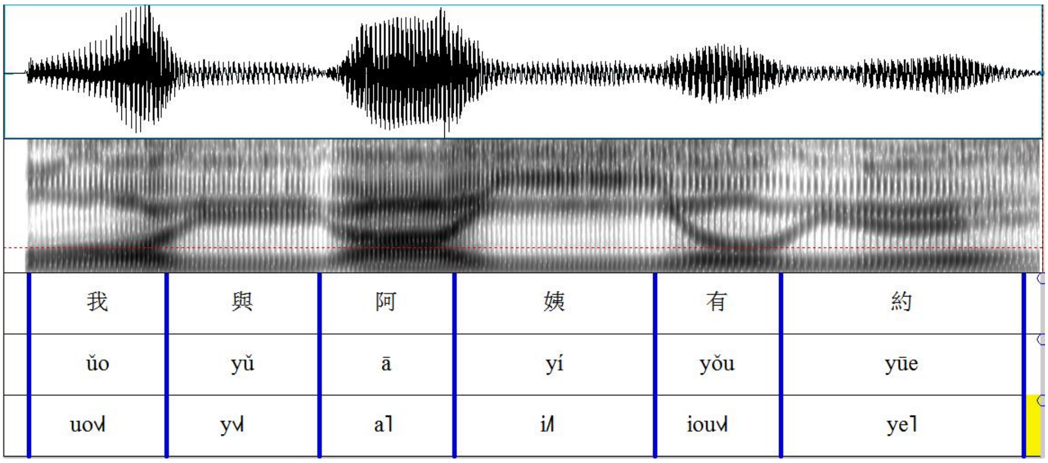

Figure 91: *Spectrogram of stimulus utterance with modal voice quality*

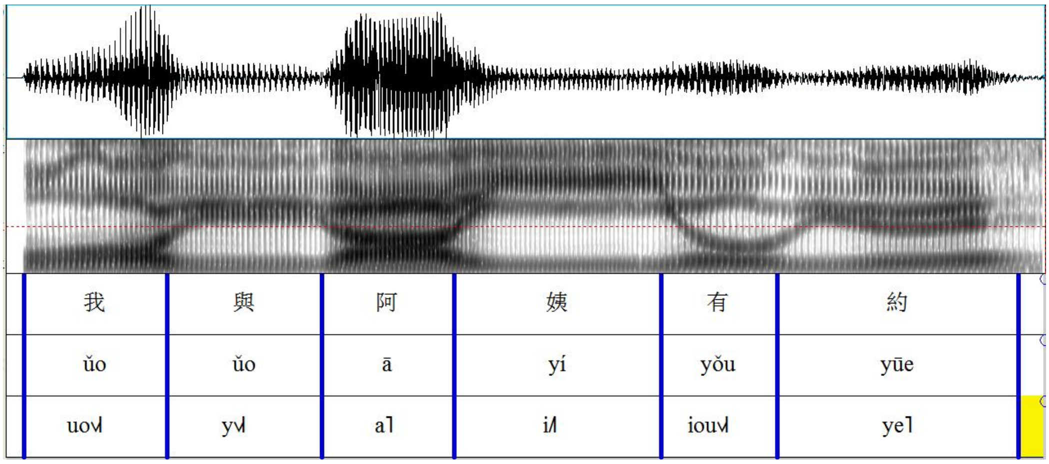

Figure 92: *Spectrogram of stimulus utterance with pressed voice quality*

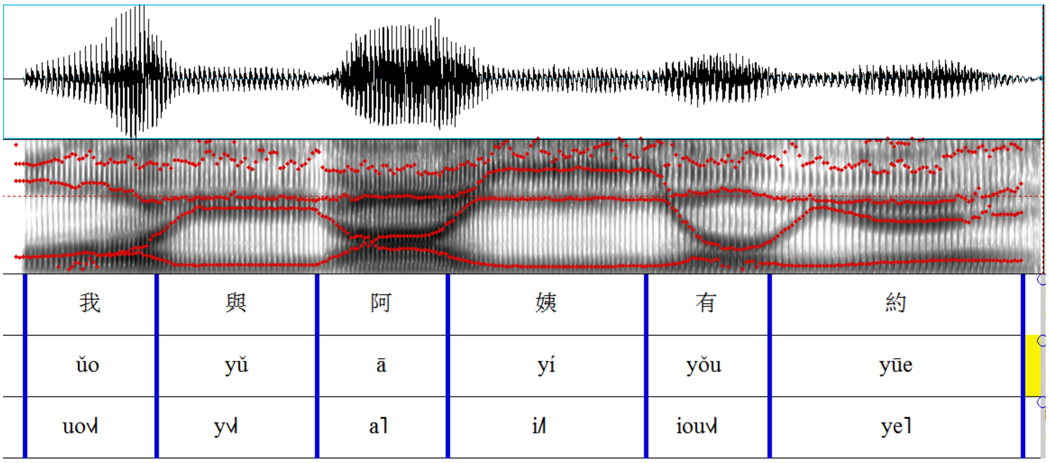

Figure 93: *Spectrogram of stimulus utterance with formant shift ratio 1.1*

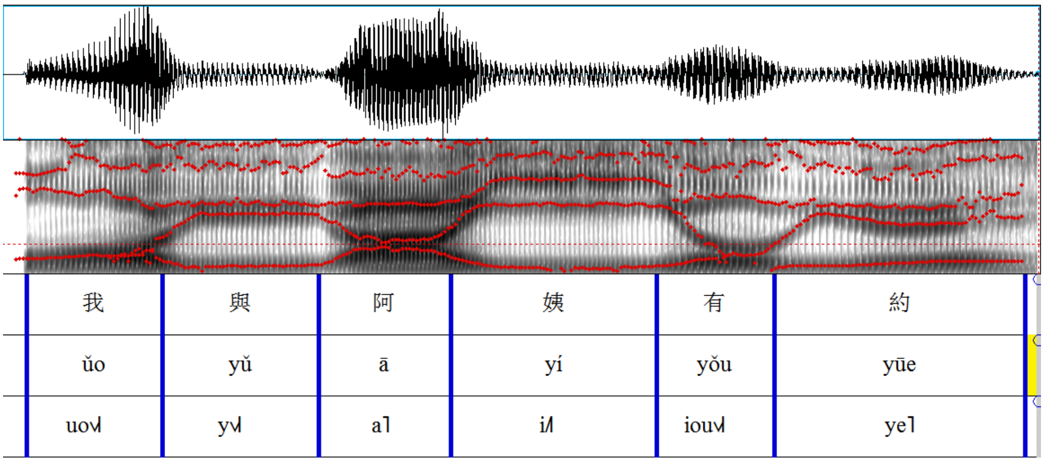

Figure 94: *Spectrogram of stimulus utterance with Formant shift ratio 1*

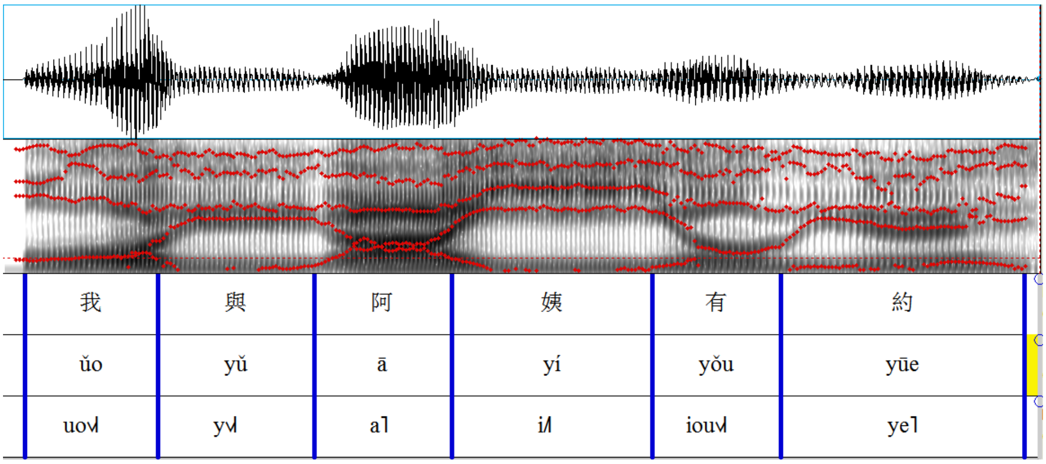

Figure 95: *Spectrogram of stimulus utterance with formant shift ratio 0.9*

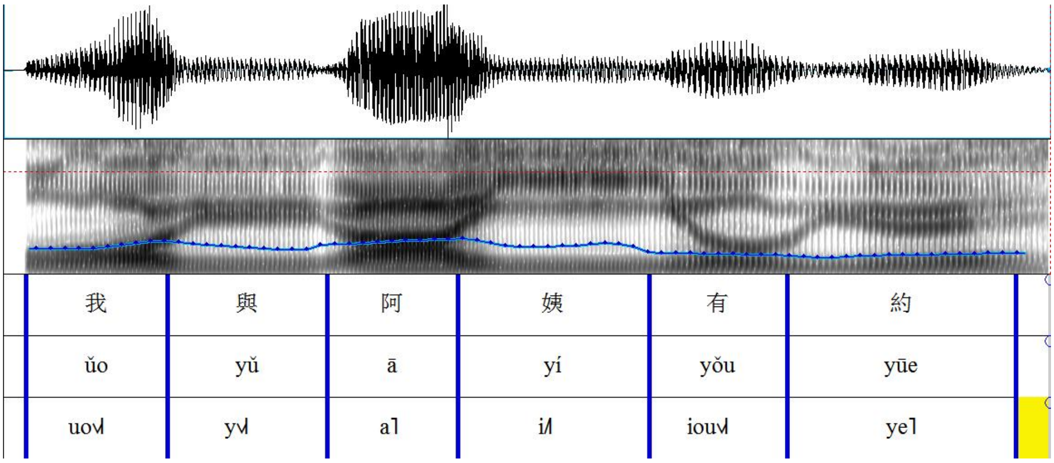

Figure 96: *Spectrogram of stimulus utterance with pitch shift +2 ST*

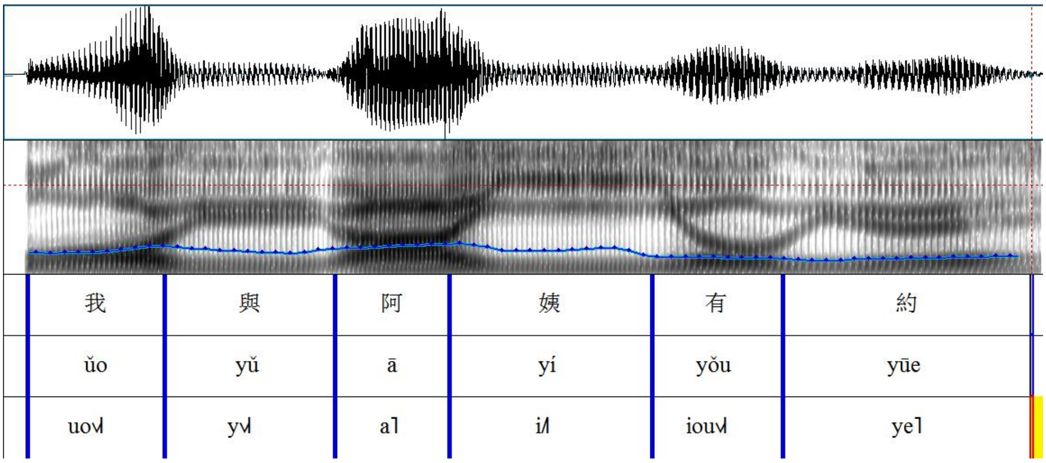

Figure 97: *Spectrogram of stimulus utterance with pitch shift 0 ST*

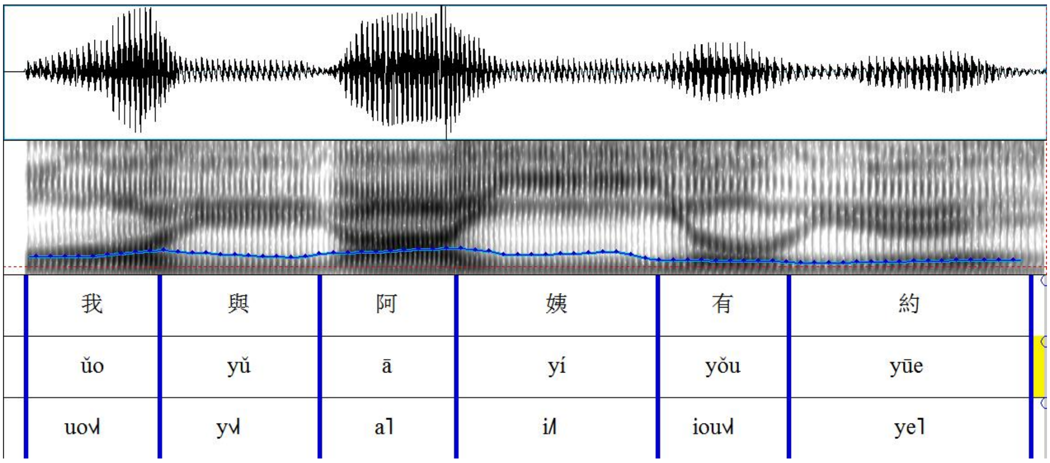

Figure 98: *Spectrogram of stimulus utterance with pitch shift -2 ST*

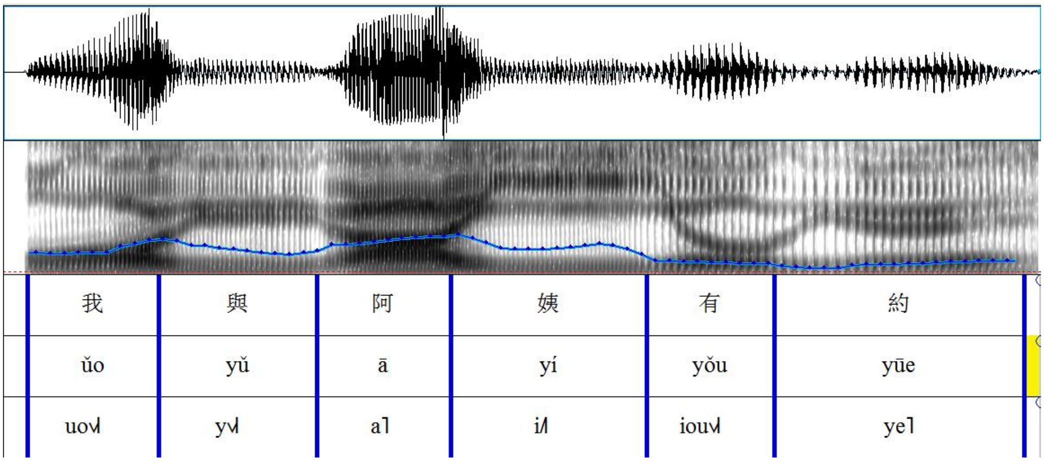

Figure 99: *Spectrogram of stimulus utterance with pitch range 2*

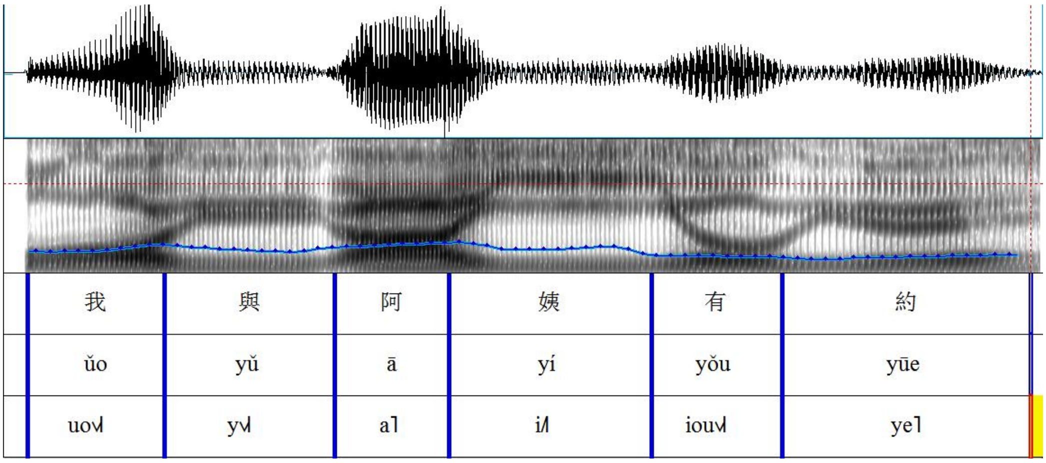

Figure 100: *Spectrogram of stimulus utterance with pitch range 1*

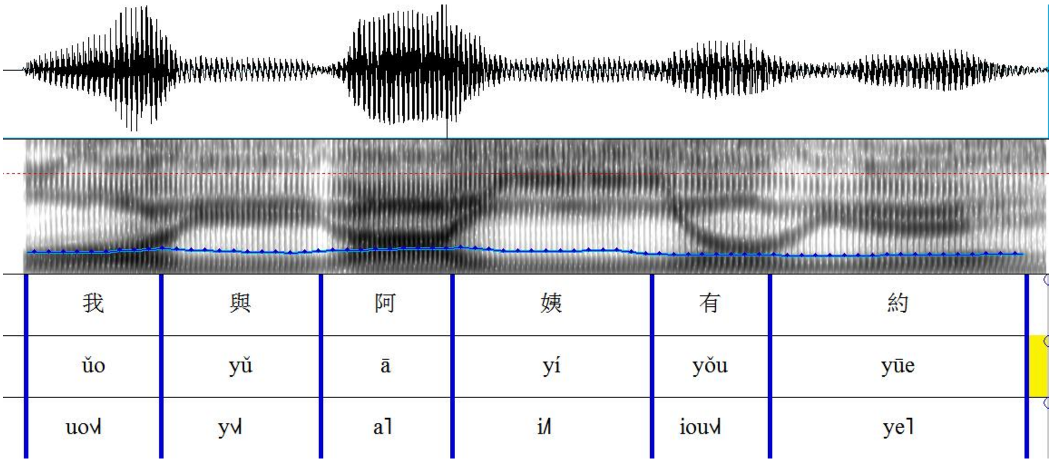

Figure 101: *Spectrogram of stimulus utterance with pitch range 0.5*