YOUNG AND OLD LISTENERS' PERCEPTION OF SPEECH IN A BACKGROUND OF ENGLISH- AND FOREIGN-ACCENTED BABBLE

Antje Heinrich[1,2], Korlin Bruhn[2], Sarah Hawkins[3]
[1] *MRC Institute of Hearing Research Nottingham, UK.*
[2] *Department of Linguistics, University of Cambridge, UK.*
[3] *Centre for Music and Science, University of Cambridge, UK.*
Email: antje.heinrich@ihr.mrc.ac.uk

Abstract

We investigated the effect of various background voice accents on the intelligibility of speech spoken in the native accent of the listener. Studies so far have only examined the influence of target speech accent on intelligibility, but not the influence of background babble accent or that of age. Our babble contained one, three or eight talkers who spoke English with a British (BE), American (AE), South Indian (Ind), or Neapolitan Italian (Ital) accent. We tested 64 young (mean age: 21.3 years) and 48 old (66.7 years) British English-speaking listeners with normal hearing. Babble was most disruptive to intelligibility when it was Ind-accented and contained more than one voice. Moreover, intelligibility decreased more in young than old listeners when the number of background voices increased. Hence, the accent of background speech affects intelligibility of natively-accented target speech and the effect of age interacts with the number of background speakers.

1) Energetic versus informational masking – the importance of speaker number and language

Understanding spoken language can be challenging when the background is noisy. However, the nature of the background noise plays an important role in determining the difficulty of the task. All background noise contains an energetic component, which will partly or completely mask the energy in the signal. In addition, when the background noise is a meaningful signal, it may be particularly difficult for a listener to identify, locate and separate the background noise from the target signal and to prevent its intrusion into the message (e.g., Schneider, Li & Daneman, 2007), a type of masking referred to as informational masking. Testing the effectiveness of masking of different types of noise for speech perception, Cooke and colleagues have shown that stationary noise is less effective than a single competing voice, which in turn is less effective than multi-talker babble (Simpson & Cooke, 2005; Garcia Lecumberri & Cooke, 2006). Presumably this result occurred because stationary noise only evoked energetic masking whereas background voices combined aspects of energetic and informational masking, though variation in background babble could also challenge the listener by making it harder to predict the shape of the babble noise and thus stream it from the signal. Simpson & Cooke (2005) also showed that once the number of voices in the background had reached a certain number (eight in their case), the effectiveness of the babble to interfere with the intelligibility of the target speech reached a plateau or decreased. They speculated that energetic and informational maskers follow separate trajectories of effectiveness and eight babble voices presented the most efficient combination of the two types of maskers; the number of voices was still small enough to exert informational masking, yet large enough to exert significant energetic masking as well.

From these experiments, the question arises of what exactly makes speech noise such an effective masker for speech. Is the adverse effect of background speech due to the

linguistic content or to the acoustic characteristics of the speech maskers? Van Engen and Bradlow (2007) examined this question by comparing the interfering effect of English vs. Mandarin babble on monolingual native-English listeners. In two-talker babble, English listeners performed worse when the babble was English speech than when it was Mandarin, whereas in the six-talker babble condition there was no difference. These data suggest that linguistic interference is more effective when the language of the background voices is known to the listener, at least in the case of two background voices. This result is consistent with Simpson & Cooke's (2005) suggestion that informational masking decreases as number of background voices increases. However, while van Engen and Bradlow's results suggest that the different effects due to English and Mandarin babble may stem mainly from linguistic differences between these babbles, it is important to remember that the two chosen languages differ not only in abstractions like vocabulary and grammar, but also in acoustic-phonetic patterns. It is unclear how much simply the unfamiliarity of these speech patterns contributed to the differences in their results. Rhebergen et al. (2005) offered evidence that non-lexical speech information can also act as an effective masker of speech. They tested Dutch and Swedish normal and reversed speech maskers for Dutch speech presented to Dutch speakers. Although not all results of this study were easy to understand they did show that reversed speech unintelligible to the listener can nonetheless interfere with intelligibility.

In all studies in the literature the crucial comparison for the effectiveness of maskers has been between maskers in the listener's native language versus in a language unknown (or made unintelligible) to the listener. This approach makes it difficult to distinguish to what degree physical versus abstract properties of the speech-language influence the masker's effectiveness. Although the current study does not solve this issue, as a first step it kept the language of the masker constant, namely English, but varied the variety used, with two native accents, and two foreign accents: Southern British English (BE), American English (AE), Italian (Ital) and Indian (Ind). Effects of number of babble voices was also explored.

2) The effect of aging on energetic and informational masking

The effect of age on energetic and informational masking is well documented and very consistent between studies. In term of energetic masking, older adults need a better signal-to-noise ratio (SNR) than young adults in order to achieve the same intelligibility (e.g., Li et al., 2004; Heinrich & Schneider, 2011 among others), but once this is taken into account, intelligibility is similar. In contrast, informational masking seems to show very few if any age-related effects (see Li et al., 2004; Agus et al., 2009; Wightman et al., 2010). Helfer and Freyman (2008), using meaningful sentences for both target and some of their masking conditions, found that young and old listeners found speech perception particularly difficult when target and masking sentences were spoken by people of the same sex, and that performance in both age groups improved equally when spatial separation was introduced. In contrast, when the masking sentences were spoken by people of opposite sexes, older listeners were disproportionately disadvantaged compared with their young counterparts (see also Humes et al., 2006). Two explanations were offered for this result: that older adults may have been less able to ignore the information from the masker, and/or to take advantage of the differences in spectral fluctuation between the voices of the two sexes.

In this study, we varied the number of talkers in background babble to investigate how age is affected by energetic and informational masking of speech. Together with Simpson and Cooke (2005) we assumed that informational masking would decrease while energetic masking increased with number of background talkers. Moreover, assuming that only energetic masking consistently affects age-related performance, we expect age differences to increase with the number of background voices. We also manipulated context-

induced predictability of the target word to explore how top-down linguistic compensation strategies may alter age effects, and, as noted above, effects of different varieties of English.

Method

Participants

Participants were 64 young and 48 old listeners. The young listeners formed four groups of 16 listeners each, and the 48 older listeners four groups of 12. Mean ages were 21.3 years (SD = 2.9) and 66.7 years (SD = 4.9) for young and old adults respectively. All young listeners were undergraduates at the University of Cambridge. The older group were local residents. All reported normal speech and hearing function. Group average for audiometric thresholds of left and right ears is in Figure 1. All participants were compensated £5.00/ hour.

Stimuli and Apparatus

Target sentences comprised 36 sentence pairs contrasting in predictability of the last word, the keyword, which was identical in each pair. Thus each sentence pair had the same final word but differed in the preceding 'sentence base' e.g. *The birds flew overhead in a huge flock; the boys knew where to look for a huge flock.* Sentences were phonetically controlled and keywords chosen so that observed differences between them would largely reflect lexical predictability. All keywords were monosyllabic with at least two cohort neighbours, and the immediate phonetic context of the keyword was identical in both sentences. Additionally, sentence bases were unique, yet natural and meaningful; pairs could only vary by one syllable or word and had similar rhythm. A further 12 unpaired sentences were used for practice.

The target sentences were recorded by a male speaker of Southern Standard British English (SSBE) aged 37 years, a trained phonetician. He read them as naturally as possible, with nuclear stress on the keyword, and the same prosodic pattern for both members of a pair.

The voices for the babble signal were recorded from 14 speakers of BE, 15 speakers of AE, 24 speakers of Tamil and Telugu and 9 from Italy (Naples). Eight from each language group were chosen to form the babble. Five of the eight BE talkers spoke SSBE, including all three speakers in the 3-talker babble. The remaining three speakers in the 8-talker babble had mild but recognisable regionalisms from Yorkshire, Coventry and Scotland. The AE speakers

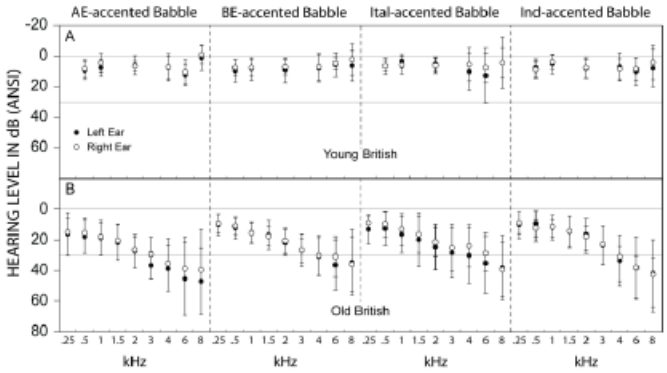

Figure 1. Average hearing levels for each of the participating groups of young (N=16 per group) and older (N=12 per group) adults.

hailed from a variety of US locations but none had a strong regional accent. Speakers from India were native speakers of Telugu, Kannada or Tamil, while speakers from Naples all had Italian as their mother tongue. All non-native English speakers were chosen for having typical local accents and being able to read English reasonably fluently. None had lived in a country where the typical mother-tongue is English.

Each speaker read from a book of their own choice for a minimum of 30 minutes. The eight recordings of the highest quality were chosen for the babble signal, with the provision that only male voices were included in the 1- and 3-talker babble. Each recording was individually edited to remove silence of more than 180 ms in order to minimize any temporal discontinuities between 1-, 3- and 8-talker babble, and to maximize the likelihood that differences in babble conditions were due to the intelligibility of single words and not the availability of amplitude dips in the signal. Moreover, each babble voice was re-scaled to a common RMS value before being combined with other voices to ensure that no voice would stand out from the babble signal.

Procedure

Testing took place in a double-walled sound attenuated chamber. Sentences were presented diotically over TDH49 headphones at a comfortable listening level that each listener was asked to choose in a two-stage process: first, a 12-talker babble of some of the recorded voices was played and the listener changed its level until it reached their most comfortable level (MCL). Listeners were encouraged to increase and decrease the presentation level several times to ensure that they had found their MCL. For young listeners, the average MCL was 82 dB SPL, for old listeners 81 dB SPL. Participants then listened to 12 practice sentences, three in quiet and three in each of the three babble conditions and were encouraged to re-adjust the sound intensity if needed. For each sentence, listeners wrote down what they heard. They were asked to guess if unsure. To equate overall performance between the age groups, SNR was -5 dB for young and -2 dB for old listeners after extensive piloting. Each listener was presented with both predictability conditions and with babble signals containing 1- 3- or 8-talkers. However, each listener heard only one babble accent (BE, AE, Ind, Ital).

Results

Intelligibility was measured as number of correct final words. Figure 2 shows the results. Prior to data analysis, the scores from each subject (3 babble x 2 predictability conditions) were converted to rationalized arcsine units (rau) to equalize the difference intervals across the whole range of scores (Studebaker, 1985). We then conducted a 3 talker-number (1, 3, 8) by 2 predictability (high, low) by 4 babble accents (AE, BE, Ind, Ital) mixed-measures ANOVA separately for young and old English listeners. Both age groups had main effects of talker number (young: $F[2, 120] = 86.71$, $p < 0.001$; old: $F[2, 88] = 18.66$, $p < 0.001$), predictability (young: $F[1, 60] = 173.02$, $p < 0.001$; old: $F[1, 44] = 85.47$, $p < 0.001$) and babble accent (young: $F[1, 60] = 5.73$, $p = 0.002$; old: $F[1, 44] = 6.09$, $p = 0.001$). Both had an interaction between talker number and babble accent (young: $F[6, 120] = 12.70$, $p < 0.001$; old: $F[6, 88] = 4.29$, $p = 0.001$). But only young listeners had an interaction between predictability and babble accent ($F[3, 60] = 7.42$, $p < 0.001$).

The main effect of *predictability* confirmed that we successfully wrote high- and low-predictability sentences; all listeners correctly heard more high- than low-predictability sentences (young: 76.5% versus 53.9 %; old: 76.6% versus 62.4%). Interestingly, old adults performed relatively better for low-predictability sentences than young ones, an interaction with age that reached significance ($F[1, 110] = 4.75$, $p = 0.03$). The main effect of *talker*

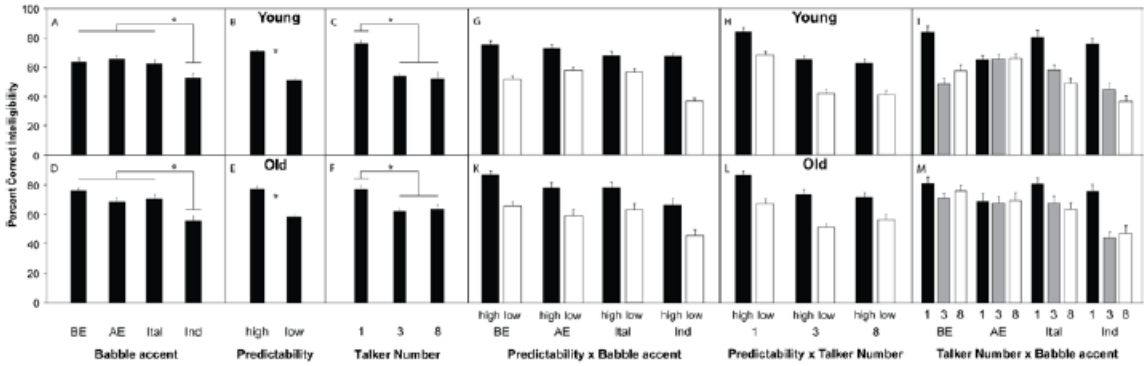

Figure 2. Intelligibility of high- and low-predictability sentences presented in 1-, 3-, or 8-talker babble of native and foreign accents of English. Main effects are displayed in panels A-C (young adults) and D-F (old adults), interactions in panels G-I (young) and K-M (old).

number was due in both age groups to a significant decrease in intelligibility when the number of talkers in the babble increased from 1 to 3. An increase to 8 talkers did not further affect intelligibility in either age group. Interestingly, the decrease in intelligibility due to increased talker numbers was more pronounced for young than old listeners, giving rise to a significant interaction with age ($F[2, 220] = 4.64$, $p = 0.01$). As for *babble accents*, it was mainly the Indian-accented babble that led to lower intelligibility compared with all other babble types; this effect was comparable in the two age groups. The *interaction between talker number and babble accent* in young adults occurred because this group showed no effect of talker number for AE babble while showing a decrease from 1 to 3 talkers for all other babble accents. Older listeners showed essentially the same pattern with the somewhat surprising addition that they showed no reliable decrease as talker number of BE-accented babble increased; this difference led to a significant talker number x babble accent x age interaction ($F[6, 208] = 2.31$, $p = 0.03$). Lastly, the young listeners' *interaction between predictability and babble accent* arose because the difference between high- and low-predictability sentences was not significant for AE babble in this age group.

Space limitations prevent us from describing the detailed effects of hearing loss on the tested variables. Suffice it to say that principle component analysis was used to create a composite score of all hearing level measures of older listeners, and this score was used as a covariate in re-calculating the original 2 predictability x 3 talker number x 4 babble accent mixed-measures ANOVA. The previous pattern of results was partially preserved in that predictability and babble accents remained significant main effects. However, the effect of talker number now interacted with predictability: while the intelligibility difference between high- and low-predictability sentences was essentially the same across all three background talker conditions for older adults with good hearing, the low-predictability condition with 3 background talkers was particularly taxing for the old listeners with poor hearing. In addition, using hearing level as a covariate produced a new interaction between babble accent and predictability. This interaction effect occurred because the smallest difference between high- and low-predictability sentences shifted from AE-accented babble for older adults with good hearing to Ital-accented babble for older adults with poorer hearing.

Discussion

The study replicated and extended a number of results in the literature. It replicated Simpson and Cooke's (2005) finding that intelligibility declines with an increase in babble voices. It

extends Rhebergen et al.'s (2008) finding in showing that factors such as unfamiliar pronunciation (including rhythm and intonation) can be effective informational maskers. Although we have yet to test the intelligibility of the masker speech, there were several indications in our results that babble signals composed of foreign-accented speech were more effective maskers, other things equal, than babble spoken by native speakers. Our results also partially replicated Helfer & Freyman's (2008) and Humes et al's (2006) finding that older listeners did not have more difficulty separating a target from a babble signal when voices had the same gender. Indeed, in our data, which used only male voices for both target and babble, the trend was for older adults to have less difficulty in 3- and 8-talker babble conditions than their young counterparts. We also found no differential age effect when the number of babble voices increased, an effect that we might have expected if effects of age increased as energetic masking levels rose. Overall, in this study older adults show remarkably similar results to young listeners despite their poorer hearing sensitivity in many cases. Moreover, the study shows that not all babble has the same masking effects, even if it is all spoken in the same language, and that experimenters should keep this in mind when opting for pre-recorded babble signals in which accent of the babble does not match the accent of the target speaker and/or listener.

Acknowledgments

This study was supported by ERA-AGE FLARE grant RG49525 to the first author. We thank Profs. B.Yegnanarayana and F. Cutugno for kindly arranging the Indian and Italian recording.

References

García Lecumberri ML & Cooke MP. (2006). Effect of masker type on native and non-native consonant perception in noise. *JASA*, *119*, 2445-2454.

Heinrich A, Schneider BA. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *QJEP, 64,* 186-205.

Helfer KS & Freyman RL (2008). Aging and speech-on-speech masking. *Ear Hear*, 29, 87–98.

Humes LE, Lee JH, Coughlin MP. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *JASA*, *120*, 2926–2937

Li L, Daneman M, Qi JG & Schneider BA. (2004). Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *JEP:HPP, 30,* 1077-1091.

Rhebergen KS, Versfeld NJ, Dreschler WA. (2005) Release from informational masking by time reversal of native and non-native interfering speech. *JASA, 118,* 1274-7.

Schneider BA, Li L, & Daneman M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *JAAA, 18,* 559-572.

Simpson S & Cooke MP. (2005). Consonant identification is N-talker babble is a nonmonotonic function of N. *JASA, 118,* 2775-2778.

Studebaker GA(1985) A "rationalized" arcsine transform. J Speech Hear Res, 28, 455-62.

Van Engen K. & Bradlow AR. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *JASA, 121*, 519-526.

Wightman FL, Kistler DJ, O'Bryan A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *JASA, 128,* 270–279.i