DOI: 10.61782/fa.2023.0295

EFFECTS OF THE AUDITORY ENVIRONMENT ON THE FUNDAMENTAL-FREQUENCY DYNAMICS OF SPEECH

P. A. Mesiano[1,2*] **J. Zaar**[1,2] **H. Relaño-Iborra**[2]
L. Bramsløw[1] **T. Dau**[2]
[1] Eriksholm Research Centre, Rørtangvej 20, 3070 Snekkersten, Denmark
[2] Hearing Systems Section, Technical University of Denmark, Ørsteds plads, 352, 2800 Kgs. Lyngby, Denmark

ABSTRACT

The fundamental frequency (F_0) is one of several voice features that talkers can alter in adverse communicative situations, e.g., when speaking in the presence of background noise or with hearing-impaired (HI) interlocutors, with the intent to produce 'clear speech' and increase intelligibility. Compared to 'conversational speech', clear speech has been shown to be characterized by a higher average F_0 and a larger F_0 dynamic range. However, these changes in F_0 have commonly been measured in laboratory simulations with normal-hearing (NH) talkers speaking in the absence of an interlocutor. In contrast, the present study explored changes in F_0 occurring in clear speech by analyzing the F_0 statistics of naturalistic dialogues between NH interlocutors as well as between NH and HI ones, conducted in quiet and in background noise. It was found that (i) in the presence of background noise, both NH and HI talkers increased their average F_0 and F_0 dynamic range, and that (ii) when speaking with a HI interlocutor, NH talkers exhibited a higher average F_0 and a larger F_0 dynamic range than when speaking with a NH interlocutor. These results provide further evidence of how talkers modify their voice and speaking style depending on the auditory environment.

Keywords: *fundamental frequency, speech production, clear speech, hearing loss*

*Corresponding author: poao@eriksholm.com.

1. INTRODUCTION

Oral communication can be challenged by factors related to the acoustic environment, such as background noise or the presence of interfering talkers, and the hearing abilities of the interlocutors (e.g., in the case of hearing-impaired individuals). To promote the intelligibility of their speech, the talker may adapt aspects of speech production to enhance the auditory and linguistic cues that can facilitate the listening task for the interlocutor, i.e., by producing the so-called 'clear speech' [1-2]. It has been demonstrated that clear speech has a positive impact on speech understanding in challenging acoustic situations as it is more intelligible than conversational speech [1].

Several acoustic properties of the voice can be altered by the talker to make their speech 'clear' in response to the acoustic environment [2]. One such property of speech is the fundamental frequency (F_0). Compared to conventional speech, clear speech has been shown to have a higher average F_0 and a wider F_0 dynamic range.

However, the available evidence for F_0 adaptations observed in clear speech has been based on results from small populations of normal-hearing (NH) talkers with a strong focus on specific communicative situations, such as mothers speaking to infants, and by analyzing speech recorded in the absence of an interlocutor, where the talker was asked to speak 'as if' they were in situations with specific communication barriers.

The present study further explored adaptations in F_0 that talkers apply to overcome certain communicative obstacles. A comparative analysis of the F_0 information measured in speech obtained from laboratory recordings of naturalistic dialogues is presented. The dialogues were conducted in pairs of NH interlocutors as well as in pairs of a NH and a HI one, in quiet and in different noisy conditions.

2. METHODS

The speech material consisted of laboratory-recordings of dialogues in the Danish language conducted by pairs performing the Diapix task [3]. 19 pairs consisted of NH interlocutors (from a group indicated as NH1) while 12 other pairs consisted of a NH interlocutor (belonging to a different group, indicated as NH2) speaking with a HI interlocutor (indicated as HI). The dialogues were conducted in quiet or with different levels of noise. The HI talkers were 'unaided', i.e., they were not using hearing aids. The recordings of the voice of each talker from the dialogues were labelled as talker-listener combination, i.e., NH1-NH1, NH2-HI and HI-NH2. The dialogues between NH1 interlocutors were conducted in quiet or in the presence of speech-modulated noise at 70-dBA sound pressure level. The dialogues between NH2 and HI interlocutors were conducted in quiet and in presence of 20-talker babble noise at 60-, 65- and 70-dBA sound pressure level. Each pair of interlocutors recorded three dialogue sessions, lasting 5':58'' on average. The recordings were obtained as part of previous studies ([4] for NH1-NH1 and [5] for NH2-HI). For this reason, the NH1-NH1 and NH2-HI recordings differed in NH population, population sizes and in the background noise conditions employed.
The F_0 of the speech material was analyzed as follows. For each talker-listener combination and background-noise condition, the F_0 trajectories from the speech signal of the talker in the different recording sessions were extracted with the software PRAAT and concatenated to obtain a unique F_0 trajectory. The long-term statistics of the F_0 of each talker in each condition were quantified in terms of average F_0 (calculated as the median value of the trajectory and indicated as $\overline{F_0}$) and F_0 dynamic range (calculated as the median absolute deviation of the trajectory and indicated here as $\sigma(F_0)$). The $\overline{F_0}$ and $\sigma(F_0)$ were compared across the different talker-listener combinations and background-noise conditions.

3. RESULTS

Fig. 1 shows the $\overline{F_0}$ of the individual talkers for each dialogue condition (talker-listener combination and background-noise condition), indicated by open black circles, as well as the corresponding group-average $\overline{F_0}$ and standard errors, indicated by filled red circles. The $\overline{F_0}$ was significantly higher in presence of 70-dBA background noise than in quiet ($p<10^{-3}$), with an average $\overline{F_0}$ difference between the two conditions of 26 Hz (2.8 semitones) for NH1-NH1, 36 Hz (3 semitones) for NH2-HI and 35 Hz (3.6 semitones) for HI-NH2 talkers. For NH2-HI and HI-NH2 (where recordings were available at three different levels of background noise), the increase in $\overline{F_0}$ was monotonic with increasing noise level. With very few exceptions, the increases in $\overline{F_0}$ for increasing level of background noise were consistent across talkers within each group. NH1-NH1 and HI-NH2 did not differ significantly in $\overline{F_0}$ in the quiet nor in the 70-dBA noise condition, but both groups had $\overline{F_0}$ values that were significantly lower than that measured for NH2-HI ($p<10^{-3}$).

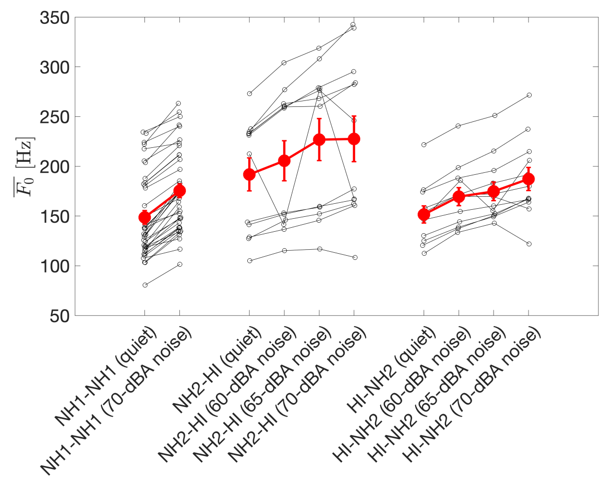

Figure 1. F_0 median of each talker speaking in different environmental conditions. Open black circles: individual talker data; filled red circles: group averages; red error bars represent standard errors.

Fig. 2 shows the $\sigma(F_0)$ for each talker-listener combination and condition, as well as group averages for each condition, with the same colors and symbols as in Fig. 1. Similarly to $\overline{F_0}$, $\sigma(F_0)$ increased in the presence of background noise, with larger increases at higher noise levels, compared to the quiet condition. The increment in $\sigma(F_0)$ measured between the quiet and the 70-dBA noise conditions was statistically significant for all groups of talkers ($p<10^{-2}$), with the largest increase measured for NH2-HI (5.3 Hz). In contrast to what was observed for the $\overline{F_0}$ at the level of the individual talker, the increment in $\sigma(F_0)$ induced by the increase in noise level was not coherent across all talkers within a group. The

$\sigma(F_0)$ of NH2-HI was significantly higher than that of NH1-NH1 ($p<10^{-2}$), while $\sigma(F_0)$ of HI-NH2 did not differ significantly from that of the other two groups of talkers.

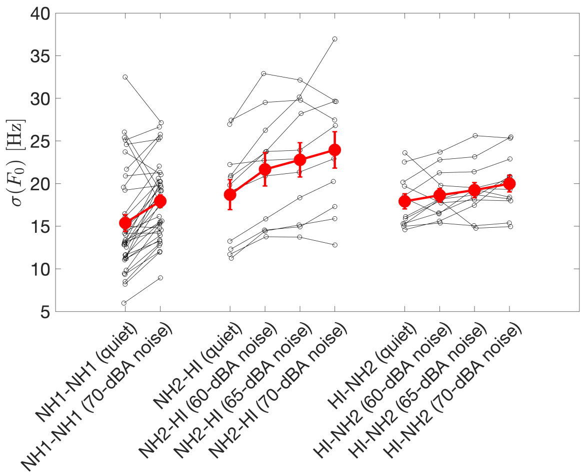

Figure 2. F_0 median absolute deviation of each talker speaking in different environmental conditions. Open black circles: individual talker data; filled red circles: group averages; red error bars represent standard errors.

4. CONCLUSIONS

The analysis of the speech recordings from naturalistic dialogues conducted in quiet and in noise between pairs of NH interlocutors and pairs of NH and HI interlocutors showed that both NH and HI listeners adapt the F_0 information of their speech in presence of communication barriers. Specifically, it was found that:

i. Both NH and HI talkers on average increased $\overline{F_0}$ and $\sigma(F_0)$ monotonically with increasing background-noise level.
ii. When speaking to a HI interlocutor, NH talkers on average increased their $\overline{F_0}$ and $\sigma(F_0)$, both in quiet and in presence of background noise.

The recordings analyzed here were obtained from previous studies and were not designed for the purpose of the present analysis. In particular, the different populations of NH interlocutors as well as the different background noise conditions employed in such previous studies, pose a limit the analysis conducted here. Further investigations are required to confirm the present findings.

5. ACKNOWLEDGMENTS

The authors would like to thank Anna Josefine Munch Sørensen for sharing the speech recordings from [4-5] analyzed in this study.

6. REFERENCES

[1] M.A. Picheny, N.I. Durlach, and L.D. Braida: “Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech,” *Journal of Speech, Language, and Hearing Research*, 29.1, pp. 96-103, 1985.

[2] M.A. Picheny, N.I. Durlach, and L.D. Braida: “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,” *Journal of Speech, Language, and Hearing Research*, 29.4, pp. 434-446, 1986.

[3] R. Baker, V.hazan: “DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs,” *Behavior research methods*, 43.3, pp. 761-770, 2011.

[4] A.J.M. Sørensen, M. Fereczkowski, and E.N. MacDonald: “Effects of Noise and Second Language on Conversational Dynamics in Task Dialogue,” *Trends in Hearing*, 25, 2021.

[5] A.J.M. Sørensen, E.N. MacDonald, and T. Lunner, “Timing of turn taking between normal-hearing and hearing-impaired interlocutors,” in *Proceedings of the International Symposium on Auditory and Audiological Research*, Vol. 7, pp.37-44, 2019.