Speech Communication 162 (2024) 103101

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

The effects of informational and energetic/modulation masking on the efficiency and ease of speech communication across the lifespan

Outi Tuomainen[a,b,*], Stuart Rosen[b], Linda Taschenberger[b], Valerie Hazan[b]

[a] Department of linguistics, University of Potsdam, Germany
[b] UCL, Speech Hearing and Phonetic Sciences, Germany

ARTICLE INFO

Keywords:
Communication efficiency
Communication ease
Interactive speech
Informational and energetic/modulation masking
Lifespan

ABSTRACT

Children and older adults have greater difficulty understanding speech when there are other voices in the background (informational masking, IM) than when the interference is a steady-state noise with a similar spectral profile but is not speech (due to modulation and energetic masking; EM/MM). We evaluated whether this IM vs. EM/MM difference for certain age ranges was found for broader measures of communication efficiency and ease in 114 participants aged between 8 and 80. Participants carried out interactive *diapix* problem-solving tasks in age-band- and sex-matched pairs, in quiet and with different maskers in the background affecting both participants. Three measures were taken: (a) task transaction time (communication efficiency), (b) performance on a secondary auditory task simultaneously carried out during *diapix*, and (c) post-test subjective ratings of effort, concentration, difficulty and noisiness (communication ease). Although participants did not take longer to complete the task when in challenging conditions, effects of IM vs. EM/MM were clearly seen on the other measures. Relative to the EM/MM and quiet conditions, participants in IM conditions were less able to attend to the secondary task and reported greater effects of the masker type on their perceived degree of effort, concentration, difficulty and noisiness. However, we found no evidence of decreased communication efficiency and ease in IM relative to EM/MM for children and older adults in any of our measures. The clearest effects of age were observed in transaction time and secondary task measures. Overall, communication efficiency gradually improved between the ages 8–18 years and performance on the secondary task improved over younger ages (until 30 years) and gradually decreased after 50 years of age. Finally, we also found an impact of communicative role on performance. In adults, the participant asked to take the lead in the task and who spoke the most, performed worse on the secondary task than the person who was mainly in a 'listening' role and responding to queries. These results suggest that when a broader evaluation of speech communication is carried out that more closely resembles typical communicative situations, the more acute effects of IM typically seen in populations at the extremes of the lifespan are minimised potentially due to the presence of multiple information sources, which allow the use of varying communication strategies. Such a finding is relevant for clinical evaluations of speech communication.

1. Introduction

There has been extensive research in recent years on the impact of noise or other types of interference on speech communication (for reviews, see Bronkhorst, 2000, 2015; Mattys et al., 2012). The degree of communication difficulty experienced in these conditions is dependent on a wide range of factors: the type of background interference, the content of the message that is transmitted and a number of characteristics linked to the individual (e.g., age, language background, hearing status, language status). In studies on the impact of background noise on speech communication, most attention has been paid to the impact on the listener, as assessed via objective or subjective speech perception tests. In other studies, the impact of background noise on the acoustic characteristics of speech production has been evaluated in studies of 'Lombard speech'. However, some recent studies have taken a broader view of the communication process and have evaluated speech communication in an interactive setting in order to evaluate the impact of background noise on both speaking and listening within a realistic communicative situation (e.g., Beechey et al., 2020; Hazan et al., 2018; Tuomainen et al., 2022). In this study, we expand on this work by

* Corresponding author at: Department of linguistics, University of Potsdam, Germany.
E-mail address: tuomainen@uni-potsdam.de (O. Tuomainen).

https://doi.org/10.1016/j.specom.2024.103101
Received 8 December 2023; Received in revised form 11 June 2024; Accepted 26 June 2024
Available online 27 June 2024

investigating the impact of different types of background interference on a number of measures reflecting the efficiency and ease of speech communication in talkers with ages ranging across the lifespan (8 to 80 years).

A strong focus of research on speech communication in challenging conditions is the relative impact of different types of background interference on individuals from different age groups. All wide-band background noises exert masking that arises from spectro-temporal overlap in the auditory periphery, either from direct interactions of energy in the target and masker (i.e., energetic masking [EM]; see Moore, 2012) or disruption of the information-carrying amplitude modulations in the target by the amplitude modulations in the masker (i.e., modulation masking [MM]; Stone et al., 2012). However, some signals in the background, not necessarily requiring spectro-temporal interaction, perceptually interfere with the target sound higher up in the auditory pathway if they contain similar information to the target. This is defined as informational masking (IM) and is the case, for example, when communicating while one or more voices can be heard in the background (e.g., Brungart, 2001; Brungart et al., 2006; Carhart et al., 1969; Shinn-Cunningham, 2008). It is a difficult task to disentangle which perceptual difficulties are due to the energetic/modulation and informational components of the maskers (e.g., see Durlach et al., 2003). When examining signal-to-noise thresholds for a given speech intelligibility level for target speech in the presence of EM or MM maskers, or maskers which are made up of a small number of voices, more favourable signal-to-noise ratios are obtained for EM/MM than IM. This suggests that the informational component of the masker has a strong impact on intelligibility. Furthermore, within informational maskers, the number of speakers heard in the background and the match or mismatch in sex between target and masker voices has been shown to be influential, with greater interference occurring when there are fewer (e.g., 2–4) speakers in the background, at least for target and masker voices of the same sex (Rosen et al., 2013). There is also increasing evidence that listening in the presence of informational masking imposes greater cognitive load and listener effort than listening in the presence of purely EM/MM masking as shown via both behavioural and physiological measures (Brungart et al., 2013; Woodcock et al., 2019).

According to the Effortful Listening Framework (FUEL), the term ‘listening effort’ refers to the level of mental workload or strain listeners experience when they are listening to speech (Pichora-Fuller et al., 2016). Effort within an individual, in turn, is modulated by several listener-external and -internal factors including the acoustic and linguistic clarity of the incoming speech, background noise type and level and the individual sensory and cognitive abilities of the listener (e.g., hearing acuity and working memory capacity). The FUEL framework therefore emphasizes that listening effort is dynamic and highly context-dependent. For example, when trying to understand speech in multi-talker environments, the listener needs to employ additional cognitive resources to focus on the desired speaker while ignoring others, and variations in working memory and attentional capacities affect how much effort is needed. According to FUEL, the outcomes of increasing listening effort can be both positive and negative: on the one hand it can lead to successful comprehension and retention of information, but on the other it can also lead to mental fatigue, disengagement and miscomprehensions. However, the drawback of the FUEL account is the fact that there is no general agreement on how to best measure listening effort, and both objective (e.g., heart rate, pupil dilation) and subjective measures (e.g., ratings of mental demand, level of exerted effort, frustration) have been used with often inconsistent intercorrelations between these different measures of effort (e.g., Koelewijn et. al., 2015; Speechmatics, 2018; Strand et al., 2018).

Relative to young adults, it has long been suggested that IM has a greater impact on speech perception for children (for reviews, see Erickson and Newman, 2017; Leibold and Buss, 2019) and older adults, even for adults who do not present with peripheral hearing difficulties (Helfer and Freyman, 2008; Rajan and Cainer, 2008; Schoof and Rosen, 2014; Tun and Wingfield, 1999). More recently, lifespan studies, which use the same test materials and procedure across a wide age range, have started to throw greater light on the time course of these changes. Sobon et al. (2019), in a study involving children, young and older adults, showed poorer speech perception thresholds for children and older adults, especially for a two-talker speech masker (IM) relative to a spectrally-matched EM/MM. In a lifespan study with children, adolescents, young adults and older adults (but excluding 30–60 year olds), Buss et al. (2019) compared the impact of speech-shaped noise (EM/MM) and a two-talker masker (IM) on the intelligibility of semantically predictable and anomalous sentences. They obtained better thresholds for young adults than for the other groups, with larger age effects for IM than EM/MM. They concluded that while there was greater interference of IM for children and older adults than for young adults, this was likely due to different sensory and cognitive factors in each age group. With an age range spanning adulthood only, Goossens et al. (2017) found an age-related decline in middle-aged and older-aged adults with typical hearing thresholds (up to 4 kHz) relative to young adults, with a greater decline for IM than EM/MM. This was despite the lack of an age effect for the short version of the Speech Spatial and Qualities of Hearing questionnaire (SSQ, Noble et al., 2013) which reports speech perception in everyday life.

The studies mentioned so far only investigate the impact of different types of maskers on speech intelligibility. However, it is also relevant for communication to consider the impact of IM and EM/MM on speech *production*. Studies of Lombard speech (Lombard, 1911), the type of speech produced in background noise, have shown changes to acoustic characteristics such as speech intensity, pitch and speaking rate when speaking in the presence of EM/MM maskers such as multi-talker babble or speech-shaped noise (see e.g., Lu and Cooke, 2008). Lombard speech is typically assessed in tasks that require participants to read sentences or repeat heard sentences. Acoustic adaptations in Lombard speech have been shown to vary between conditions involving EM/MM and IM (Meekings et al., 2016).

However, typical communicative situations are far removed from purely perceptual tasks or tasks that involve reading sentences. First, they involve an individual as both talker and listener, and the presence of social interaction has been shown to lead to changes in speech adaptations (Cooke and Lu, 2010; Hazan et al., 2018) and in neural activation in the brain (Rice and Redcay, 2016). Second, the aim of much communication is to exchange information and in most cases, the message to be decoded is typically in the context of a known situation and prior context. As such, even limited perception of the acoustic signal may be sufficient to understand the message, suggesting that typical intelligibility tests which estimate the percentage of keywords or sentences correctly perceived may seriously underestimate how successful natural communication may be at similar signal-to-noise levels (see e.g., Petersen, 2024). Thirdly, speech that is produced in an interaction is shaped by the process of interaction itself. In situations where communication becomes difficult, talkers naturally adapt their speaking style to be more easily understood with adaptations made not only at the speech acoustic level (for reviews, see Smiljanic and Bradlow, 2009; Mattys et al., 2012) but also in terms of the lexical and semantic context and syntactic structures used (Genovese et al., 2020; Harmon et al., 2021). Finally, a large proportion of our everyday interactions take place when we are doing something else at the same time (driving a car, navigating a route, using media) resulting in an increase in cognitive load relative to most laboratory-based experiments where participants are asked to perform one task at a time (speaking, listening). Speech communication is therefore highly adaptive and dynamic and talkers continuously monitor their interlocutors in order to meet their informational needs while investing the appropriate amount of articulatory effort (i.e., not too little, not too much) to maintain effective communication as suggested by Lindblom’s Hyper-Hypo Model of Speech Production (Lindblom, 1990).

In recent years, there has therefore been a move towards evaluating speech communication in challenging conditions in more ecologically-valid tasks that involve communicative exchanges between pairs or groups of talkers. One such approach is *diapix*, a picture-based 'spot the difference' problem-solving task (Van Engen et al., 2010; Baker and Hazan, 2011); other problem-solving tasks such as Sudoku (Lu and Cooke, 2008; 2010; Aubanel and Cooke, 2013) and tangram-based puzzles (Beechey et al., 2019) have also been used. In *diapix,* analyses typically evaluate the change in task transaction time or in speech characteristics that occur when pairs of participants communicate in the presence of different types of interference relative to when they can hear each other normally (e.g., Hazan and Baker 2011; Van Engen et al., 2010).

In a study that focused more on the relative effects of EM/MM and IM on speech production within a communicative task, Cooke and Lu (2010) showed that, when completing a Sudoku puzzle task in the presence of a competing-speaker masker, participants showed adaptations consistent with Lombard speech, although changes in energy, mean F0 and spectral tilt were smaller for IM (competing speaker) than for EM/MM (speech-shaped noise). In comparing speech production when the Sudoku game was done alone and together with a partner, they also showed that greater acoustic adaptations were made to speech in the collaborative task. Aligning with the acoustic-phonetic results of Cooke and Lu (2010), Pham & Karuza (2023), using a picture description task via Zoom, showed that in background noise participants adapted the syntactic complexity of their speech both for their own benefit as well as to facilitate comprehension in the listener. Together these findings reinforce the notion that tasks involving communicative intent produce speech that differs from that produced without communicative intent (e.g., Garnier et al., 2010; Hazan and Baker, 2011).

Even though IM has been shown to have a greater impact on speech intelligibility in children and older adults, it is unclear whether this is the case in a broader communicative situation. In this way, we build on and extend the findings of Cooke and Lu (2010) who only analysed the speech of young adults. Furthermore, evaluating interlocutors' performance in these interactive communicative tasks is often more challenging than measuring performance in purely perceptual tasks using various accuracy-based intelligibility metrics. In most cases interactive performance has been evaluated via both objective and subjective measures. As mentioned above, objective assessments have often included measures of task transaction time (e.g., how long it took to finish the task; e.g., Hazan and Baker, 2011; Hazan et al., 2018), acoustic-phonetic characteristics of speech (e.g., speech rate, f0, intensity; e.g., Beechey et al., 2018; Hazan et al., 2018; Tuomainen et al., 2022) and turn-taking behaviour (e.g., durations of silent interval between two speech turns; e.g., Sorensen et al., 2021). The more subjective measures have usually consisted of questionnaires and ratings that assess talkers' subjective experiences of the task, communicative setting and interaction (e.g., evaluations of difficulty, engagement and motivation; e.g., Beechey et al., 2019). These metrics, however, are rarely acquired in studies across the lifespan, making it difficult to assess how interactive speech communication develops as a function of age from childhood to older age.

Here we expand this work and measure efficiency and ease of speech communication in different types of background noise in pairs of talkers aged 8–80 years. In a related paper, acoustic measures of speech productions on these same participants and tasks as in the current study showed that when conversing in background noise, older adults (at 50+ years) increase vocal effort regardless of the noise type (IM vs EM/MM) whereas children (8–13 years) increase vocal effort only in IM (Tuomainen et al., 2022). However, albeit involving interactive speech by pairs of interlocutors, Tuomainen et al. (2022) focused mostly on acoustic-phonetic adaptations made by the "lead talker" of the pair (i.e., the person who was instructed to speak more).

In summary, the aim of this study was to investigate the relative impact of EM/MM and IM in an ecologically-valid task involving interactions between pairs of participants covering a wide age range from children to older adults. More specifically, we investigated whether IM had a greater impact on communication efficiency and ease and whether this was modulated by talker age. Our hypothesis, based on the previous literature, was that a greater impact of IM than EM/MM on both communication efficiency and ease would be seen in communicative interactions involving young children or older adults. However, we also expected that the difference between IM and EM/MM would be less marked than has previously been shown in the literature for speech intelligibility tests. We also investigated whether being in primarily talker or listener role had an impact on measures of communication ease; we expected that the participant taking the more active role would show greater effects due to increased processing load.

2. Method

2.1. Participants

A total of 114 monolingual native speakers of Standard Southern British English aged between 8 and 80 participated in the study. This age range was chosen because our previous studies had shown that *diapix* could successfully be used without any alterations of the picture sets with participants at each extreme of this age range (Hazan et al., 2016; 2018).

Participants were recruited within age bands, to ensure a balance of participants across age, even though all analyses were conducted on age as a continuous variable. These age bands were: 8–12 years (M=10.34), 13–17 years (M=15.94), 18–29 years (M=21.82), 30–49 years (M=42.98), 50–64 years (M=59.30) and 65–80 years (M=71.19). Age-ranges within each band are based on our previous studies with children (Hazan et al., 2016) and later adulthood (Hazan et al., 2018) and capture age-related changes in speech production. Each of the six age bands included 20 participants (10F) apart from the 13–17 band due to recruitment difficulties (N=14, with only 4 males).

Participants were tested in sex- and age-band-matched pairs and were not known to one another. In the two youngest and oldest groups, pairs were matched by age as closely as possible (e.g., an 8-year old was never paired up with a 12-year-old). All participants were tested for hearing thresholds using pure tone audiometry and were classified as normal hearing up to 4 kHz, achieving a better ear average of <25 dB HL across the 0.25–4 kHz octave frequencies. Some participants in the older age range had higher thresholds at 4 and 8 kHz, albeit still within the limits of mild hearing loss (≤ 41 dB). Participants reported no history of speech and language impairments or neurological trauma (assessed with an in-house questionnaire). All participants aged over 65 passed the Montreal Cognitive Assessment screening test (MoCA; Nasreddine et al., 2005).

Ethical approval was obtained from the University College London (UCL) Research Ethics Committee, and informed written consent was obtained from each participant and a parent/guardian of each child.

2.2. Diapix task

The communicative task used to obtain spontaneous speech interactions between two conversational partners was *diapix* (van Engen et al., 2010), an interactive problem-solving 'spot the difference' picture task. The two participants doing the task sat in adjacent acoustically-shielded rooms and communicated via headsets fitted with a cardioid microphone (Beyerdynamic DT297) whilst playing *diapix* on a desktop PC using the DiapixUK picture sets (Baker and Hazan, 2011). The participants could not see each other. Participants were each given a different version of the same picture scene (beach, farm or street; see Baker and Hazan, 2011 for details) and had 10 min to find the 12 differences between the pictures. A total of eight different *diapix* picture pairs were used and the pictures were randomised across participants within each age band. *Diapix* was carried out in four listening conditions affecting both participants: i) in quiet (QUIET), eliciting unmarked conversational speech, ii) EM/MM with no informational content (speech-shaped noise, SPSN), iii) IM that is semantically related to the picture task (IMRE; three voices talking about the same *diapix* picture), and iv) IM that is semantically unrelated to the task (IMUR; three voices talking about a different *diapix* picture). Two different IM conditions were used to evaluate the degree to which the semantic content of the masker would affect the degree to which it would interfere with communication. We expected that hearing the same picture being discussed by the voices in the background (IMRE condition) would cause greater interference than hearing an unrelated picture being discussed (IMUR condition).

Both IMRE and IMUR were three-talker maskers consisting of a male, a female and a child talker. The same three talkers were used for all listeners. The three talkers were reading pre-scripted descriptions of each picture pair that were based on conversations in the previous *diapix* corpora (as reported in Hazan et al., 2016; 2018). To create the maskers, silences, speech errors and hesitations were edited out of the recordings and the order in which the picture was discussed was changed. For the SPSN condition, a speech-shaped noise file was created from the long-term spectrum of the recording from all voices included in the IM masker.

To give a more natural listening environment in terms of the space and background noise levels, we used Spatial Audio Simulation System software (Audio 3D; available at https://www.phon.ucl.ac.uk/resource/audio3d/) that mimics real room acoustics (room dimensions: width 4 m, length 3 m, height 2.5 m) combined with head-related transfer functions in real-time (height of talker: 1.25 m). The maskers (three voices for IMUR and IMRE, and three matched speech-shaped noise signals for SPSN) and the voice of the interlocutor were delivered via headphones and spatially separated by 1 meter from both each other and the "live" talker via the Audio 3D simulator. The intensity of the three maskers was normalised to 72 dB SPL. The intensity level of the talkers was set to approximate a signal-to-noise ratio of 0 dB when speaking normally. Full details of the configuration files are available on github (https://github.com/outepi/Diapix-virtual-room).

The picture and noise condition orders were randomised (QUIET was always presented first and participants were informed about whether they were about to hear an EM/MM or IM condition). Participants were recorded on separate audio channels at a 44.1 kHz (16 bit) sampling rate using a Fireface audio interface and Audacity audio software. One of the participants was randomly assigned to lead the interactions ('Talker A'). The other ('Talker B') was a more passive participant, mainly responding to queries by Talker A. All participants completed the four test conditions in both talker roles with the same conversational partner, giving us an assessment, within a communicative context, of a participant's communicative efficacy when in Talker A role and their perceptual ability in Talker B role. Participants were not given any instructions as to what strategies they should use or which speaking style they should adopt. No prior exposure to the maskers was given before the start of the recording.

To assess whether communication in background noise was affected by differences in cognitive control, a secondary task was carried out by both participants during all conversations. This secondary task informs about distractibility and the effects of competing cognitive resources across the lifespan. At random intervals, either participant heard one out of two possible auditory cues (dog bark or car horn) which were edited to have the same duration (200 ms) and intensity (33 dB in quiet and + 8 dB SNR in noise backgrounds). These cues were incorporated into the audio masker files to appear at quasi-randomly assigned intervals which were fixed across participants but varied across the two roles (A,B) within the dyad and across the different listening conditions. Participants had to react by either pressing a bell (dog bark) or inhibiting a response (car horn). Participants were instructed to respond as quickly as possible and were reminded of this secondary task before every condition. The number of cues presented was dependent on the duration of the task (e.g., in a 5 min sequence, there were 6 ignore trials and 9 attend trials). Full information about the method, with example masker files and audio3D configuration files can be found on Github (https://github.com/outepi/Diapix-virtual-room).

After completing each *diapix* task, both participants completed a paper-based questionnaire, answering four questions using an 11 point Likert scale, concerning four aspects of their experience: 1. *Concentration*: 'Did you have to concentrate very hard to understand your partner?' (0-concentrate hard 10-no need to concentrate) 2. *Effort*: 'Did you have to put in a lot of effort to understand your partner?' (0-lots of effort 10-no effort) 3. *Noisiness*: 'On average, how noisy did you experience the background noise you heard during the task?' (0-quiet 10-very noisy) 4. *Interference*: 'Could you easily ignore the background noise?' (0-not easily ignore 10-easily ignore). The scale was ordered in the same way as those sections assessing effort and concentration (Section 3) in the SSQ questionnaire (Gatehouse and Noble, 2004) which they had previously completed. Questions 1, 2 and 4 were adapted from the SSQ questionnaire ('qualities of hearing' -section) to match the current interactive setting. Question 3 was added as a broader measure of subjective perception of "noisiness", as the concept of "noisy" is easy to understand across the lifespan but at the same time it can be influenced by individual perceptual sensitivities and age (e.g., Miedema and Vos, 1999). In the ratings, participants could place a cross at any point along the scale. Note that for 3 of the 4 scales, lower values indicate greater degrees of *effort, concentration* and *interference*, but for *noisiness*, the scale ran in the other direction. Participants were not given questions 3 and 4 in QUIET.

2.3. Procedure

The participant pairs took part in two test sessions, either carried out on different days or within the same day separated by a lengthy break (e.g., first session in the morning and the second session in the afternoon). At the first session, they completed a series of background sensory and cognitive tasks individually (not reported here) and received training in *diapix* and in the secondary task. The training started with the hardcopy version of the *diapix* task with both participants in the same room. After approximately 3–5 min of practise face-to-face, participants continued the training on a PC computer now in separate booths and communicating via headsets. For this second training session they were also familiarised with the secondary task (in quiet background). This training continued for approximately 3–4 min until participants were performing near-ceiling in the secondary task (for both attend and ignore trials). In the second session, participants took part in four *diapix* tasks, in both Talker A and Talker B roles (i.e., 8 *diapix* tasks in total). Each *diapix* task was immediately followed by a rating task. Even though this study involved a broad age range, no adjustments to the procedure were needed. For the post-*diapix* ratings, the experimenter read the questions to the child participants one by one and reminded them of the meaning of each end of the scale. Adult participants carried out the ratings alone but these were checked by the experimenter who looked at the ratings and would check with participants if they thought they had read the scale wrongly (e.g., giving a 'lots of effort' rating for the QUIET condition). This sometimes resulted in a participant changing their rating). The complete data collection across the two sessions took approximately 2.5 h.

2.4. Data processing

All recordings were automatically transcribed using a cloud-based transcription system (Speechmatics, 2018) and then manually corrected for word-level errors and audio-transcription misalignments in Praat (Boersma and Weenink, 2018). Bell presses were manually labelled in the transcriptions. From the audio recordings, we calculated measures that reflect communication efficiency and cognitive control. Scores for perceived concentration, effort, noisiness and interference were manually entered by the participant after each transaction. While the communication efficiency measure reflects the joint performance of the dyad, the cognitive control and ratings measures were calculated separately for each participant in Talker A and Talker B roles.

The communication efficiency measure ($time_1$) was the average time (in seconds) it took the pair of talkers to find a difference in *diapix* (i.e., total file duration divided by number of differences found). This reflects how effective the talkers were in communicating with their conversational partner, as frequent requests for repetitions or misunderstandings would lead to more time needed to complete the task. Assessment of communication ease consisted of both the cognitive control measures and subjective ratings. The cognitive control measure was performance on the secondary task, a discriminability (d') measure in Talker A and Talker B roles. For the post-*diapix* ratings, means were obtained for each listening condition and in each role (Talker A, Talker B).

2.5. Data analysis and results

First, to verify that Talker A was the dominant talker in the interaction, we calculated the number of words produced by Talker A and Talker B in each dialogue, after exclusion of segments labelled as unfinished words, hesitations, fillers, and agreements. The mean number of words produced by Talker A was 801 words and by Talker B was 338 words; also, Talker A produced more words than Talker B in 97.4 % of all individual dialogues. It is clear, therefore, that Talker A dominated the interaction, as intended.

Communication efficiency ($time_1$) was analysed with Multivariate Adaptive Regression Splines (MARS) using the R package 'earth' (Milborrow, 2011). Here, MARS performs a segmented regression over *age*, determining the minimum number of separate line segments that are necessary to describe the data, as well as whether the predictor of listening *condition* (QUIET, IMRE, IMUR, SPSN) improves the model. For the secondary task and the ratings, linear mixed-effects regressions were used in the main analyses investigating the effects of *age*, listening *condition* and *role* (Talker A/B). Age was treated as a continuous predictor even though recruitment was performed through age bands. Starting from saturated models with participant as random intercepts and all main effects and interactions of the fixed effects as predictors, terms were removed iteratively if their removal did not significantly worsen the fit of the model, resulting in a so-called minimal model. Term elimination started with the highest-order interactions and no term was eliminated if it was included in a higher-order interaction. All calculations used the R packages 'lme4' (Bates et al., 2015) and 'lmerTest' (Kuznetsova et al., 2017). Data and code for statistical analyses in R are available at Open Science Framework: https://osf.io/tkve8/.

2.5.1. Communication efficiency

Fig. 1 shows the measure of communication efficiency ($time_1$) as a function of age and condition. As the form of the data suggested linear segments with clear breakpoints, we used segmented regression (MARS)

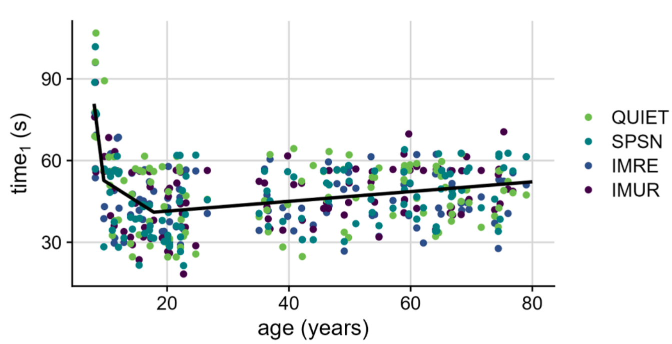

Fig. 1. Best fitting Multivariate Adaptive Regression Splines (MARS) for the average time (in seconds) taken to find one difference in diapix. Individual coloured dots represent the four different listening conditions.

over age and investigated whether the predictor of condition improves the model. As the best-fitting MARS line in Fig. 1 shows, there was no significant effect of listening condition.

Overall, the dyads found the differences as fast in quiet as in background noise conditions. However, as seen in the regression line segments in Fig. 1, communication efficiency changed as a function of age: performance improves most quickly from age 8 to 10 years (by 17.9 s per year), it then continues to improve up to age 18 years, but at a slower rate (1.4 s per year). Finally, performance worsens from 18 to age 80 years, but at a very slow rate (0.18 s per year). This is only a change of 10.8 s over the 60-year age range from 20 to 80 years, as opposed to the 26.4 second improvement in the children from 8 to 10 years.

2.5.2. Communication ease: secondary task

For each combination of participant, condition and role, a 2 × 2 response matrix was constructed tallying up the number of hits, false alarms, correct rejections and misses. From these matrices, d' values were calculated using the method of Hautus (1995). There were six values of d' below −1, most of which were significantly different from chance (using Fisher's exact test on the response matrix). Negative d' values indicate a consistent mislabelling of responses (pressing a bell for the car horn instead of the dog bark), so these values were eliminated, leaving 906 valid cases.

Our primary interest in this task is the d' value exhibited by participants as a function of age, condition and role. An initial inspection of the trends in the data across age revealed a clear quadratic component, so the saturated model included a quadratic term for age along with the predictors of (linear) age, condition and role. After term elimination, the resulting minimal model (see Table S1 in supplementary materials) included highly significant main effects of age, age^2, condition and role, along with significant interactions of role x age and role x age^2 (all p's<0.005).

Fig. 2 shows the predictions of the minimal model along with the observed data. Overall, the results showed that age had a significant impact on the secondary task performance. It can be clearly seen in Fig. 2 that the quadratic term in age arises because performance improves over younger ages, reaches a peak at about 30–50 years, and then gradually decreases thereafter. The interaction of role with the quadratic term in age reflects the fact that role has little effect on performance for the children but it has an effect for adults aged 20–70 years. Note that the effect of role disappears after the age 70 years, so it only applies to a small number of the oldest participants. As expected, the Talker B role leads to superior performance when role has an effect.

Furthermore, listening condition significantly affected performance in the secondary task (see supplementary materials, Fig. S2). As

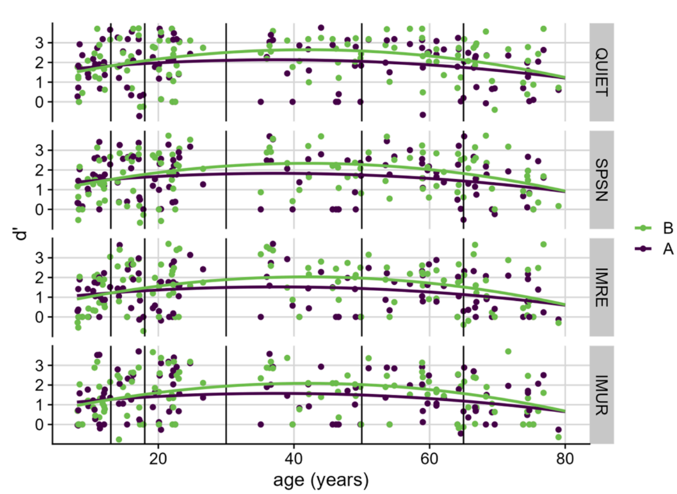

Fig. 2. The d' values in the secondary bell press task as a function of age and role (Talker A,B). The four listening conditions are displayed in separate panels.

expected again, performance is best in QUIET, worst in IMRE and IMUR, and intermediate in SPSN (all comparisons $p < 0.001$). Performance was similar in the two IM conditions (IMRE, IMUR; $p \cong 0.4$). However, against our predictions, the effect of listening condition was not modulated by age.

2.5.3. Communication ease: post-diapix subjective ratings

Unsurprisingly, the 4 rating scales were moderately to highly correlated with one another (with the magnitude of correlation ranging from 0.40 to 0.87), and the highest correlations were observed between ratings of *effort* and *concentration*. Therefore, we executed a Principal Components Analysis (PCA) to extract components which best accounted for the variability in the ratings using the R package 'FactoMineR' (Lê et al., 2008). Because ratings were not gathered for every listening condition (i.e., *noisiness* and *interference* were not rated in QUIET), analyses were done separately for two sets of ratings, with a main focus on the two scales that were available for all four listening conditions (*effort* and *concentration*).

Given the high correlation between *effort* and *concentration*, it is only reasonable to extract a single dimension, which is equivalent to a mean of the two scales after converting each to a standard score. This dimension accounted for ~93 % of the variance. Similar to the d' measure, an initial inspection of the trends in the rating data across age revealed a clear quadratic component, so the saturated model included a quadratic term for age along with the predictors of (linear) age, condition and role. After term elimination, the resulting minimal model (see Table S3 in supplementary materials) included highly significant main effects of condition and role, and moderately significant main effects of age and age^2. The only significant interactions were of condition x age and condition x age^2 (see Table S3).

Fig. 3 shows the predictions of the minimal model along with the observed data. In terms of the effects of listening condition, putting aside the extent to which the difference between conditions varies across age, it is clear in Fig. 3 that QUIET results in the lowest ratings of concentration/effort with the two IM conditions (IMRE/IMUR) being very similar and achieving the highest concentration/effort ratings. As with the d` measure of cognitive load, the non-speech masking condition (SPSN) lies between the QUIET and IMRE/IMUR conditions. The significant quadratic age term indicates a general trend for rated concentration/effort to decrease from childhood to middle age, and then increase again in older adults. Interestingly, the trends in the two IM conditions are flatter across age than in QUIET and SPSN conditions, accounting for the significant condition x age^2 interaction. Note here that it appears that IMRE and IMUR have slightly different trajectories which results in them diverging for the oldest listeners. However, it is crucial to realise that not everything that looks different in these plots is statistically different, because once an interaction of condition with age and age^2 is necessary, then the fitted lines will follow the data within each category independently, emphasising small difference in outcomes. In fact, a model in which IMRE and IMUR are collapsed into one category fits the data as well as one in which those are separate categories ($p > 0.2$).

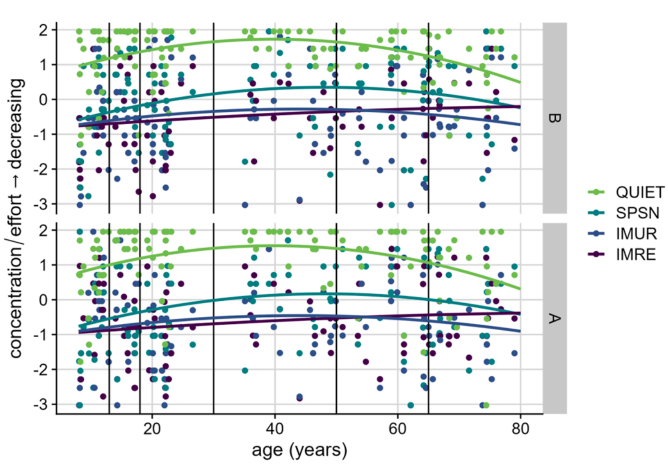

Fig. 3. A subjective measure of concentration/effort obtained from post-diapix subjective ratings as a function of age and condition, with role (Talker A,B) in separate panels. Note that higher values on the y-axis indicate lower degrees of concentration/effort.

As to the effects of role within the dyad, as expected, higher concentration/effort ratings were achieved in the more active (Talker A) role. Although consistent and highly statistically significant, this effect was relatively small (see supplementary Fig. S4). Because the extracted PCA dimension is in units that are not easy to interpret, we calculated the simple means of the concentration/effort scales, which also results in an 11-point scale as used by the participants (and which correlates with the PCA dimension at 0.999). This allows one to better interpret the size of the effect of role in comparison to the effect of condition. As Table 1 shows, the two IM conditions are rated, on average, about 0.9 of a scale point more effortful than SPSN, which is itself about 2.7 scale points more effortful than QUIET. The main effect of role, however, is only about 1/3rd of a scale point.

We turn now to analyses of the three masked conditions only, in which all 4 rating scales were available (concentration, effort, interference, noisiness). An initial PCA showed the first two extracted dimensions to account for 69 % and 16.5 % of the variance respectively. The first dimension (PCA-1) was essentially a mean of all 4 rating scales, with slightly varying weights (ranging from 0.68 - 0.91), and correlation with a simple mean of the 4 scales at 0.999. This dimension also correlated highly with the dimension extracted from just concentration and effort at 0.934. To simplify the discussion, we refer to lower scores here as being more 'difficult' whether it refers to more concentration/effort, more interference, or a judgement that the situation was 'noisier'.

A similar mixed-effects analysis for PCA-1 as was done for all conditions above was repeated here, with very similar results in terms of the difference in conditions (the two IM conditions being very similar, and SPSN being somewhat less 'difficult'), and with a minimal effect of role (although again statistically highly significant; see Table S5 and Fig. S6 in supplementary materials). The main difference in outcomes was that a quadratic term in age was not necessary, implying that trends with age in the QUIET condition in the previous analysis drove the inclusion of the quadratic term. Here, age, although significant, had only a small effect in the older participants indicating small or no linear changes in the direction of less difficulty across age.

For two reasons we do not do a full-blown analysis of the second extracted dimension (PCA-2). For one thing, it only accounted for a small proportion of the variance. Secondly, this dimension is not easy to interpret, being primarily a weighted *difference* between the ratings of noisiness, and those of concentration and effort (with little contribution from interference). What is interesting is that this measure is dominated by the noisy rating (correlated with it at 0.72), implying that judgements about noisiness seem to involve aspects of the experience that are at least somewhat different to those involved in the other 3 scales. Given that

Table 1
Mean ratings for concentration/effort averaged within conditions (top) and role (bottom) on the 11-point likert scale (higher number indicates less concentration/effort).

condition	mean	s.d.
IMRE	4.94	2.44
IMUR	4.95	2.50
SPSN	5.87	2.46
QUIET	8.55	1.80
role	**mean**	**s.d.**
A	5.90	2.80
B	6.26	2.68

this dimension is primarily a judgement about 'noisiness', it is perhaps not surprising that participant role appeared to be irrelevant. Note also that the non-speech masker condition (SPSN) is perceived as less 'noisy' than the two IM maskers, which had very similar ratings.

In summary, subjective ratings showed IM to have a greater impact on communication ease than EM/MM. The participant who was the more active talker experienced a (somewhat) greater effect than the more passive talker, although not for judgements related to noisiness. There was a general trend for the youngest listeners to rate the task generally as more 'difficult'.

3. Discussion

The main objective of this study was to evaluate whether the greater impact of IM (speech) relative to EM/MM (speech-shaped noise), which is typically seen in controlled intelligibility tests and to an extent in speech production tasks, remains within the context of a more ecologically-valid task involving speech communication between participant pairs. We also evaluated whether this effect was modulated by participant age given previous findings in controlled intelligibility tests, that children and older adults were more greatly affected by IM than young adults. Finally, we also evaluated if the participant's role in the interaction (active or more passive) had an impact on their overall communication ease.

In summary, several measures provided clear evidence of an additional impact of IM on ease of speech communication relative to the EM/MM condition. Although participants were not less efficient when in challenging conditions, that is, did not take longer to complete the task in background noise, effects of the two types of maskers were clearly seen across other measures. In IM conditions, relative to the EM/MM and quiet conditions, participants were less able to attend to the secondary task, reported greater effects of the masker on their perceived degree of effort/concentration, difficulty and noisiness. However, the semantic content of the informational masker (i.e., related/unrelated to the picture pair they were describing) did not have any effect on any of the measures.

There was also an impact of communicative role. In general, the person taking the lead in the task, and who therefore spoke the most, performed worse on the secondary task. However, this was only evident in the adult data and no such effect was found for children. Albeit a very small effect, the more active speaker also reported a greater impact on effort/concentration and difficulty than the person who was mainly in a 'listening' role and responding to queries. Contrary to the effects of age found in the majority of studies measuring speech intelligibility in background noise, the effects of age were less pronounced in the broader measures of communication efficiency and ease. Communication efficiency improved most quickly between ages 8–10 years and it continued improving more gradually until the age of 18 years. For communication ease, secondary task performance, reflecting cognitive load, peaked between the ages 30–50 years with poorer performance at the younger and older age ranges. Also, a (shallow) inverted u-shaped trend across age was observed in some of the subjective rating scales, with increasing effort/concentration ratings for younger and older participants. However, this effect of increased effort/concentration was mostly driven by the condition where participants could hear each other normally. In sum, we found clear effects of listening condition in many of our measures of communication ease with greater impact of IM and EM/MM on communication. However, contrary to what is typically reported in the more standard speech-in-noise perception literature, the effects of listening condition were mostly independent of age, that is, we found no evidence of decreased communication efficiency and ease in IM for children and older adults relative to EM/MM when broader measures of interaction are considered.

Studies separately investigating speech perception and speech production have shown that IM causes greater interference than EM/MM but, to our knowledge, this is one of few studies to examine the relative impact of IM and EM/MM on efficiency and ease of communication within a collaborative task. Cooke and Lu (2010) used a communicative task to examine the relative impact of IM and EM/MM but focused on the acoustic modifications made by the young adult participants rather than on changes in communication ease. Similarly, in a related study involving the same participants across the lifespan and the *diapix* task, we assessed the effects of IM and EM/MM on vocal effort (operationalised as a correlation between median pitch and intensity) in the more active speaker of the pair (Tuomainen et al., 2022). Expanding the metrics from Tuomainen et al. (2022) to broader communicative measures and to both interlocutors, our study shows a greater effect of IM than EM/MM on processing load within a communicative interaction. In addition, higher load was experienced by adults taking a more active role in the interactions. It could have been the case that in communicative situations which are rich in contextual information and where it is not necessary to perceive every keyword, this IM vs. EM/MM difference would have been greatly diminished but the effect is robust enough to be shown in a setting that is closer to natural communication.

Against our predictions, the measure of task transaction time ($time_1$) did not differentiate the listening conditions. However, transaction time was also not shown to be sensitive enough to detect differences across IM & EM/MM conditions in a Sudoku collaborative task (Aubanel and Cooke, 2013) and across different and less severe challenging conditions using *diapix* with older and younger adults (e.g., 8-talker babble, Hazan et al., 2018). Note that significant transaction time differences were found when more severe challenging conditions were used, such as a 3-channel noise vocoder or a simulation of a severe-to-profound hearing loss (e.g., Hazan and Baker, 2011; Hazan et al., 2018) likely due to increases in hesitations, disfluencies and questions from the more passive talker.

As shown by our related study on articulatory adaptations in the same participants, in less severe adverse conditions, talkers are able to overcome these adverse conditions by increasing the clarity of their speech. This was evident for the oldest groups of participants in listening conditions involving both IM and EM/MM and for the youngest group of children in IM (see Tuomainen et al., 2022). This increase in vocal effort in background noise might have led to increased intelligibility of their speech, thus improving their overall communication efficiency (*diapix* transaction time).

However, as shown by the rating scales in our current study, and in accordance with Lindblom's hyper-hypo model of speech production (Lindblom, 1990), this is at a cost to themselves in terms of increased effort and difficulty. Alternatively, it is possible that the dyads were less efficient during the first picture pair (due to learning effects and getting to know their interlocutor), which in our case was always the quiet condition, and this diminished the effect of the listening condition on how quickly they completed the task.

Together these combined results from vocal effort and communication efficiency and ease align with the predictions of the FUEL model (Pichora-Fuller et al., 2016) which states that effortful listening arises from multiple sensory, cognitive, and contextual factors. Here we showed increased self-rated listening effort in background noise across the lifespan, with greater influences for cognitively more demanding IM than for EM/MM. We did not find the expected age-related differences at the ends of the lifespan. However, this could have been, at least partially, due to the increase in vocal effort that we observed in children and older adults that would have led to increases in speech clarity. Furthermore, as the FUEL states, task importance and motivation can also influence both perceived effort and expenditure of effort. Here, we used the *diapix* task that is a simple goal-oriented and point-based task of relatively short duration ("find 12 differences in 10 min"). Both the motivation to locate differences and the knowledge of the maximum duration of each conversation might have encouraged participants to increase listening effort in order to do well in the task. In everyday settings, however, conversations rarely have such clear objectives and known durations, and an investment of listening effort is not always the best strategy as it can lead to fatigue. Therefore, we might have observed more age-related differences at the ends of the lifespan had the *diapix* tasks lasted for longer or at least for unpredictable durations.

Even though the greater effect of IM than EM/MM on communication was conclusive, the most prominent finding of this lifespan study was the lack, for most measures, of a greater impact of IM in children and older adults relative to young adults, which would have been evidenced by condition by age interactions. We expected such age effects in the performance on the secondary task if there was an increased processing load for these populations. There were no prominent effects of age interacting with listening conditions in most of our measures. In some of the measures (e.g., secondary task performance and subjective ratings of effort/concentration) we reported inverted u-shaped curves as a function of age but the performance peaks/valleys did not correspond to the age ranges previously reported in more controlled laboratory studies with regards to the effects of IM and EM/MM on performance. One potential explanation for this difference between current and previous findings is the fact that in the current study, the background noise levels were kept at fairly moderate levels (corresponding approximately to 0 dB SNR) which better corresponds to situations individuals encounter in the outside world. Most accuracy-based intelligibility metrics, however, measure performance at much lower SNRs, potentially over-emphasising any age-related differences found between children/older adults and younger adults.

Furthermore, this lack of age effect may also be ascribed to the masking conditions in our study, which differed from many 'speech-on-speech' perception studies that consider the impact of a single or two masker voices of the same or different sex as the target voice (e.g., Rosen et al., 2013). For two-talker maskers, older adults experience greater difficulty than younger adults with a relatively larger difference between age groups for different-sex maskers (Helfer and Freyman, 2008; Humes et al., 2006). While children also show poorer performance than young adults (e.g., Wightman and Kistler, 2005; McCreery et al., 2020), they appear to be able to take advantage of a mismatch in sex between target and masker (e.g., Wightman and Kistler, 2005). Our maskers included three voices (male, female, and child) so it is more difficult to predict whether the groups at the extremes of the age range would have been differentially affected by this masker. The decrease in interference that occurs when the masker voice is a different sex to the target (e.g., Brungart, 2001) does not apply here as for all participants, there would have been a masker voice similar to that of the target and two further masker voices that differed.

In our study, Audio 3D software was used to simulate normal communicative conditions, by introducing spatial separation of the three masker voices relative to the target voice as well as a simulation of room acoustics. Such spatial separation provides some release from masking effects (for a review, see Litovsky, 2012), with a greater release shown for maskers involving two or three voices than maskers involving up to ten voices (Freyman et al., 2004). Spatial separation has also been shown to reduce listener effort in purely perceptual tasks. Spatial release from masking seems to be immature until around 14 years (e.g., Freyman et al., 2001) and reduced in older adults even in the absence of hearing loss (Zobel et al., 2019). However, here, such age effects in the degree of spatial release of masking were not reflected in age effects on communication ease for the youngest and oldest age ranges.

Age effects in the differential impact of IM vs. EM/MM may also be reduced relative to those identified in laboratory experiments because the task used here was closer to natural communication, and enabled participants to use a wide range of strategies to overcome the adverse listening/speaking conditions, including prior context (Kalikow et al., 1977; Drager and Reichle, 2001). To complete the task, participants had to understand the gist of the interaction and some keywords in order to locate the differences in the pictures, but contextual information was present both linguistically and visually, with items seen in the picture. Children and older adults may have compensated for greater difficulties in the communicative situation by making greater use of such contextual information. There is evidence that older adults are more adept at using contextual information to compensate for increased difficulties in hearing in adverse conditions (Pichora-Fuller, 2008; Lash et al., 2013), with these effects more prominent in cases of less severe signal degradation such as those used here (Pichora-Fuller et al., 1995). We therefore suggest that the effects of greater IM interference that are obtained in highly-structured laboratory-based perception tests may not be representative of what occurs in natural communication.

An objective of this study was to establish whether simple measures of communication efficiency and ease could reflect the difference in interference between IM and EM/MM conditions. Measures of acoustic adaptations by talkers when communicating in adverse conditions have been shown to be successful in differentiating between adverse conditions using *diapix* (e.g., Hazan and Baker, 2011) or other problem-solving tasks such as tangram puzzles (Beechey et al., 2019) but they can be time-consuming to obtain (but see Beechey et al. (2018) for a successful use of automated acoustic analyses). Here, despite a large degree of individual variability, both the secondary task measure and subjective ratings of effort/concentration and difficulty were successful both in differentiating the effects of IM and EM/MM, but also in showing (somewhat) greater impact of interference on the more active participant in the conversational interactions.

Subjective effort ratings have been used extensively in purely listening tasks (Johnson et al., 2015; see also e.g., Pichora-Fuller et al., 2016 for a review), also to investigate the relative effects of IM and EM/MM on speech perception in different conditions of spatial location of maskers and other masker characteristics (Zekveld et al., 2014; Rennies et al., 2019). Such rating scales were also used by Beechey et al. (2019) with younger and older adults to evaluate effort in a problem-solving task with naturalistic conditions in which participants were not assigned specific roles. Our results therefore provide further evidence of the usefulness of subjective ratings in reflecting overall communication effort, measures which are quick and easy to obtain and analyse.

It had been expected that the presence of related semantic content (talking about the same *diapix* picture) in the masker would cause greater interference, especially in age groups with poorer selective attention (see e.g., Müller et al., 2008), but no statistically significant differences between the IMUR and IMRE conditions were found for any of the measures. Informal feedback from participants on completion of the task suggests that the presence of semantic content could have had different effects across participants. Although the majority of the participants reported noticing if the background talkers discussed the same/different picture to theirs, for some, hearing keywords in the masker that were related to the picture they were analysing might have had a facilitative effect and primed participants to focus on those aspects of the picture denoted by those keywords. For other participants, hearing related content in the masker could have had the expected distracting effect and have caused interference in the task. These differing strategies may have cancelled out any condition effects.

Finally, in the literature on communication in adverse conditions, much attention has been focused on the impact on speech perception, but the separate subjective ratings and secondary task measures obtained here for the more active and more passive participants in the *diapix* task enable us to get a relative measure of talker and listener effort within the same communicative interaction. The results suggest that, when more naturalistic settings are used, age-related differences in speech comprehension are less severe than what has been previously reported. Also, at least in adults, the processing load involved in planning how the problem-solving task can be resolved and in formulating questions leads to poorer performance in the secondary task. Such findings reinforce the need to consider the impact of adverse conditions within a communicative situation with the individual as both speaker and listener, as advocated by Hazan et al. (2018) and Beechey et al. (2019).

Funding

This work was supported by the United Kingdom Economic and Social Research Council (Grant No. ES/P002803/1).

CRediT authorship contribution statement

Outi Tuomainen: Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Stuart Rosen:** Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Linda Taschenberger:** Writing – review & editing, Project administration, Investigation. **Valerie Hazan:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

Valerie Hazan, Outi Tuomainen and Stuart Rosen report financial support was provided by Economic and Social Research Council (ESRC). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared a link to our data in the revised manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.specom.2024.103101.

References

Aubanel, V., Cooke, M., 2013. Strategies adopted by talkers faced with fluctuating and competing-speech maskers. J. Acoust. Soc. Am. 134 (4), 2884–2894. https://doi.org/10.1121/1.4818757.

Baker, R., Hazan, V., 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. Behav. Res. Methods 43 (3), 761–770. https://doi.org/10.3758/s13428-011-0075-y.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beechey, T., Buchholz, J.M., Keidser, G., 2018. Measuring communication difficulty through effortful speech production during conversation. Speech Commun. 100, 18–29. https://doi.org/10.1016/j.specom.2018.04.007.

Beechey, T., Buchholz, J.M., Keidser, G., 2019. Eliciting naturalistic conversations: a method for assessing communication ability, subjective experience, and the impacts of noise and hearing impairment. J. Speech Lang. Hear. Res. 62 (2), 470–484. https://doi.org/10.1044/2018_JSLHR-H-18-0107.

Beechey, T., Buchholz, J.M., Keidser, G., 2020. Hearing impairment increases communication effort during conversations in noise. J. Speech Lang. Hear. Res. 63 (1), 305–320. https://doi.org/10.1044/2019_jslhr-19-00201.

Boersma P., & Weenink D. (2018). Praat: doing phonetics by computer [Computer program]. 6.0.38 retrieved 29 March 2018 from http://www.praat.org/.

Bronkhorst, A.W., 2000. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acust. Acta Acust. 86, 117–128.

Bronkhorst, A.W., 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Atten. Percept. Psychophys. 77 (5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9.

Brungart, D.S., 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. J. Acoust. Soc. Am. 109 (3), 1101–1109. https://doi.org/10.1121/1.1345696.

Brungart, D.S., Chang, P.S., Simpson, B.D., et al., 2006. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J. Acoust. Soc. Am. 120 (6), 4007–4018. https://doi.org/10.1121/1.2363929.

Brungart, D., Iyer, N., Thompson, E., Simpson, B.D., Gordon-Salant, S., Shurman, J., Grant, K.W., 2013. Interactions between listening effort and masker type on the energetic and information masking of speech stimuli. Proc. Meet. Acoust. 19, 060146 https://doi.org/10.1121/1.4800033.

Buss, E., Hodge, S.E., Calandruccio, L., Leibold, L.J., Grose, J.H., 2019. Masked sentence recognition in children, young adults, and older adults: age-dependent effects of semantic context and masker type. Ear Hear. 40 (5), 1117–1126. https://doi.org/10.1097/AUD.0000000000000692.

Carhart, R., Tillman, T.W., Greetis, E.S., 1969. Perceptual masking in multiple sound backgrounds. J. Acoust. Soc. Am. 45 (3), 694–703.

Cooke, M., Lu, Y., 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. J. Acoust. Soc. Am. 128 (4), 2059–2069. https://doi.org/10.1121/1.3478775.

Drager, K., Reichle, J., 2001. Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: applications for AAC. J. Speech Lang. Hear. Res. 44 (5), 1052–1057. https://doi.org/10.1044/1092-4388(2001/083.

Durlach, N.I., Mason, C.R., Kidd, G., Arbogast, T.L., Colburn, H.S., Shinn-Cunningham, B.G., 2003. Note on informational masking. J. Acoust. Soc. Am. 113 (5), 2984–2987. https://doi.org/10.1121/1.1570435.

Erickson, L.C., Newman, R.S., 2017. Influences of background noise on infants and children. Curr. Dir. Psychol. Sci. 26 (5), 451–457. https://doi.org/10.1177/0963721417709087.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. 109 (5), 2112–2122. https://doi.org/10.1121/1.1354984.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2004. Effect of the number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Am. 115 (4), 2246–2256. https://doi.org/10.1121/1.1689343.

Garnier, M., Henrich, N., Dubois, D., 2010. Influence of sound immersion and communicative interaction on the Lombard effect. J. Speech Lang. Hear. Res. 53 (3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138.

Gatehouse, S., Noble, W., 2004. The speech, spatial and qualities of hearing scale (SSQ). Int. J. Audiol. 43 (2), 85–99. https://doi.org/10.1080/14992020400050014.

Genovese, G., Spinelli, M., Romero Lauro, L.J., Aureli, T., Castelletti, G., Fasolo, M., 2020. Infant-directed speech as a simplified but not simple register: a longitudinal study of lexical and syntactic features. J. Child Lang. 47 (1), 22–44. https://doi.org/10.1017/S0305000919000643.

Goossens, T., Vercammen, C., Wouters, J., van Wieringen, A., 2017. Masked speech perception across the adult lifespan: impact of age and hearing impairment. Hear. Res. 344, 109–124. https://doi.org/10.1016/j.heares.2016.11.004.

Harmon, T.G., Dromey, C., Nelson, B., Chapman, K., 2021. Effects of background noise on speech and language in young adults. J. Speech Lang. Hear. Res. 64 (4), 1104–1116. https://doi.org/10.1044/2020_JSLHR-20-00376.

Hautus, M.J., 1995. Corrections for extreme proportions and their biasing effects on estimated values of D'. Behav. Res. Methods Instrum. Comput. 27 (1), 46–51. https://doi.org/10.3758/BF03203619.

Hazan, V., Baker, R., 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. J. Acoust. Soc. Am. 130 (4), 2139–2152. https://doi.org/10.1121/1.3623753.

Hazan, V.L., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., Brungart, D., 2018. Clear speech adaptations in spontaneous speech produced by young and older adults. J. Acoust. Soc. Am. 144 (2), 1331–1346. https://doi.org/10.1121/1.5053218.

Hazan, V., Tuomainen, O., Pettinato, M., 2016. Suprasegmental characteristics of spontaneous speech produced in good and challenging communicative conditions by talkers aged 9–14 years. J. Speech Lang. Hear. Res. 59 (6), S1596–S1607. https://doi.org/10.1044/2016_jslhr-s-15-0046.

Helfer, K.S., Freyman, R.L., 2008. Aging and speech-on-speech masking. Ear Hear. 29 (1), 87–98. https://doi.org/10.1097/AUD.0b013e31815d638b.

Humes, L.E., Lee, J.H., Coughlin, M.P., 2006. Auditory measures of selective and divided attention in young and older adults using single-talker competition. J. Acoust. Soc. Am. 120 (5), 2926–2937. https://doi.org/10.1121/1.2354070.

Johnson, J., Xu, J., Cox, R., Pendergraft, P., 2015. A comparison of two methods for measuring listening effort as part of an audiologic test battery. Am. J. Audiol. 24 (3), 419–431. https://doi.org/10.1044/2015_AJA-14-0058.

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am. 61 (5), 1337–1349.

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B.G., Zekveld, A.A., Kramer, S.E., 2015. The pupil response reveals increased listening effort when it is difficult to focus attention. Hear. Res. 323, 81–90. https://doi.org/10.1016/j.heares.2015.02.004.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest package: tests in linear mixed effects models. J. Stat. Softw. 82 (13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lash, A., Rogers, C.S., Zoller, A., Wingfield, A., 2013. Expectation and entropy in spoken word recognition: effects of age and hearing acuity. Exp. Aging Res. 39 (2), 235–253. https://doi.org/10.1080/0361073X.2013.779175.

Leibold, L.J., Buss, E., 2019. Masked speech recognition in school-age children. Front. Psychol. 10, 1981. https://doi.org/10.3389/fpsyg.2019.01981.

Lê, S., Josse, J., Husson, F., 2008. FactoMineR: an R package for multivariate analysis. J. Stat. Softw. 25 (1), 1–18.

Lindblom, B., Hardcastle, W.J., Marchal, A., 1990. Explaining phonetic variation: a sketch of the H & H theory. Speech Production and Speech Modelling. Kluwer Academic Publishers, Dordrecht, pp. 403–439.

Litovsky, R.Y., 2012. Spatial release from masking. Acoust. Today 8, 18–25. https://doi.org/10.1121/1.4729575.

Lombard, E., 1911. Le signe de l'elevation de la voix ("The sign of the elevation of the voice"). Ann. Maladies Oreille, Larynx, Nez, Pharynx 37, 101–119.

Lu, Y., Cooke, M., 2008. Speech production modifications produced by competing talkers, babble, and stationary noise. J. Acoust. Soc. Am. 124 (5), 3261–3275. https://doi.org/10.1121/1.2990705.

Mattys, S.L., Davis, M.H., Bradlow, A.R., Scott, S.K., 2012. Speech recognition in adverse conditions: a review. Lang. Cogn. Process. 27 (7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006.

McCreery, R.W., Miller, M.K., Buss, E., Leibold, L.J., 2020. Cognitive and linguistic contributions to masked speech recognition in children. J. Speech Lang. Hear. Res. 63 (11), 3525–3538. https://doi.org/10.1044/2020_JSLHR-20-00030.

Meekings, S., Evans, S., Lavan, N., Boebinger, D., Krieger-Redwood, K., Cooke, M., Scott, S.K., 2016. Distinct neural systems recruited when speech production is modulated by different masking sounds. J. Acoust. Soc. Am. 140, 8–19. https://doi.org/10.1121/1.4948587.

Miedema, H., Vos, H., 1999. Demographic and attitudinal factors that modify annoyance from transportation noise. J. Acoust. Soc. Am. 105, 3336–3344.

Milborrow, S., Hastie, T., Tibshirani, R., 2011. Derived from mda:mars by. earth: Multivariate Adaptive Regression Splines. R package. http://CRAN.Rproject.org/package=earth.

Moore, B., Moore, B.C.J., 2012. Frequency selectivity, masking and the critical band. An Introduction to the Psychology of Hearing, 6th ed. Brill, Bingley, UK, pp. 67–101.

Müller, V., Brehmer, Y., Von Oertzen, T., Li, S.C., Lindenberger, U., 2008. Electrophysiological correlates of selective attention: a lifespan comparison. BMC Neurosci. 9, 18. https://doi.org/10.1186/1471-2202-9-18.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H., 2005. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53, 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x.

Noble, W., Jensen, N.S., Naylor, G., Bhullar, N., Akeroyd, M.A., 2013. A short form of the speech, spatial and qualities of hearing scale suitable for clinical use: the SSQ12. Int. J. Audiol. 52, 409–412. https://doi.org/10.3109/14992027.2013.781278.

Pham, C.T., Karuza, E.A., 2023. Noise-induced differences in the complexity of spoken language. Q. J. Exp. Psychol. 76 (7), 1609–1631. https://doi.org/10.1177/17470218221124869.

Petersen, E.B., 2024. Investigating conversational dynamics in triads: effects of noise, hearing impairment, and hearing aids. Front. Psychol. 15, 1289637 https://doi.org/10.3389/fpsyg.2024.1289637.

Pichora-Fuller, M.K., 2008. Use of supportive context by younger and older listeners: balancing bottom-up and top-down information processing. Int. J. Audiol. 47 (Suppl. 2), S144–S154. https://doi.org/10.1080/14992020802307404.

Pichora-Fuller, M.K., Kramer, S.E., Eckert, M.A., Edwards, B., Hornsby, B.W., Humes, L. E., et al., 2016. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). Ear Hear. 37 https://doi.org/10.1097/aud.0000000000000309.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M., 1995. How young and old adults listen to and remember speech in noise. J. Acoust. Soc. Am. 97, 593–608. https://doi.org/10.1121/1.412282.

Rajan, R., Cainer, K., 2008. Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. Neuroscience 154 (2), 784–795. https://doi.org/10.1016/j.neuroscience.2008.03.067.

Rennies, J., Best, V., Roverud, E., Kidd, G., 2019. Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort. Trends Hear. 23, 1–21. https://doi.org/10.1177/2331216519854597.

Rice, K., Redcay, E., 2016. Interaction matters: a perceived social partner alters the neural processing of human speech. Neuroimage 129, 480–488. https://doi.org/10.1016/j.neuroimage.2015.11.041.

Rosen, S., Souza, P., Ekelund, C., Majeed, A.A., 2013. Listening to speech in a background of other talkers: effects of talker number and noise vocoding. J. Acoust. Soc. Am. 133 (4), 2431–2443. https://doi.org/10.1121/1.4794379.

Schoof, T., Rosen, S., 2014. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. Front. Aging Neurosci. 6. https://doi.org/10.3389/fnagi.2014.00307.

Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12 (5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003 (Regul. Ed.).

Smiljanić, R., Bradlow, A.R., 2009. Speaking and hearing clearly: talker and listener factors in speaking style changes. Lang. Linguist. Compass. 3 (1), 236–264. https://doi.org/10.1111/j.1749-818x.2008.00112.x.

Sobon, K.A., Taleb, N.M., Buss, E., Grose, J.H., Calandruccio, L., 2019. Psychometric function slope for speech-in-noise and speech-in-speech: effects of development and aging. J. Acoust. Soc. Am. 145, EL284. https://doi.org/10.1121/1.5097377.

Stone, M.A., Füllgrabe, C., Moore, B.C., 2012. Notionally steady background noise acts primarily as a modulation masker of speech. J. Acoust. Soc. Am. 132, 317–326. https://doi.org/10.1121/1.4725766.

Strand, J.F., Brown, V.A., Merchant, M.B., Brown, H.E., Smith, J., 2018. Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. J. Speech Lang. Hear. Res. 61 (6), 1463–1486. https://doi.org/10.1044/2018_JSLHR-H-17-0257.

Sørensen, A.J.M., Fereczkowski, M., MacDonald, E.N., 2021. Effects of noise and second language on conversational dynamics in task dialogue. Trends Hear. 25. https://doi.org/10.1177/23312165211024482.

Speechmatics (2018). Speech-to-text conversion program. Available at: https://www.speechmatics.com/.

Tun, P.A., Wingfield, A., 1999. One voice too many: adult age differences in language processing with different types of distracting sounds. J. Gerontol. B Psychol. Sci. Soc. Sci. 54 (5), P317–P327. https://doi.org/10.1093/geronb/54B5.P317.

Tuomainen, O., Taschenberger, L., Rosen, S., Vl, H., 2022. Speech modifications in interactive speech: effects of age, sex and noise type. Philos. Trans. R. Soc. B B377. https://doi.org/10.1098/rstb.2020.0398.

Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M., Bradlow, A.R., 2010. The Wildcat Corpus of native- and foreign-accented English: communicative efficiency across conversational dyads with varying language alignment profiles. Lang. Speech 53, 510–540. https://doi.org/10.1177/0023830910372495.

Wightman, F.L., Kistler, D.J., 2005. Informational masking of speech in children: effects of ipsilateral and contralateral distracters. J. Acoust. Soc. Am. 118, 3164–3176. https://doi.org/10.1121/1.2082567.

Woodcock, J., Fazenda, B., Cox, T.J., Davies, W.J., 2019. Pupil dilation reveals changes in listening effort due to energetic and informational masking. In: Proceedings of the 23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019. Aachen, Germany. At.

Zekveld, A., Rudner, M., Kramer, S.E., Lyzenga, J., Ronnberg, J., 2014. Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masked speech. Front. Neurosci. 8, 88. https://doi.org/10.3389/fnins.2014.00088.

Zobel, B.H., Wagner, A., Sanders, L.D., Başkent, D., 2019. Spatial release from informational masking declines with age: evidence from a detection task in a virtual separation paradigm. J. Acoust. Soc. Am. 146 (1), 548–566. https://doi.org/10.1121/1.5118240.

Further reading

Audio3D virtual room software: 2024 www.phon.ucl.ac.uk/resource/audio3d/.

Bolia, R.S., Nelson, W.T., Ericson, M.A., Simpson, B.D., 2000. A speech corpus for multitalker communications research. J. Acoust. Soc. Am. 107 (2), 1065–1066. https://doi.org/10.1121/1.428288.

Brown, D.K., Cameron, S., Martin, J.S., Watson, C., Dillon, H., 2010. The North American listening in spatialized noise-sentences test (NA LiSN-S): normative data and test-retest reliability studies for adolescents and young adults. J. Am. Acad. Audiol. 21 (9), 629–641. https://doi.org/10.3766/jaaa.21.10.3.

Corbin, N.E., Bonino, A.Y., Buss, E., Leibold, L.J., 2016. Development of open-set word recognition in children: speech-shaped noise and two-talker speech maskers. Ear Hear 37 (1), 55–63. https://doi.org/10.1097/aud.0000000000000201.

Füllgrabe, C., Moore, B.C.J., Stone, M.A., 2015. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. Front. Aging Neurosci. 6, 347. https://doi.org/10.3389/fnagi.2014.00347.

Hall, J.W., Grose, J.H., Buss, E., et al., 2002. Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. Ear Hear 23 (2), 159–165. https://doi.org/10.1097/00003446-200204000-00008.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimault, N., Perrin, F., Collet, L., 2007. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech Commun 49 (11–12), 905–916. https://doi.org/10.1016/j.specom.2007.05.008.

Magimairaj, B.M., Nagaraj, N., Benafield, N.J., 2018. Children's speech perception in noise: evidence for dissociation from language and working memory. J. Speech Lang. Hear. Res. 61, 1294–1305. https://doi.org/10.1044/2018_jslhr-h-17-0312.

Seeman, S., Sims, R., 2015. Comparison of psychophysiological and dual-task measures of listening effort. J. Speech Lang. Hear. Res. 58 (6), 1781–1792. https://doi.org/10.1044/2015_JSLHR-H-14-0180.