Timing of turn taking between normal-hearing and hearing-impaired interlocutors

A. JOSEFINE MUNCH SØRENSEN[1,*], EWEN N MACDONALD[1], THOMAS LUNNER[1,2]

[1] *Department of Health Technology, Technical University of Denmark (DTU), DK-2800 Kgs. Lyngby, Denmark*

[2] *Eriksholm Research Centre, Oticon A/S, DK-3070 Snekkersten, Denmark*

Having a conversation requires more resources than just understanding speech. Previous studies of the timing of turn taking in conversations suggest that in order to sustain normal, fluid turn taking, interlocutors have to predict the end of each other's turns. Thus, while noise and hearing loss should make understanding speech more difficult, it should also reduce the resources available for speech planning and possibly reduce the saliency of cues used to predict turn ends, resulting in delayed and more variable turn taking. We recorded conversations between 12 pairs of native-Danish young normal-hearing (NH) and older hearing-impaired (HI) listeners with mild presbyacusis in quiet and multitalker babble at three levels. The interlocutors conducted a Diapix task, finding differences in two near-identical pictures. Both HI and NH talkers responded more slowly and with more variability with increasing noise level, and the HI with more variability than the NH. We saw indications that the younger NH adopted a more careful communication strategy, likely to ease the effort on their older HI interlocutor, by adapting their speech rates to their interlocutor and overlapping less.

INTRODUCTION

Traditionally, speech understanding and production is studied in isolation where people are passively listening and reporting back what they heard, or producing speech with no addressee. However, real communication is not just the sum of production and listening, it is an interaction between two or more participants who use dynamic feedback and adaptation to increase understanding and information sharing. Recent studies, however, seek to measure speech understanding and production simultaneously in studies of conversational interaction (e.g., Beechey *et al.*, 2018; Hadley *et al.*, 2019). In this study, we investigated conversational turn-taking between younger normal-hearing (NH) and older hearing-impaired (HI) interlocutors solving the Diapix task (Baker and Hazan, 2011) in quiet and in three levels of a multitalker babble noise: 60, 65, and 70 dBA SPL. Earlier studies of conversational interactions suggest that interlocutors predict the end of their partner's turn to sustain normal, rapid turn-taking (e.g., Levinson and Torreira, 2015). We hypothesised that hearing loss and noise interference should increase listening difficulty, reducing the resources available

*Corresponding author: ajso@dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

for speech planning and reducing the saliency of predicting cues, resulting in delayed and more variable response times.

METHOD

Participants

Twelve unacquainted mixed- and same-gender pairs of younger normal-hearing (NH) and older hearing-impaired (HI) interlocutors were recruited (9 females, 7 mixed-gender pairs). The NH participants ($\mu = 26$ years, $\sigma = 2.7$ years) had hearing threshold levels below 20 dB HL between 125 Hz and 8 kHz. The HI participants ($\mu = 73$ years, $\sigma = 4.4$ years) had mild presbyacusis (see Figure 1 for their audiograms), and were unaided during the experiment. All participants provided informed consent and the experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). The participants were compensated for their time.

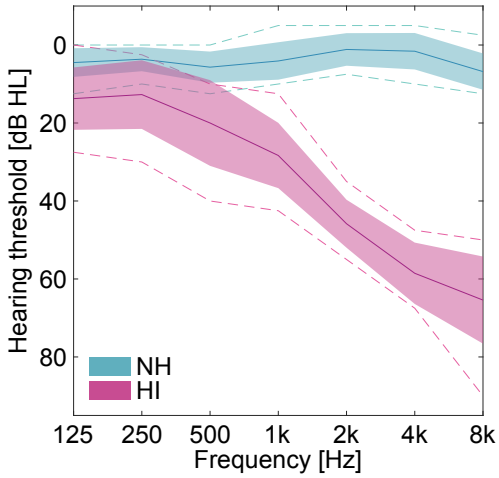

Fig. 1: Audiometric thresholds of the younger normal-hearing and older hearing-impaired listeners. The solid line indicates the mean hearing threshold, the coloured regions indicate one standard deviation, and the dotted lines indicate minimum and maximum measured thresholds.

Setup

Seated in separate booths, the participants wore Shure PGA31 wireless cardioid microphones (transmitted by Shure GLXD14 wireless system) and Sennheiser HD650 open headphones, over which they communicated with each other. The gains were calibrated such that the resulting presentation levels over the headphones were the same as the A-weighted broadband levels one meter away from the talker in the same room. A 20-talker babble was created by taking 20 minutes of recordings from 20 talkers balanced in genders from the NH/NH recordings from Sørensen *et al.* (2020).

The recordings were normalized to the same RMS level as the recording with lowest RMS level, and pauses were removed using voice activity detection (VAD). Finally, they were added together. Auditive verification ensured it was impossible to resolve any content from the individual talkers.

Task and procedure

Similarly to the task in Sørensen *et al.* (2020), participants solved the DiapixUK task (Baker and Hazan, 2011) to elicit dialogue. A training round was conducted outside the booths to familiarize the participants with the task. Inside the booths, the participants had another test round with 65 dBA SPL background noise to familiarize them with the setup and procedure. In the test, the participants solved the Diapix task in three replicates of four conditions: quiet, 60, 65 and 70 dBA SPL background noise. The order of the conditions was randomized within each replicate, and they had a break in between each replicate. The participants were given a maximum of 10 minutes to find 10 differences between the Diapix.

Analysis of recordings

During a turn-taking there is a change in the conversational floor termed a floor-transfer. The duration of such a floor-transfer is termed a floor-transfer offset (FTO) measured from the offset of one person's speech to the onset of the next person's speech. This can either be negative, termed an overlap-between, or positive, termed a gap. Following the procedure in Sørensen *et al.* (2020), each of the conversations were categorized into conversational states: 1) gaps, 2) overlaps-between, 3) utterances, which are speech tokens separated by silence of less than 180 ms, 4) pauses, which are joint silence not followed by a floor-transfer, and 5) overlaps-within, which are joint speech during utterances of one talker that does not result in a floor-transfer.

Mixed-effects regression models were fit to the variables in *R* using the *lme4* package, with background, hearing and replicate as main effects, and pair as random intercept. Denominator degrees-of-freedom were Satterthwaite approximated for the F-tests for the fixed effects. Pairwise comparisons were computed using the *lsmeans* function (*lmerTest* package) comparing least-squares means of the significant effects using the Satterthwaite approximated df.

RESULTS

The average speaking levels of the participants is seen in Figure 2, left panel. A random intercept for participants was added to the mixed effects model. All participants increased their speaking levels significantly in background noise $[F(3,258) = 584.95, p < 2.2\text{e-}16]$, and there was a significant interaction between hearing status and background $[F(3,269) = 14.54, p < 8.91\text{e-}9]$. A multiple comparison post-hoc analysis revealed that the difference was driven by the level differences in quiet between NH and HI, where the HI spoke significantly louder than the NH $[t(24.7) = -2.091, p < 0.047]$.

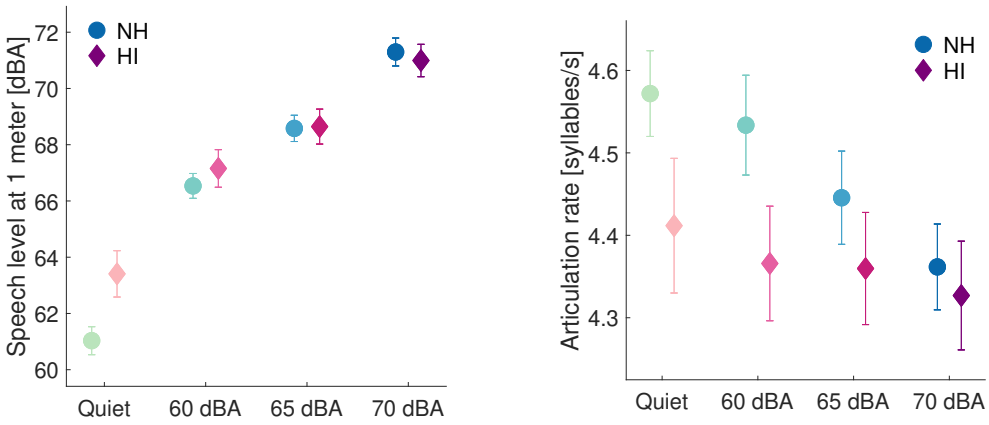

Fig. 2: Speech level (left panel) and articulation rate (right panel) in the four backgrounds (quiet, 60, 65 and 70 dBA SPL) for younger normal-hearing and older hearing-impaired listeners. The bars indicate standard error.

In Figure 2, right panel, the participant's articulation rates for the NH and HI in the four backgrounds is plotted (computed using the Praat script presented in de Jong and Wempe (2009) with default parameter settings). Again, a random intercept for participants was added. There was a significant main effect of background $[F(3,258) = 10.97, p < 8.88\text{e-}7]$, and a significant interaction between hearing and background $[F(3,258) = 2.75, p < 0.043]$. With increasing noise level, the articulation rates of both groups of talkers decreased, and the articulation rates of the NH talkers approached those of the HI.

As an indication of who tended to dominate the conversation, the average proportion of time each person in the two hearing status groups was speaking was computed and can be seen in the left panel of Figure 3. The proportion is measured as the total duration of active speech from the participant (determined by VAD) divided by the total duration of active speech in the conversation from both participants. There was a statistically significant effect of hearing, with the HI speaking more than the NH: $[F(1,286) = 80.4, p < 2.2\text{e-}16]$.

The HI group produced more overlaps-within than the NH and for both groups the rate of overlaps-within decreased with increasing background noise level, confirmed by a statistically significant main effect of both hearing $[F(1,272) = 5.44, p < 0.0204]$ and background $[F(3,272) = 7.47, p < 8.05\text{e-}5]$.

The FTO distributions, collapsed across participants within the NH and HI groups, in the four backgrounds, are seen in the left panel of Figure 4. The distributions were estimated using 100 ms bin widths. By visual inspection, the distributions seem broader for the HI group than the NH group, and slightly broader with increasing noise level. A two-sample Kolmogorov-Smirnov test rejected the null-hypothesis that the samples came from the same distributions for the NH and HI in the four conditions:

[$D = 0.103, p <$ 2.2e-16] in quiet, [$D = 0.103, p <$ 1e-13] in 60 dBA noise, [$D = 0.088, p <$ 1.18e-10] in 65 dBA noise and [$D = 0.096, p <$ 1.4e-12] in 70 dBA noise.

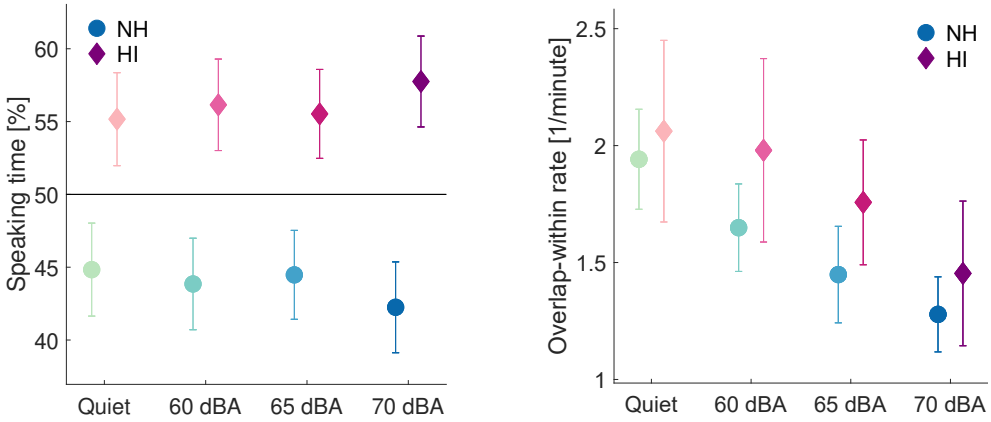

Fig. 3: Speaking time (the percentage of time a person speaks during the conversation) (left panel) and rate of occurrence of overlaps-within (i.e., turns from one talker that occur completely within a turn of the other talker) (right panel) in the four backgrounds (quiet, 60, 65 and 70 dBA SPL) for younger normal-hearing and older hearing-impaired listeners. Note that the rate has been normalized by the total phonation time rather than duration of the conversation. The bars indicate standard error.

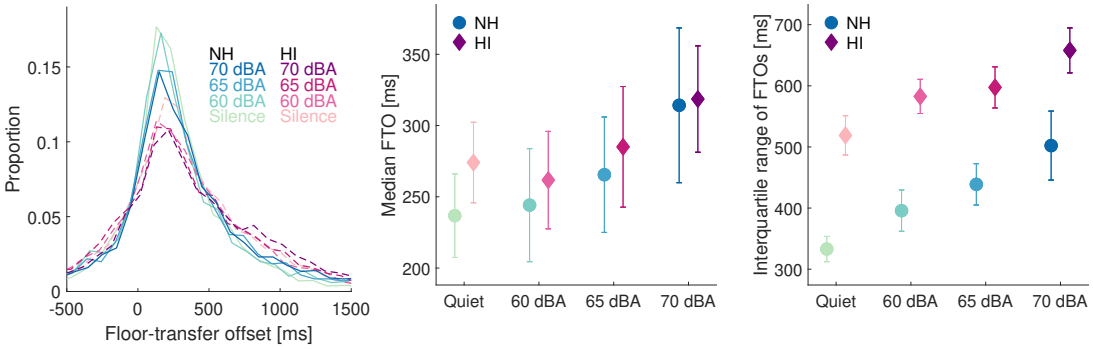

Fig. 4: Normalized distributions (left panel) of floor-transfer offsets (FTOs) along with the median (middle panel) and interquartile range (right panel) for the four combinations of language and noise. The bars indicate standard error.

In the middle and right panels of Figure 4, the median FTO and interquartile range (IQR) of FTOs are plotted. There was a statistically significant main effect of background on the median FTO [$F(3,261) = 10.89, p <$ 9.14e-7], but no significant main effect of hearing status or replicate and no interactions. For the IQR, there

were significant main effects of both background [$F(1,272) = 15.4, p < 2.78\text{e-}9$] and hearing status [$F(1,272) = 135.4, p < 2.2\text{e-}16$], confirming the visual impression that the distributions were broader for the HI group.

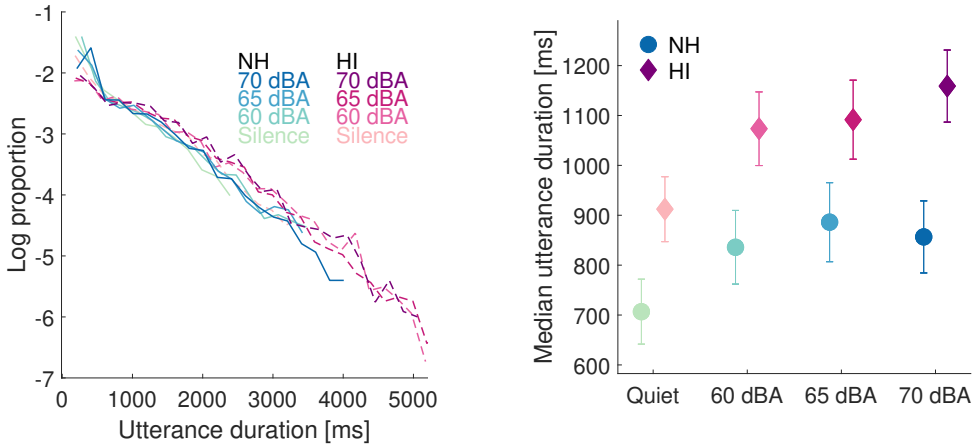

Fig. 5: Normalized distributions (left panel) and median (right panel) of the duration of utterances for younger normal-hearing and older hearing-impaired talkers in the four backgrounds. The bars indicate standard error. The density in the left panel has been log-transformed to more easily compare the slopes.

The distributions of utterances (estimated using 200 ms bin widths) along with the median utterance duration for the two hearing statuses and four backgrounds are seen in the left and right panels of Figure 5, respectively. There was a significant main effect of both hearing status [$F(1,272) = 52.48, p < 4.48\text{e-}12$] and background [$F(3,272) = 7.49, p < 7.7\text{e-}5$] on the median utterance duration, with both groups increasing their utterance durations in noise and the HI group producing about 25 % longer utterances than their NH interlocutor.

DISCUSSION

Speaking levels

On average, the older HI spoke about 2.5 dB louder than the younger NH when there was no background noise, but the two groups increased their speaking levels to achieve almost the same SNR in background noise. With increasing noise level, there was a decrease in the SNR. At 60, 65, and 70 dBA SPL, the average SNR was 7, 3.5 and 1 dB, respectively. It is physically strenuous to speak at a high sound pressure level, so talkers may trade off speech understanding and physical effort. However, in all conditions the participants spoke at positive SNRs, whereas in the NH/NH conversations in Sørensen *et al.* (2020) the participants spoke at -2.5 dB SNR in 70 dBA noise. This shows adaptive behavior from the younger NH talker to their older HI interlocutor, adjusting to their hearing difficulty. The HI may both speak at a

positive SNR for their own auditory feedback, but also to signal difficulty so that their interlocutors increase their voice level.

Utterances

Similarly to the experiments with NH/NH of Sørensen *et al.* (2020) and Watson *et al.* (2020), the participants lengthened their utterances in noise. This may be a strategy to give themselves and their interlocutor more time to plan their response. It may also be a strategy to meet appropriate response times by initiating turns while still not fully planned, resulting in lengthened turns. While the NH increased their duration of utterances to the same extent as those in Sørensen *et al.* (2020), the HI lengthened their turns significantly more. This supports the interpretation that the older HI talkers are more challenged than the younger NH talkers. From the distributions of utterances it seems that the overall utterance duration of the older HI are longer, indicated by the shallower slope. An immediate interpretation is that it could be explained by the lower articulation rate of the older HI. However, the articulation rates of NH are similar to those of the older HI in the 70 dBA condition, yet utterance durations remain different.

Articulation rates

In other studies, when talking to an NH partner, NH talkers increased their articulation rates in noise (Sørensen *et al.*, 2020; Watson *et al.*, 2020). However, in this study we saw the opposite trend: with increasing noise level, the NH participants decreased their articulation rates. In general, the HI talkers spoke slower than the NH talkers. This may just be an effect of age, but it may also be a signalling strategy to their NH interlocutors to slow down articulation to ease speech understanding and reduce the communication challenge for the HI interlocutor.

Overlaps-within

Another indication of the NH's adaptive and accommodating behaviour is the rate at which overlaps-within occur. With increasing background noise level, the rate goes down for both groups. In both Sørensen *et al.* (2020) and Watson *et al.* (2020), when talking to an NH partner, the NH increased their rate of overlaps-within in background noise, regardless of whether they were unacquainted or not, or if they spoke in free conversation or solved the Diapix task. This was attributed to an increased stress. However, more overlaps are likely to decrease the information transmission and increase the cognitive load on participants. It was observed in Watson *et al.* (2020) that when participants spoke freely, they had a higher rate of overlaps-within than when they solved the Diapix task, where information transfer is presumably more crucial. Here, the younger NH may have adopted an even more careful turn taking strategy in conversations with older HI talkers to increase information transfer. This may also be why we found a delay in FTOs, not just because of increased cognitive load, but also to actively reduce overlaps that likely reduce speech understanding.

SUMMARY

For both the younger NH and older HI talkers, floor-transfer offsets were longer and more variable in background noise, and the older HI showed more variability than the NH talkers. There were indications that the NH adapted their speech to accommodate their interlocutor's difficulty. For example, they adapted their speaking rates to speak slower with increasing noise level, opposite to what was found in our previous studies with NH interlocutors (Sørensen *et al.*, 2020; Watson *et al.*, 2020). Moreover, both groups decreased their rates of overlaps-within with increasing noise level, but the NH decreased their rates significantly more than the older HI. The NH also produced significantly shorter utterances than the HI and spoke less of the time than the HI.

ACKNOWLEDGEMENTS

A.J.M.S. and a portion of this study was supported by the William Demant Foundation (16-3968).

REFERENCES

Baker, R., and Hazan, V. (**2011**). "DiapixUK: Task Materials for the Elicitation of Multiple Spontaneous Speech Dialogs," Behav. Res. Methods, **43**(**3**), 761–70. doi: 10.3758/s13428-011-0075-y.

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

de Jong, N. and Wempe, T. (**2009**). "Praat script to detect syllable nuclei and measure speech rate automatically," Behav. Res. Methods, **41**, 385-90. doi: 10.3758/BRM.41.2.385.

Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (**2019**). "Speech, movement, and gaze behaviours during dyadic conversation in noise," Sci. Rep., **9**(1), 10451. doi: 10.1038/s41598-019-46416-0.

Levinson, S. C., and Torreira, F. (**2015**). "Timing in turn-taking and its implications for processing models of language," Front. Psychol., **6**, 731. doi: 10.1038/s41598-019-46416-0.

Sørensen, A. J. M., Fereczkowski., M, and MacDonald, E. N. (**2020**). "Effects of noise and L2 on the timing of turn taking in conversation," Proc. ISAAR, **7**, 85-92.

Watson, S., Sørensen, A. J. M., and MacDonald, E. N. (**2020**). "The effect of conversational task on turn taking in dialogue," Proc. ISAAR, **7**, 61-68.