Preliminary investigation of the categorization of gaps and overlaps in turn-taking interactions: Effects of noise and hearing loss

A. JOSEFINE SØRENSEN[1,2,*], ADAM WEISSER[2], AND EWEN N. MACDONALD[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *GN ReSound A/S, Ballerup, Denmark*

Normal conversation requires interlocutors to monitor the ongoing acoustic signal to judge when it is appropriate to start talking. Categorical thresholds for gaps and overlaps in turn-taking interactions were measured for normal-hearing and hearing-impaired listeners in both quiet and multitalker babble (+6 dB SNR). The slope of the categorization functions were significantly shallower for hearing impaired listeners and in the presence of background noise. Moreover, the categorization threshold for overlaps increased in background noise.

INTRODUCTION

In normal conversation, talkers take turns speaking in a manner that is flexible (i.e., not organized in advance) and often rapid. Over the decades since the seminal paper by Sacks *et al.* (1974) was published, many models have been proposed to explain why the switching between interlocutors remains fluid with rapid transitions between speakers. In general, all of these models involve interlocutors monitoring aspects of the ongoing acoustic signal to judge when the current talker will stop or has stopped talking (for further discussion and review including the distribution of acoustic overlaps and gaps in normal turn taking see Heldner and Edlund, 2010; Levinson and Torreira, 2015). Background noise and hearing impairments are known to reduce many aspects of auditory perception (e.g., speech intelligibility) and can reduce cognitive spare capacity (i.e. resources for higher-level processing of speech; Rudner and Lunner, 2014). Thus, it is possible that noise and/or hearing loss may alter normal communication dynamics by altering the perception of acoustic cues monitored by interlocutors (for a discussion of these cues, see Gravano and Hirschberg, 2011).

As a first step towards investigating this, the present study focused on the perception of turn-taking interactions in a manner similar to Heldner (2011). Specifically, normal-hearing and hearing-impaired listeners were asked to listen to pre-recorded turn-taking interactions and categorize whether the interactions were perceived as overlaps (i.e., the second talker started before the first had finished), gaps (i.e., there was

*Corresponding author: ajs@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

silence/pause between when the first talker stopped and the second talker started), or neither, when the acoustic interval between the speech offset of the first talker and the speech onset of the second talker was systematically varied from −500 ms to 500 ms.

METHOD

The participants in this study were 24 normal-hearing (mean age 36 years) and 7 hearing-impaired (mean age 72 years) native Danish listeners. The hearing impaired group had primarily moderate hearing loss (see Fig. 1). The procedure was approved by the Science-Ethics Committee for the Capital Region of Denmark. A speech corpus was developed based on the dialogue from a Danish translation of the play "Educating Rita" (Author: Willy Russell, translator: Riri Ianke Firing), consisting of a conversation between a man and a woman. In addition to portions of the script, 11 turn-taking interactions were created from everyday conversation topics. Overall, the corpus consisted of 85 turn-taking interactions, 43 where the woman starts and the man takes over and 42 in the opposite order. Of these turn-taking interactions, 44 were in the form of a question-answer, 7 in the form of statement-question, 29 in the form statement-answer, and 5 in the form of statement-statement.

The corpus was recorded in an audiometric sound booth, with the male and female native Danish talkers standing approximately 1 m apart from each other. The speech was recorded using Cubase Elements 7 (Steinberg) with microphones (AKG C451B) with pop filters placed approximately 10 cm from each talker and connected to a Fireface UFX (RME) soundcard. The talkers were instructed to read the sentences as naturally as possible, but to pause for a second or two during the turn-taking interaction to avoid crosstalk. Each turn-taking interaction was repeated twice. The recordings were reviewed to select the best utterances for each turn-taking interaction. These were the utterances that were relatively constant in level and for which the interaction sounded the most natural to the first author. After selecting the best utterances, they were segmented by hand using Praat (Boersma and Weenink, 2002). The mean duration of each utterance was 1.35 s (standard deviation 0.49 s) and varied between 0.5 and 2.95 s. After editing, the sentences were normalized to have the same RMS level.

Categorization of overlaps, gaps, and no-gap-no-overlap was obtained using a 3-alternative forced-choice paradigm for 17 acoustic intervals, ranging from -500 to 500 ms. Here, we use the convention that a negative acoustic interval corresponds to an overlap (i.e., the onset of speech from the second talker occurs before the offset of speech from the first talker) and a positive interval corresponds to a silence. We used a three-category procedure rather than conducting a two-category procedure twice (i.e., gap vs. no gap; overlap vs. no overlap) to ensure listeners would maintain the same internal criteria they use when listening to regular conversations. On each trial, participants listened to a turn-taking interaction that was composed from an edited pair of utterances that were shifted in time and then mixed to create the desired

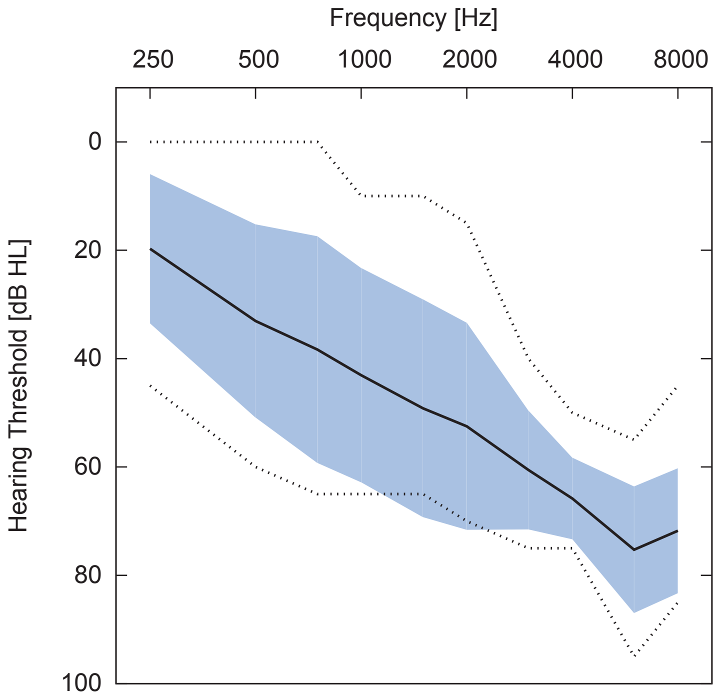

Fig. 1: Audiometric thresholds of the hearing-impaired listeners. The solid black line indicates the mean hearing threshold, the shaded region indicates one standard deviation, and the dotted lines indicate minimum and maximum measured thresholds.

acoustic interval. After listening to a turn-taking interaction, the participant pressed one of three buttons on a computer screen. The buttons were labeled with Danish text corresponding to the English terms overlap, gap, and no-gap-no-overlap. For each listener, five judgments of each interval in both quiet and in a background of seven-talker English babble[1] (mixed to achieve an SNR of +6 dB) were obtained. The order in which the acoustic intervals were presented was randomized across trials. Thus, the utterances judged for each acoustic interval varied across listeners. Half of the listeners judged turn-taking interactions in quiet before judging turn-taking interactions in babble. The other half judged the two conditions in the opposite order.

Stimuli were presented over headphones (Sennheiser HD 650) in a sound booth. For the normal hearing listeners, the stimuli were presented at 65 dB SPL. To compensate for reduced audibility, the stimuli (i.e., speech and noise in the babble condition and speech alone in the quiet condition) was further amplified using the Cambridge Formula (Moore and Glasberg, 1998) for each individual hearing impaired listener. Overall, it took each listener approximately 15-20 minutes to complete the experiment.

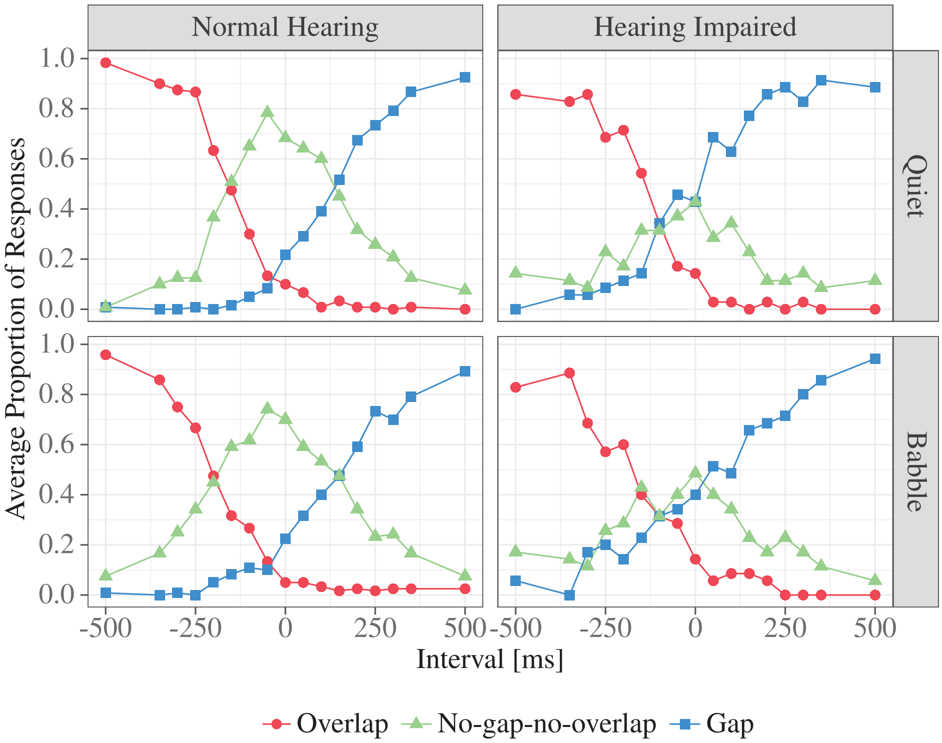

Fig. 2: The average proportion of responses of overlap (squares), gap (circles), and no-gap-no-overlap (triangles) as a function of acoustic interval. A negative acoustic interval indicates acoustic overlap (i.e., an interval where both talkers are speaking) while a positive interval indicates an acoustic gap. The top and bottom panels present the results for the normal and impaired listeners, respectively. The left and right panels present results when the turn-taking stimuli were presented in quiet and in multitalker babble (with an SNR of +6 dB), respectively.

RESULTS

The average proportion of responses of overlap, gap, and no-gap-no-overlap as a function of acoustic interval is plotted in Fig. 2. To estimate categorical thresholds, cumulative Gaussian functions were fitted to individuals' data after smoothing using a simple moving average of responses from three neighbouring acoustic intervals. The average mean (which is used as the estimated category threshold for that individual) and standard deviation of the fitted Gaussian functions are plotted in Fig. 3. Note that in this figure, the sign of the means for overlap results have been inverted to better compare their magnitudes with the results for gap categorization.

Repeated measures ANOVAs were conducted on the fitted means of categorizing overlaps and gaps, with background condition (quiet vs. babble) as within-subject

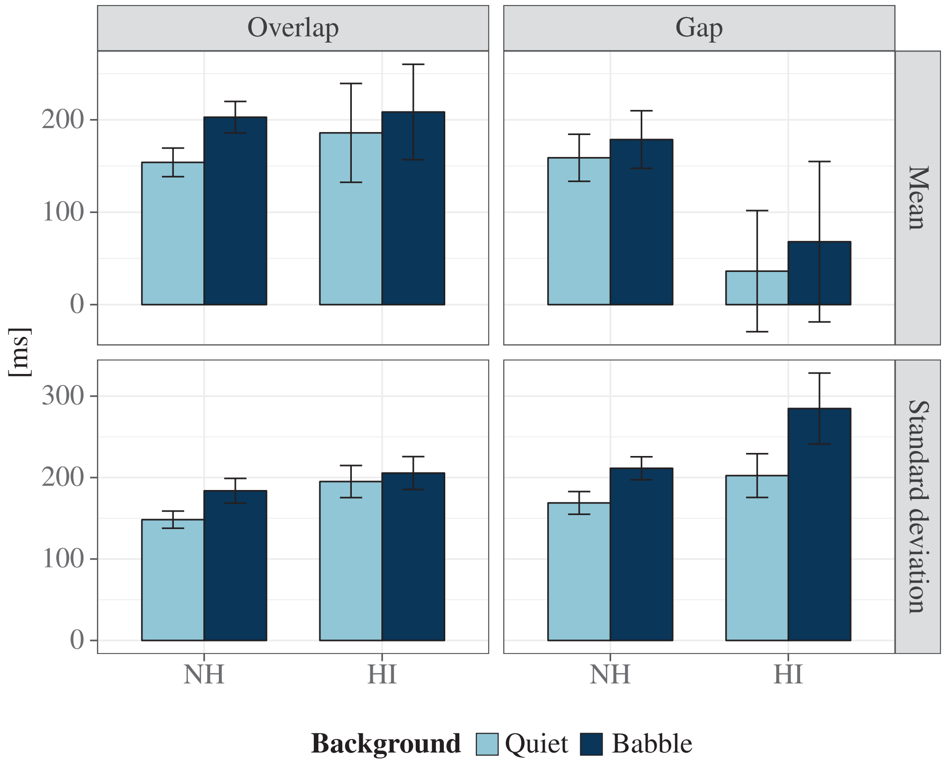

Fig. 3: Average mean (top panels) and standard deviation (bottom panels) of cumulative Gaussian functions fitted to individual's proportion of responses. The left and right panels present results for overlap and gap categorizations averaged within groups of normal hearing (NH) and hearing impaired (HI) listeners respectively. Here, the sign of the means for overlap results have been inverted to better compare their magnitudes with the results for gap perception. The bars indicate one standard error.

and hearing status (normal hearing vs hearing impaired) as between-subjects factors. For the overlap categorization, only the main effect of background condition was significant [$F(1,29) = 5.782, p < 0.023$]. For the gap categorization there was no significant effect of either background condition, or hearing status, but there was a trend for the effect of hearing status [$F(1,29) = 3.329, p < 0.078$]. None of the interactions were significant.

A repeated measures ANOVA on the fitted standard deviations (slopes) of the categorization functions for overlaps and gaps was conducted with background condition (quiet vs. babble) and category (gap vs. overlap) as within-subject factors and hearing status (normal hearing vs hearing impaired) as between-subjects factor. Main effects of background condition [$F(1,29) = 8.455, p < 0.01$], category [$F(1,29) =$

$11.565, p < 0.01$], and hearing status were significant [$F(1,29) = 4.524, p < 0.05$]. None of the interactions were significant.

When compared with the results from the normal hearing listeners, the proportion of responses for the no-gap-no-overlap from the hearing-impaired listeners is greatly reduced (see Fig. 2). To quantify this, the average proportion of responses of no-gap-no-overlap for acoustic intervals ranging from -150 to 150 ms (i.e., acoustic intervals between the average categorical thresholds of overlap and gap in quiet by normal-hearing listeners) were calculated for each individual in each condition. A repeated measures ANOVA with background condition as within- and hearing status as between-subjects factors confirmed a main effect of group [$F(1,29) = 9.174, p < 0.005$] such that the hearing-impaired listeners used the no-gap-no-overlap category less frequently than the normal-hearing listeners.

DISCUSSION

Virtually all models of turn taking in conversation involve interlocutors monitoring the ongoing acoustic signal. Thus, it is possible that the presence of background noise or hearing loss could influence conversational dynamics by disrupting the perception of acoustic cues monitored by interlocutors. In the present study, categorical thresholds for perceiving a turn-taking interaction with an acoustic overlap as an overlap increased when listening in the presence of a background noise. This effect was observed, even though the level of the multitalker babble was relatively low (SNR of +6 dB) and should not have decreased the intelligibility of the turn-taking interaction utterances. The slopes of the categorization functions of the hearing-impaired listeners were shallower than those of the normal-hearing listeners.

In a previous study, Heldner (2011) measured gap and overlap thresholds separately using a two-alternative forced-choice procedure (i.e., turn-taking interactions with positive acoustic intervals were judged as either gap or no-gap, whereas turn-taking interactions with negative acoustic intervals were judged as either overlap or no-overlap). In the present study, the results of the normal-hearing listeners suggested they perceived three clear categories (see Fig. 2) and the categorical thresholds were consistent with the detection thresholds reported by Heldner (2011). The hearing-impaired listeners exhibited a much lower proportion of responses for no-gap-no-overlap than the normal hearing listeners for small acoustic intervals (i.e., between -150 and 150 ms). Thus, it is not clear if this between-group difference is because the hearing-impaired listeners perceived only two categories (i.e., either gap or overlap) or the task was too difficult (i.e., choosing between three rather than two categories). It should be noted that the hearing impaired listeners were much older than the normal hearing listeners. Thus, the differences observed could also be due to aging effects (both in auditory perception and cognition) rather than or in addition to hearing loss. The hearing-impaired group's small detection thresholds for gaps was therefore not attributed to a higher sensitivity, but rather an effect of poor categorization as indicated by their significantly shallower slopes and their inability to use the no-gap-no-overlap category. A less cognitively demanding follow up study using two 2-alternative forced-choice paradigms, as in Heldner (2011), is needed to help disentangle these effects.

The corpus used in this study was recorded from scripted dialog rather than the spontaneous conversations used in previous studies into the perception of gaps in turn-taking interaction (e.g., Walker and Trimboli, 1982; Heldner, 2011). Further, in the present study, the recorded turn-taking interactions were edited to systematically vary the acoustic interval. In contrast, previous studies have presented the original acoustic recordings, which contained a wide range of acoustic intervals. Nevertheless, the thresholds measured in this study are consistent with these previous studies. For normal-hearing listeners, the average threshold for categorizing a gap in quiet was approximately 160 ms. This is close to the thresholds of 120 ms reported by Heldner (2011) and 180 ms estimated[2] from the data published by Walker and Trimboli (1982). Similarly, for normal-hearing listeners, the average threshold for categorizing an overlap in quiet was approximately 155 ms, which is close to the 120 ms reported by Heldner (2011) and indicates the same symmetry between the gap and overlap thresholds (i.e., a duration of about one syllable in order to perceive it as either a gap or an overlap).

CONCLUSION

In the present study, thresholds for perceiving overlaps and gaps were obtained for both normal-hearing and hearing-impaired listeners in both quiet and noise. The results indicated that the threshold for perceiving an overlap increased in the presence of background noise. Furthermore, the categorization functions for both gaps and overlaps were shallower in the presence of background noise and for hearing impaired compared to normal hearing listeners. The gap categorization was very different for the hearing-impaired group, and it is not clear if it is an effect of the paradigm or some higher order processing. Thus it is suggested that a follow-up study using a different paradigm is needed to explore this. In conclusion the presence of background noise influences the perception of acoustic cues used to judge turn-taking interaction. This suggests that comparing turn taking in quiet and noisy conditions may be useful for validating proposed models of turn taking, particularly in relation to language processing (e.g., Levinson and Torreira, 2015). Furthermore, it suggests that parts of the language processing system is affected even at SNRs well above the typical SRT, and that these effects are not captured by standard speech intelligibility tests.

ACKNOWLEDGEMENTS

This experiment was conducted as a bachelor project by the first author in collaboration with GN ReSound. The authors would like to thank Andreas Schousboe from GN ReSound A/S for his help in facilitating the project, as well as Johan Gabriel Selby, Lotte Hagen Hernvig, and Brent Kirkwood, all from GN ReSound A/S, for their help with the experiment.

ENDNOTES

[1] Seven-talker babble with both male and female voices was created from an unpublished dialogue and trialogue corpus created previously by the second author. The levels of all talkers in the babble were adjusted to have the same RMS.

[2] The categorical threshold corresponding to 50% proportion of responses was estimated from a linear interpolation of the proportion of responses for 100 and 200 ms (42% and 52%, respectively).

REFERENCES

Boersma, P., and Weenink, D. (**2002**). "Praat, a system for doing phonetics by computer," Glot Int., **5**, 341-345.

Gravano, A., and Hirschberg J. (**2011**). "Turn-taking cues in task-oriented dialogue," Comput. Speech Lang., **25**, 601-634. doi: 10.1016/j.csl.2010.10.003

Heldner, M., and Edlund, J. (**2010**). "Pauses, gaps, and overlaps in conversations," J. Phonetics, **38**, 555-568. doi: 10.1016/ j.wocn.2010.08.002

Heldner, M. (**2011**). "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," J. Acoust. Soc. Am., **130**, 508-513. doi: 10.1121/1.3598457

Levinson, S.C., and Torreira, F. (**2015**). "Timing in turn-taking and its implications for processing models of language," Front. Psychol., **6**, 731. doi: 10.3389/fpsyg.2015.00731

Moore, B.C.J., and Glasberg, B.R. (**1998**). "Use of loudness model for hearing-aid fitting. I. Linear hearing aids," Br. J. Audiol., **32**, 317-335. doi: 10.3109/03005364000000083

Rudner, M., and Lunner, T. (**2014**). "Cognitive spare capacity and speech communication: A narrative overview," BioMed Res. Int., 869726. doi: 10.1155/2014/869726

Sacks, H., Schegloff, E.A., and Jefferson, G. (**1974**). "A simplest systematics for the organization of turn-taking for conversation," Language, **50**, 696-735. doi:10.2307/412243

Walker, M.B., and Trimboli, C. (**1982**). "Smooth transitions in conversational interactions," J. Soc. Psychol., **117**, 305-306. doi: 10.1080/00224545.1982.9713444