SPEECH COMMUNICATION IN BACKGROUND NOISE: EFFECTS OF AGING

Outi Tuomainen, Valerie Hazan, Linda Taschenberger

Department of Speech Hearing and Phonetic Sciences, UCL, UK
o.tuomainen@ucl.ac.uk, v.hazan@ucl.ac.uk, l.taschenberger@ucl.ac.uk

ABSTRACT

Most communication in everyday life takes place in less than ideal listening conditions and the presence of noise or other voices in the background is often challenging for older adults. This study investigated how the informational content of the background noise affects communication in age-matched younger and older talker pairs. We used an interactive 'spot-the-difference' task to elicit spontaneous interactive speech in younger (20-27 years, N=10) and older (58-76 years, N=20) female talkers with normal hearing. The task was done in quiet and in three different masker conditions varying in informational content: no content (speech-shaped noise), unrelated speech, related speech. Communication efficiency, self-rated listening effort, and acoustic-phonetic features of speech were measured for all conditions. Even though younger and older talkers did not differ in task efficiency or self-rated effort in the masker conditions, some of these conditions elicited speech adaptations consistent with increased speaking effort in older talkers.

Keywords: speech production, speech-in-noise, spontaneous speech, speech adaptations, aging.

1. INTRODUCTION

Our ability to communicate successfully with others can be strongly affected by the presence of noise and other voices in the environment, and older adults (OAs) can be more greatly affected than young adults (YAs) in these situations [1]. The interference (or *masking*) that is caused by one sound on the perception of another depends on the characteristics of the masker: whether energetic (EM) or informational masking (IM, e.g., speech-on-speech masking). The weighting between the effects of EM and IM can vary with listener age, with some studies showing increased effect of IM in older adult listeners [11, 14]. However, most studies have focused on the effects of background noise onto speech perception and relatively few have investigated the interference of these noise types on speech production [4]. One such study [6] showed that while completing interactive Sudoku-tasks, young adult talkers adapted the way they speak in proportion to the energetic masking capacity of the background noise by increasing fundamental frequency and speech intensity especially in the higher frequency range (i.e., reducing spectral tilt). Furthermore, they showed that for speech-on-speech masking, talkers reduced the degree of temporal overlap between the masker and their own speech to increase the separation between the two competing sound sources for the interlocutor. These results suggest that, first, talkers adopt different strategies to overcome the effect of background noise depending on the types of noise they are exposed to. Second, when interacting with another person, talkers actively monitor their background ("listening while speaking") and exploit pauses in the maskers to maximise ease of communication in difficult listening conditions.

Taking a different perspective on interactive speech in noise, a recent study [9] focused on the effects of aging and (mild) age-related hearing loss on communication efficiency and acoustic-phonetic adaptations made by older and younger adults in good and adverse listening conditions. In this study, pairs of OA (either with normal hearing, OANH, or with hearing loss, OAHL) and YA talkers completed collaborative 'spot-the-difference' picture tasks [15] with a YA interlocutor in different listening conditions, including a condition where both talkers interacted in 8-talker babble noise. Even though adding background noise did not affect communication efficiency (i.e., task completion time) in any of the three groups, the OAHL group differed from the YA and OANH groups in the speech adaptations made to compensate for the effects of background noise: when speaking in babble noise, only the OAHL group made adaptations that were consistent with an increase in speaking effort.

In the current study, we expanded the studies by [6] and [9], by investigating communication efficiency, listening and speaking effort in normally hearing older adults, communicating with an age-matched interlocutor, under good and adverse listening conditions that differ in terms of their energetic masking potential and informational content. We recorded spoken interactions taking place (a) in quiet, (b) in noise with no informational content ("pure" EM), (c) background speech related to the task (IM), and (d) background speech unrelated to the task (IM). To elicit spontaneous interactive speech, we used the same problem-solving task as in [9] under these four different listening conditions. We measured communication efficiency (i.e., time it took to find the differences), self-rated listening effort and speaking effort (i.e., acoustic-phonetic adaptations of the speech).

We predicted that OAs would be less efficient communicators in background noise, and that it would take them longer to complete the task than YAs. Second, we predicted that the presence of noise in the background would lead to an increase in both listening effort and speaking effort in older adults. Lastly, we predicted that older adults would be more affected by the informational content (i.e., unrelated versus related speech) of the interfering masker sounds than YAs. This would lead to a decrease in communication efficiency and increase in speaking effort in the related speech condition.

2. METHOD

2.1. Participants

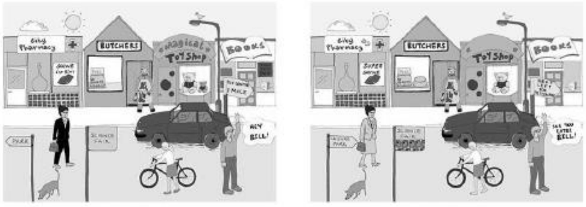
Speech was audio recorded from 30 age-matched female pairs of monolingual native speakers of Standard Southern English aged between 20-27 years (Younger Adults, YAs, N=10, Mean age 22.5 years) and 58-76 years (Older Adults, OAs, N=20, Mean age 66.3 years). All participants had normal hearing thresholds (<20 dB HL) across the 0.25-4 kHz range and reported no history of speech and language impairments or neurological trauma. All participants aged over 65 years passed a Montreal Cognitive Assessment (MOCA) –dementia screening test [12].

2.2. Procedure

During the audio recordings, participants sat in separate acoustically-shielded rooms and communicated via headsets fitted with a cardioid microphone (Beyerdynamic DT297) whilst playing interactive ‘spot-the-difference’ Diapix games [3] on a desktop PC. Participants were given different versions of the same picture scenes (Fig. 1) and told that they had 10 minutes to find the 12 differences between the pictures. Each talker was recorded on a separate channel at a 44 100 Hz (16 bit) sampling rate using a Fireface audio interface and Audacity audio software. One of the talkers was instructed to lead the interactions; all analyses reported here focus on the lead talker.

The picture task was carried out in four listening conditions affecting both participants: both speakers in i) quiet (“normal”, NORM), ii) EM with no informational content (speech-shaped-noise, SPSN), iii) IM that is semantically related to the picture description task (IM-RE; i.e., talking about the same picture), and iv) IM that is semantically unrelated to the task (IM-UR; i.e., talking about a different picture). Both IM-RE and IM-UR were 3-talker maskers consisting of a male, a female and a child speaker. The picture and noise condition orders were randomised. In order to increase real-life validity of these adverse conditions, we used a Spatial Audio Simulation System software (Audio 3D) [2] that mimics real room acoustics combined with head-related transfer functions in real-time via headphones. The maskers and the voice of the interlocutor were spatially separated by 1 meter from both each other and the “live” talker. These virtual room simulations were presented over headphones at 72 dB in the noise conditions.

Figure 1: A Diapix picture pair.



Street 3A Street 3B

2.3. Data processing

All recordings were automatically transcribed using a speech recognition system by Speechmatics [7] and then manually corrected for word-level errors and audio-transcription misalignments. From the recordings we calculated three measures that reflect i) communication efficiency (i.e., time in seconds from start to finding 8th difference), ii) listening effort (11-point scale: “Did you have to put in a lot of effort to understand your partner?”, 0=lots of effort, 10= no effort) and iii) three acoustic-phonetic measures that reflect speaking effort (i.e., articulation rate, median F0 and intensity; see [9] for details of the analysis).

Articulation rate was calculated as the number of syllables divided by the total duration in seconds of the speech regions (syllables/second). Syllable counts were calculated using the qdap package in R [12], after exclusion of segments labelled as unfinished words, hesitations, fillers, and agreements.

For median F0, a Praat [5] script was used to concatenate all speech; F0 calculations were then done using the “pitch” function, with a time step of 100 pitch values per second. A formula by [8] was used to calculate ceiling and floor limits specific to each talker and median F0 values were calculated per condition. Values in Hertz were converted to semitones relative to 1 Hz.

For the intensity measure, long-term average spectra (LTAS) were calculated in Praat. First, the intensity of labelled speech segments was calculated (values above 88 dB were excluded, as likely instances of shouting). The remaining segments were concatenated and scaled to 75 dB. The signal was then bandpass filtered between 1 and 3 kHz and the mean intensity of the resulting waveform calculated for a measure of the amount of energy in the 1–3 kHz frequency range relative to the total energy in the spectrum (ME13). An increase in the relative energy in this mid-frequency band reflects a reduction of spectral tilt. For evidence of increased speaking effort, we expected to see, first, a reduction in speaking rate and second, a significant positive correlation between the median F0 and ME13 measures [9].

3. RESULTS

Repeated measures ANOVAs were carried out for within-subject factor listening Condition (4: NORM, IM-RE, IM-UR, SPSN) and a between-subject factor Group (2: YA, OA) separately for communication efficiency and listening effort. To quantify speaking effort we, first, ran a repeated measures ANOVA (as above) for articulation rate, F0 median and ME13 measures and, second, Pearson’s bivariate correlations between the median F0 and ME13 measures separately for the YA and OA groups. For the correlations, the F0 and ME13 values were normalised as percentage change relative to NORM.

3.1. Communication efficiency

For communication efficiency, there were no significant effects of listening Condition, F(3, 84)=2.14, p=.102, or Group, F(1,28)=0.34, p=.505, and the interaction between Group and Condition was also not significant, F(3,84)=0.83, p=.482 (see Table 1 for mean scores). Overall, across groups, neither adding noise nor the energetic and informational content of the background masker affected the time it took to find differences in the Diapix task.

Table 1: Mean (in seconds) and SDs for time it took to find the 8th difference in the Diapix task for the younger (YA) and older (OA) talkers.

	YA	OA
NORM	237 (60)	274 (75)
IMRE	257 (70)	279 (68)
IMUR	277 (102)	268 (62)
SPSN	284 (85)	302 (117)

3.2. Listening effort

Self-rated listening effort increased significantly in all noise conditions, F(3,84)=17.97, p<.001 (see Table 2), but neither the main effect of Group (p=.141) nor Group and Condition (p=.368) interaction were significant. All three noise conditions were rated as significantly more effortful than NORM (p≤.001 all comparisons) and the informational masking conditions were rated as significantly more effortful than the SPSN condition (p≤.028). The two informational masking conditions did not differ from each other (p=.711).

Table 2: Means and standard deviations for self-rated effort (10=”no effort”, 0= “lots of effort”) in the Diapix task in different listening conditions for the younger (YA) and older (OA) talkers.

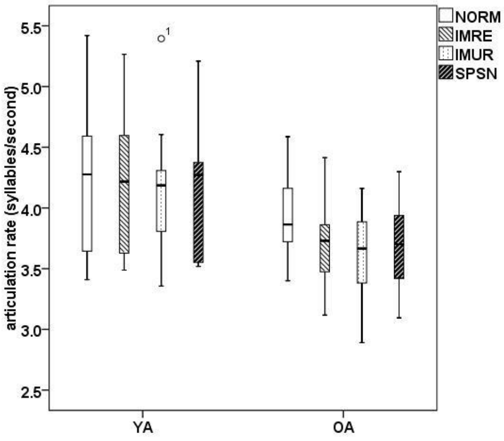
	YA	OA
NORM	8.20 (2.35)	8.18 (2.66)
IMRE	3.70 (1.89)	5.58 (3.04)
IMUR	3.68 (2.26)	5.25 (2.87)
SPSN	5.05 (2.31)	6.18 (2.78)

3.3. Speaking effort

3.3.1 Articulation rate

There were significant main effects of Condition, F(3,84)=11.98, p<.001 and Group, F(1,28)=7.77, p=.009 and the Group and Condition interaction approached significance, F(3,84)=2.49, p=.066; see Fig. 2). OAs (M=3.73 s/s) were slower speakers than YAs (M=4.19 s/s), and articulation rate was faster for NORM than for adverse conditions (all comparisons, p<.001). The follow-up comparisons for the near-significant Group and Condition interaction revealed that this difference in articulation rate between listening conditions was driven by the OA group: articulation rate was not significantly different between listening conditions for the YA talkers (p=.194) whereas it was for the OA talkers (p<.001, see Fig. 2).

Figure 2: Articulation rate for younger adult (YA) and older adult (OA) talker groups.



3.3.2 Median F0, ME13 and correlations between F0 median and ME13 measures

There were no significant group differences in the F0 median and ME13 measures ($p \geq .207$). Both YA and OA groups significantly raised median F0, $F(3,84)=118.21$, $p<.001$, and ME13, $F(3,84)=89.75$, $p<.001$, in background noise with largest changes for the SPSN condition ($p \leq .001$). For speaking effort, a significant positive correlation between the F0 median and ME13 measures were obtained for the OA talkers in the IMRE and SPSN conditions and a marginally significant correlation in the IMUR condition (see Table 3). No such relationship between these measures was found in the YA group in any of the noise conditions.

Table 3: Correlations between F0 median and ME13 (% change relative to NORM) in younger (YA, N=10) and older (OA, N=20) talkers for different noise conditions.

	YA	OA
IMRE	r=.-.150, p=.679	r=.579, p=.008**
IMUR	r=.-.303, p=.395	r=.433, p=.057
SPSN	r=.-.062, p=.865	r=.493, p=.027*

4. DISCUSSION

We investigated if older adult talkers are less efficient and need to exert more listening and speaking effort when communicating in background noise with an age-matched interlocutor. We also investigated if the informational content of the background noise is more disruptive for older talkers than it is for younger talkers. Against our predictions, background noise had no effect on communication efficiency as measured by the completion time of a picture description task. When speaking in noise, however, both younger and older listener groups reported increased listening effort in the noise conditions with highest effort rating for the related and unrelated informational masking conditions. For the older talker group, background noise also increased speaking effort with slower articulation rate and concomitant changes in pitch and mid-frequency energy measures in noise relative to quiet. Against our predictions, however, the semantic relatedness of the informational background noise did not play a role in communication difficulties or the speech adaptations made to compensate for these difficulties in older talkers.

These results conflict with previous findings by [9] who showed increased speaking effort in background noise only in older adults with age-related hearing loss. In the study by [9], normally hearing older adults did not need to exert speaking effort when speaking in 8-talker babble noise, a more “static” type of masker, where the semantic content is unintelligible. The current study differed from [9] in two important aspects: i) we used a more intelligible 3-talker babble with spatially separated sound sources, and ii) both groups of talkers interacted with age-matched interlocutors. More intelligible maskers can cause more interference and may require exerting more effort, especially for older talkers [14]. Also, talkers adapt their speech to meet the needs of their interlocutor [10], and interacting with another older adult, who is likely to be more affected by the background noise than a younger adult interlocutor, may require investing more effort to enable effective communication. Furthermore, the fact that there was no difference in communication efficiency between good and adverse listening conditions can also reflect different allocation of effort between younger and older talkers. Younger adults may not find these noise conditions particularly challenging whereas older adults may only achieve the same level of efficiency between good and adverse conditions by investing more speaking effort in the noise conditions.

In summary, these results suggest that normally hearing younger and older adults find interacting in background noise effortful, especially if background noise was speech, but only older adults need to recruit more speaking effort to maintain successful communication with largest effects observed for background speech that was semantically related to the picture description task and for speech-shaped-noise.

The current study has some limitations, however, as effort can be expressed in different ways by different listeners and talkers and also at different linguistic levels (i.e., effort is not necessarily restricted to only acoustic-phonetic speech adaptations [4,6]). Thus, further investigations on different talker groups (e.g., male talkers, older adults with age-related hearing loss) and levels (linguistic adaptations, effects of listener feedback, temporal adjustments) is required to gain a more comprehensive picture of communication difficulties in older adults.

5. ACKNOWLEDGEMENTS

This work was supported by the Economic and Social Research Council [grant number ES/P002803/1].

5. REFERENCES

[1] Agus TR, Akeroyd MA,. (2009). An analysis of the masking of speech by competing speech using self-report data. *J. Acoust. Soc. Am*, 125, 23-26.

[2] Audio3D virtual room software: www.phon.ucl.ac.uk/resource/audio3d/

[3] Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43, 761-770.

[4] Beechey, T, Buchholz, JM, Keidser, G. (2018). Measuring communication difficulty through effortful speech production during conversation. *Speech Communication*, 100, 18-29.

[5] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer. http://www.praat.org/

[6] 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, 2059-2069.

[7] Cloud transcription service. https://www.speechmatics.com

[8] De Looze, C., Hirst, D. J. 2008. Detecting key and range for the automatic modelling and coding of intonation, *Actes de Speech Prosody* 2008, Conference, Campinas, 135-138.

[9] Hazan, V., 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, 1331-1346

[10] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W.J. & Marchal, A. (eds.), *Speech production and speech modelling*. Dordrecht, the Netherlands: Kluwer Academic, 403–439.

[11] Mattys, SL., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, 59, 203-243

[12] Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 53, 695–699.

[13] Rinker, T. W. 2013. qdap: Quantitative discourse analysis package. version 1.3.1. Buffalo, NY: University at Buffalo. http://github.com/trinker/qdap.

[14] Tun, PA., O’Kane, G., Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychol Aging*, 17, 453-467.

[15] 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. *Language and Speech*, 53, 510-540.