Intelligibility of spontaneous casual speech in background noise across the lifespan

Outti Tuomainen[1,2], Amanda Soengadie[2], Linda Taschenberger[1], Valerie Hazan[1]

Speech, Hearing and Phonetic Science, University College London (UCL)[1], Linguistics, University of Potsdam[2]
tuomainen@uni-potsdam.de

ABSTRACT

This study investigated speech-in-noise intelligibility of child, young, middle aged and older adult talkers (N=48; 24 Female). The aim was to investigate whether age and sex differences in intelligibility remain when masking noise is individually tailored to talker characteristics. A total of 83 native English listeners took part in an online listening study. The listeners heard short sentences extracted from spontaneous diapix conversations; these were mixed with speech-shaped noise that was tailored to each talker's long-term average spectrum. The number of correct keywords was calculated. There were no significant differences between talker groups based on sex or age apart from the middle-aged male talkers who were less intelligible than other talker groups. These results suggest that some of the previous findings with regards to intelligibility differences between different talker groups (female vs male; younger vs older adults) may be driven by the masking potential of background noise.

Keywords: Speech intelligibility, speech perception, speech-in-noise, lifespan

1. INTRODUCTION

How well or poorly a particular talker is understood by a listener is determined by multiple parameters. For example, it has been shown that individual talker characteristics (e.g., talker sex and age), listener characteristics (e.g., listener's hearing acuity, degree of shared language experience with the talker), speaking style (casual vs clear speech), and the listening environment (quiet vs noisy backgrounds and background noise type and level) can all contribute to perceived speech intelligibility (for a review see [1]). Therefore, one could argue that speech intelligibility is highly constrained by the individual parameters that were manipulated in each perceptual assessment of intelligibility.

Assessment of speech intelligibility in background noise, in particular, has gained much attention in past research [2, 3, 4, 5, 6]. Overall, multiple studies have shown that with regards to effects of age and sex, for casual everyday speech heard in noisy backgrounds, female talkers are more intelligible than male talkers [7], and that younger adults are more intelligible than older adults (60+ years) [e.g., 6]. For children, it has been shown that children acquire intelligible speech gradually in early childhood. For example, the intelligibility of casual speech heard in quiet, is at 96% for speech produced by 3 year olds, and 99% for speech produced by 8 year olds [8]. For older children and casual speech presented in background noise, studies have shown that speech intelligibility for 11-12-year-old child talkers does not differ from adults [3, 5].

The interesting question then is that what explains these differences in intelligibility between different talker groups. There are anatomical and physiological differences between female and male talkers and younger and older talkers that result in differences in their acoustic-phonetic profiles: for example, in fundamental frequency (f0) and temporal characteristics, as well as in the amount of mid-frequency (1-3 kHz) energy in their speech [e.g., 2, 9]. However, despite these well-documented age- and sex-related differences in speech intelligibility and acoustics, studies have often failed to identify any systematic acoustic-phonetic correlates of perceived intelligibility. For example, some studies have highlighted speaking rate as a factor, with slower speech facilitating intelligibility in different multitalker and reverberant environments [2] whereas other studies have reported that f0 characteristics [7] or the amount of energy within the 1-3 kHz frequency range [4,6] are better predictors of intelligibility than speaking rate. Furthermore, these three acoustic cues are also shown to be important for speech perception in noisy environments as there is evidence that listeners use these cues to help segregate target and competing talkers in various multitalker contexts [e.g., 10].

However, one potential confound for the unclear and conflicting results with regards to predictors of perceived speech intelligibility in background noise is the plethora of different background noise types and levels used between these different studies. Because some background noise types might be less effective in masking talkers from different age groups, it is hard to draw parallels between the results from these different studies. For example, multi-talker babble that is based on a mixture of adult female and male voices may be less effective in masking child voices as the target talkers. Furthermore, signal-to-noise ratios that are either very hard [11] or very easy [3] can lead to floor and ceiling effects in listener performance which can make it difficult to statistically assess the true extent of listeners' abilities. Finally, to our knowledge, no other study has previously investigated speech intelligibility in a more lifespan design, that is, mapping differences between children and younger, middle aged and older adults in the same study.

Therefore, in the current study, we extend some of the previous findings by investigating the intelligibility of spontaneous casual speech in background noise across the lifespan (here: 10-76 years of age) at moderate difficulty levels. Here, we focus on how deriving the background noise from the individual long-term average spectrum (LTAS) of the target talker influences the perceived intelligibility of their speech. We predict that if age- and sex-related differences in perceived speech intelligibility are mainly related to the degree to which the background noise is masking the acoustic characteristics of the talker, these group differences related to age and sex will disappear, or be significantly reduced, if the background noise provides a similar degree of masking for each talker. However, if perceived intelligibility is primarily driven by individual talker characteristics such as speaking rate, mid-frequency energy and f0, we expect to replicate previous findings with regards to these sex- and age-related differences.

2. METHODS

2.1. Listeners

Participants were recruited via Prolific (www.prolific.co) [12]. Only participants with a Prolific score (i.e., participant quality rating) of 70/100 or higher were invited to the study. A total of 83 native monolingual English listeners (recruitment criteria for country of birth: United Kingdom) aged 20-42 years completed the study (*M*=29 years, SD=6.6). All participants self-declared no history of language/hearing impairments, and they passed the initial online automated screening for headphone use and the auditory attentional screening during experimental testing (see below for details). Ethical approval was obtained from the UCL Research Ethics Committee.

2.2. Speech materials

The speech materials used were taken from the LifeLUCID corpus [13], which consists of interactive speech from native Southern British English talkers across the lifespan (8-80 years) communicating with an unfamiliar partner of the same sex and similar age. The pair of talkers had to complete a spot-the-difference picture-based task (diapix; [14]) within 10 minutes or less. One person in the pair was told to take the lead in the task. We extracted speech samples from child (CH), young (YA), middle-aged (MA), and older talkers (OA) communicating in favourable listening conditions (i.e. no noise in background), thereby eliciting a conversational speaking style. The samples were short sentences or part-sentences containing between 4 and 10 words (e.g., “there's a dog sitting on a blue seat” or “the big green bush”) containing 3 keywords. Most samples were simple descriptions of the diapix pictures containing high frequency lexical items similar to BKB (Bamford-Kowal-Bench) sentences [15]. Criteria for the selection of the samples were that they be equally spaced throughout the task, did not contain any disfluencies and were not produced in response to a request for repetition by the interlocutor. Each speech sample was extracted from the original audio recording using Praat (v 6.0.21). In total, 288 speech samples were extracted from 48 lead talkers (12 CH, 10-17 years old, *M* = 13.85 years; 12 YA, 20-26 years old, *M* = 22.32 years; 12 MA, 35-49 years old, *M* = 42.95 years; 12 OA, 59-76 years old, *M* = 66.50 years; 6 female per group). Six speech samples were taken from each talker, resulting in 72 samples per age group. For the male CH group, three pre-puberty and three post-puberty talkers were included. This classification was based on the talker's median f0 with pre-pubescent males ranging from 94-99 semitones (relative to 1 Hz) and post-pubescent from 80-85 semitones.

A Praat script was used to produce tailored speech-shaped noise for each talker. This was done by creating noise with a spectrum matching their individual long-term average spectrum, which was calculated from their complete diapix speech recording (duration of noises 20 sec). These tailored noises were then mixed with the speech samples and root-mean-square normalised to 70 dB in Matlab (2016b). Initial piloting showed a fixed signal-to-noise ratio (SNR) of -4 dB to achieve an average intelligibility score (% keywords correct) of around 50-70% (avoiding floor/ceiling effects in intelligibility).

2.3. Procedure

The perception test was run online, using the Gorilla platform to run and host the experiment [www.gorilla.sc]. Data was collected between 28 May and 1st June 2020. Samples were divided into six lists with each list comprised of one sample from each of the 48 talkers. Additionally, each list included eight cross-reference trials and 4 attentional trials. The latter consisted of one sample presented in quiet from each age group (2 female, 2 male). The cross-reference trials, which were identical for all listeners, were included to check consistency in intelligibility scores across listener groups. These were additional masked samples from the existing 48 talkers (1 female, 1 male from each age group). This resulted in each participant hearing one sample from each of 48 talkers (disregarding attentional and cross-reference trials); six participants therefore made up a complete data set for each talker. Trials were randomised with attentional trials occurring every 12 trials.

After hearing each sample online once, participants were asked to type out the target sentence as accurately as possible. They were then presented with the keywords on a following screen and had to select those that they had correctly identified. Participant self-scorings were manually checked and corrected for any errors before statistical analyses.

2.4. Data analysis

Data are presented as raw scores of correctly recognised keywords where a score of 18 represents 100% correct (3 keywords x 6 talkers). Statistical analysis were conducted in R (v 1.4.1106) using an linear mixed-effects models (*lmer* function) for main effects and interactions of Talker sex (F,M) and Talker age (CH, YA, MA, OA) with listener as random effect. Post hoc comparisons were conducted using the *emmeans* package in R (Tukey p-value adjustment).

3. RESULTS

3.1 Attentional and cross-reference trials

All participants were near or at ceiling in the attentional trials (overall performance 96%). Furthermore, the performance for the cross-reference trials did not differ between the six different lists (p=.185), indicating that there were no significant differences for identical samples across listener groups who were hearing different sample lists.

3.2 Sentences-in-noise results

Raw scores (out of 18) for keywords correctly identified for female and male and CH, YA, MA and OA groups separately are displayed in Fig. 1. The best fitting model included significant interaction between Talker sex and Talker age ($\chi2$ (3)=64.40, p<.001) and, therefore, the main effects of Talker sex and Talker age were retained in the final model.

For Talker sex, female talkers (*M*=10.7, SD=3.05) were more intelligible than male talkers (*M*=9.3, SD=3.44; p<.001).

For Talker age, we found that OA and CH talkers achieved highest overall intelligibility and YA and MA talkers achieved the lowest overall intelligibility (see Fig. 1). MA talkers were the least intelligible group and they were significantly less intelligible than CH and OA talkers (both comparisons; p<.001). YA talkers were also less intelligible than OA talkers (p<.001; see Fig. 1). MA and YA talkers did not differ from each other (p=0.393).

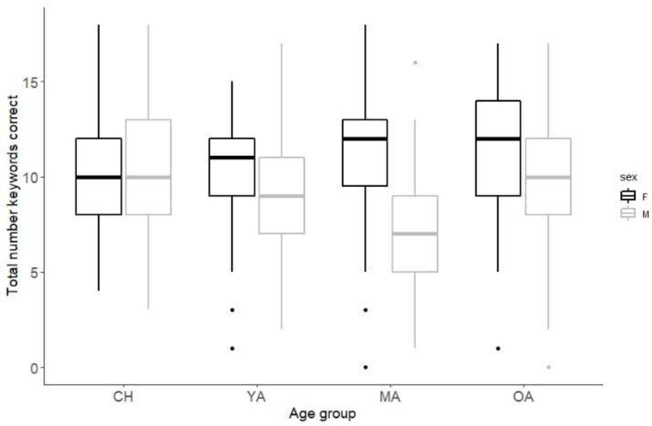

Figure 1. Number of keywords correct (out of 18) for the child (CH), young adult (YA), middle aged adult (MA) and older adult (OA) talkers separated by Talker sex (Female, Male).

However, Talker sex * Talker age interaction indicated that these sex- and age-related differences were mainly driven by the MA male talker group. First, post-hoc comparisons between the four different age groups separately for the female and male talkers revealed that none of the age groups differed in female talkers (all comparisons, p≥.110) whereas in the male talkers, the MA group was significantly less intelligible than the CH, YA and OA groups (all comparisons p<.001). Second, post-hoc comparisons between female and male talkers within each of the four age groups revealed that female talkers were significantly more intelligible than male talkers only in the MA group (p<.001); all other comparisons p≥.134).

4. CONCLUSIONS

Our data supports the hypothesis that age- and sex-related differences in perceived speech intelligibility are, at least partially, related to the properties of the background masking noise. When the masking noise was individually tailored to the individual talker characteristics with respect to its long-term average spectrum, we generally found very little differences in perceived intelligibility in different-aged talker groups.

Contradicting many previous studies, we found here that child and older adult talkers were (at least marginally) easiest to understand and middle aged adults were most difficult to understand in individually tailored background noise. On the other hand, aligning with many previous findings, we reported here too that female talkers were more intelligible than male talkers. However, in the current study these group differences with regards to age and sex were primarily driven by the low score achieved for the middle aged male talkers. Although male talkers achieved a lower score than female talkers in all four groups, the difference generally remained relatively small at one to two fewer keywords correctly identified (out of 18 total), whereas the intelligibility score for the middle aged male talkers was approximately four to five keywords correct lower than that for their female peers.

The reasons for this difference in this particular age group are unclear. Perceptually, it appeared that, despite tailoring the background noise masker to the acoustics of individual talkers, some target talkers might still be masked better by the noise than others. For example, it is possible that for some of the talkers the masking noise better covered lower frequency regions (that are rich in phonetic detail for word identity) leaving fewer audible phonemes/words to aid recognition [16]. Furthermore, as stated earlier, speech intelligibility is very likely a result of multiple parameters involving the individual characteristics of both the listener and the speaker as well as the content of the message itself. Therefore, we plan to further investigate these individual differences and look at which talkers are more/less intelligible and what, if any, acoustic parameters might help explain these results and predict speech intelligibility in our dataset. We will also conduct qualitative analyses of the content of the speech samples used in the study to further rule out the effect of sentence predictability on speech intelligibility in our sample.

Lastly, it is worth noting that although the talkers in this study all spoke the Southern British English variety, we have no information about the regional accents of the listeners. However, listener characteristics are an unlikely explanation for our results as the different listener groups performed consistently and did not differ on the cross-reference trials. Furthermore, despite the fact that it has been well established that shared language and accent background between the speaker and the listener (e.g., in terms of phonology) increase perceived intelligibility [see e.g., 17], it is nevertheless unlikely that a discrepancy between the accents influence intelligibility of one particular sub-group of talkers, such as middle aged male talkers.

In summary, this study highlights the importance of carefully considering masking characteristics in studies of speech in noise that investigate correlates of speech intelligibility.

5. REFERENCES

[1] Smiljanić, R., & Bradlow, A. R. 2009. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*, *3*(1), 236-264.

[2] Cox, R., Alexander, G., and Gilmore, C. 1987. Intelligibility of average talkers in typical listening environments *The Journal of the Acoustical Society of America*. 81, 1598 – 1608.

[3] Hazan, V., & Markham, D. 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, *116*(5), 3108–3118. https://doi.org/10.1121/1.1806826

[4] Hazan, V., Tuomainen, O., Tu, L., Kim, J., Davis, C., Brungart, D., Sheffield, B. 2018. How do aging and age-related hearing loss affect the ability to communicate effectively in challenging communicative conditions?, *Hearing Research*, 369, 33-41.

[5] Smiljanic, R., & Gilbert, RC. 2017. Intelligibility of Noise-Adapted and Clear Speech in Child, Young Adult, and Older Adult Talkers. *Journal of Speech, Language, and Hearing Research*, 60(11), 3069–3080. https://doi.org/10.1044/2017_JSLHR-S-16-0165.

[6] Tuomainen, O., Hazan, V., Davis, C., & Kim, J. 2019. Intelligibility of conversational and clear speech in young and older talkers as perceived by young and older listeners. *The Journal of the Acoustical Society of America*, *146*(1), EL28.

[7] Bradlow, AR., Torretta, GM., Pisoni, DB. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication*, 20(3), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5

[8] Flipsen P. 2006. Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics*, 20(4), 303–312.

[9] Tuomainen, O., Taschenberger, L., Rosen, S., & Hazan, V. 2022. Speech modifications in interactive speech: effects of age, sex and noise type. *Philosophical transactions of the Royal Society of*

London. Series B, Biological sciences, 377(1841), 20200398. https://doi.org/10.1098/rstb.2020.0398
[10] Cushing, IR., & Dellwo, V. 2010. The role of speech rhythm in attending to one of two simultaneous speakers. In *proceedings of speech prosody*, Chicago.
[11] Smiljanic R. 2013. Can older adults enhance the intelligibility of their speech?. *The Journal of the Acoustical Society of America*, 133(2), EL129–EL135. https://doi.org/10.1016/j.heares.2018.06.009.
[12] www.prolific.co. London, UK (accessed March 2020).
[13] Tuomainen, O, Taschenberger, L., Hazan, V. 2022. LifeLUCID Corpus: Recordings of Speakers Aged 8 to 85 Years Engaged in Interactive Task in the Presence of Energetic and Informational Masking, 2017-2020. Data Collection. Colchester, Essex: UK Data Service. 10.5255/UKDA-SN-854350.
[14] Baker, R., Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43, 761–770
[15] Bench, J., Kowal, A. and Bamford, J. 1979. The BKB (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children. *British Journal of Audiology*, 13, 108-112. http://dx.doi.org/10.3109/03005367909078884
[16] Cooke M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://doi.org/10.1121/1.2166600
[17] Hansen Edwards, JG., Zampini, ML., & Cunningham, C. (2018). The accentedness, comprehensibility, and intelligibility of Asian Englishes. *World Englishes*, 37(4), 538-557