DOI: 10.61782/fa.2023.0149

forum acusticum 2023

EFFECTS OF FUNDAMENTAL-FREQUENCY DYNAMICS ON SENTENCE INTELLIGIBILITY IN COMPETING-TALKER SCENARIOS

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

Differences in the fundamental-frequency (F_0) dynamic range between competing voices (i.e., an F_0-dynamic-range contrast) have recently been shown to facilitate the perceptual separability of the target speech from the interfering speech. This effect was observed when pairing natural voices with different levels of intonation and using a fixed combination of target and masking talkers. The present study extended these findings by using a larger variety of talkers, F_0-dynamic-range levels and F_0-dynamic-range contrasts. Target-speech intelligibility was measured in young normal-hearing listeners as a function of the F_0-dynamic-range contrast between two competing sentences (spoken by the same talker and digitally manipulated in F_0) that had either the same or different average F_0. Speech intelligibility (i) was only moderately affected by the F_0-dynamic-range contrast, independent of the difference in average F_0, (ii) was lowest when both sentences had a small F_0 dynamic range, and (iii) increased when a moderate F_0-dynamic-range was introduced in at least one of the sentences, regardless of the F_0-dynamic-range contrast between them. These findings suggest that the qualitative dissimilarity between the F_0 trajectories of competing sentences, rather than their F_0-dynamic-range contrast, aids speech intelligibility.

[*] ***Corresponding author***: poao@eriksholm.com.

Keywords: *fundamental frequency, competing talkers, speech intelligibility*

1. INTRODUCTION

In competing-talker scenarios, where a target voice is masked by one or several interfering voices, the fundamental-frequency (F_0) differences between the voices can provide useful auditory cues that aid speech intelligibility. It has been largely demonstrated that when two competing voices differ in the time-average of their F_0, target-speech understanding is facilitated [1-2]. More recently, it was shown that differences in the F_0 dynamics of the competing voices, i.e., an F_0-dynamic-range contrast, can also help target-speech intelligibility for normal-hearing listeners in such complex acoustic scenarios [3].

The only available study on this topic [3] used an experimental design with two interfering voices and female voices only (both as target and interferers). Furthermore, they obtained the desired F_0-dynamic-range contrast by pairing speech signals recorded with different speaking styles (e.g., monotonous/sad or excited style). Such method may not allow to isolate the effects of F_0 dynamic range and F_0-dynamic-range contrast on speech intelligibility, as also other acoustic features of speech such as amplitude modulations may have been modified by the talkers when changing their speaking style.

The present study aimed at extending the investigation on the effects of the F_0 dynamic range and the F_0-dynamic-range contrast on speech intelligibility in competing talker scenarios for normal-hearing listeners, by (i) using a larger number of talkers, with a wide variability of F_0 dynamic

ranges and their contrast and (ii) using a processing method of the speech stimuli that targeted only their F_0 information.

2. METHODS

Nineteen young, native Danish, NH listeners participated in the study. Speech intelligibility was measured by presenting pairs of Danish HINT sentences [4], spoken by the same talker (which varied over trials) diotically over headphones. The pairs of sentences were presented using the competing-voices test procedure developed by [5]. The F_0-dynamic-range contrast between competing sentences was calculated as

$$R = \log \frac{\sigma(F_0^{target})}{\sigma(F_0^{masker})} \quad (1)$$

where $\sigma(F_0)$ is the F_0 dynamic range of each sentence, measured as median absolute deviation of their F_0 trajectories. Pairs of sentences with R values ranging from -1.8 to 1.8, in 0.6 steps, were generated by combining $\sigma(F_0^{target})$ and $\sigma(F_0^{masker})$ values, randomly chosen within the natural range of values found in the recordings of each talker. The average F_0 difference ($\Delta\overline{F_0}$) between the two sentences was either 0 or 6 semitones. The desired $\Delta\overline{F_0}$ and R values were applied to the pair of sentences by manipulating their F_0 information with the software PRAAT. In each pair, the sentences were mixed at a target-to-masker ratio (TMR) of -4 dB. Each combination of $\Delta\overline{F_0}$ and R conditions was tested with 20 pairs of sentences.

3. RESULTS

Fig. 1 shows speech intelligibility calculated as the proportion of correctly recognized words from the target sentence, averaged across listeners and sentences as a function of R, for the two $\Delta\overline{F_0}$ conditions of 0 semitones (blue circles) and 6 semitones (red squares). Overall, no significant effect of R was observed, in either of the $\Delta\overline{F_0}$ conditions tested. When averaged across R conditions, speech intelligibility was found to be significantly higher ($p<10^{-5}$) by 10 percentage points for $\Delta\overline{F_0}$=6 semitones than for $\Delta\overline{F_0}$=0 semitones.

Fig. 2 illustrates an additional analysis of the results, which shows the interaction of the effects of R, $\sigma(F_0^{target})$ and $\sigma(F_0^{masker})$. In the figure, the stimuli are divided into nine regions, each one characterized by a specific combination of $\sigma(F_0^{target})$ and $\sigma(F_0^{masker})$.

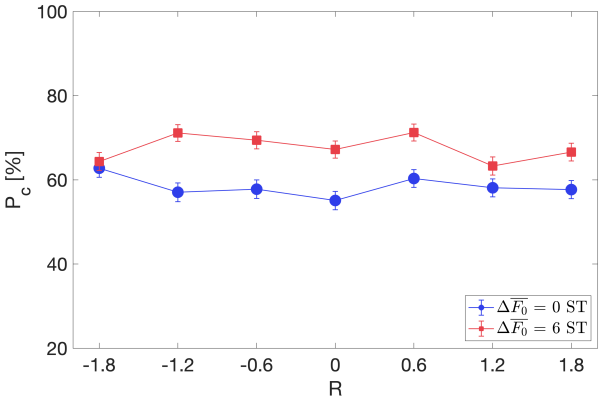

Figure 1. Speech intelligibility scores as a function of F_0-dynamic-range contrast (R), averaged across listeners and sentences. The blue circles represent the results for $\Delta\overline{F_0}$=0 semitones and the red squares for the $\Delta\overline{F_0}$=6 semitones. Error bars represent standard errors.

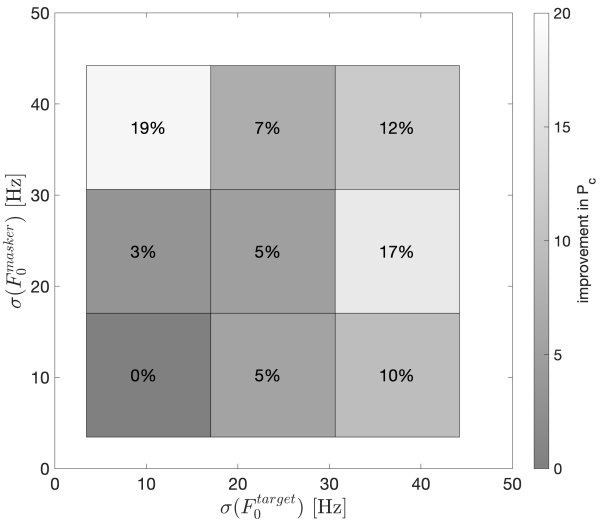

Figure 2. Analysis of average change in speech intelligibility score for groups of stimuli characterized by specific combinations of $\sigma(F_0^{target})$ and $\sigma(F_0^{masker})$, in the $\Delta\overline{F_0}$=0 semitones condition. The bottom-left corner represents the baseline.

Average speech intelligibility was calculated over the stimuli belonging to each region. In each region, the figure shows the average speech-intelligibility change calculated with respect to the region where both target and masker had relatively low F_0 fluctuations, i.e., the bottom-left corner region (baseline condition). This analysis was only conducted for the $\Delta\overline{F_0}$=0 semitones condition for illustration purposes. Speech-intelligibility improvements with respect to this region were observed in connection with an increase of either $\sigma(F_0^{target})$, $\sigma(F_0^{masker})$ or both, regardless of the *R* value they generated.

4. CONCLUSIONS

The speech-intelligibility experiment conducted in this study showed that:

i. The F_0-dynamic-range contrast between two competing voices has a negligible effect on target-speech intelligibility.
ii. The average F_0 difference between two competing sentences was beneficial for speech intelligibility, as already largely shown in previous studies.
iii. The presence of relatively large F_0 dynamic range in at least one of the sentences provided a substantial benefit to speech intelligibility, regardless of the F_0-dynamic-range contrast between sentences and even in absence of an average F_0 difference.

5. REFERENCES

[1] C.J. Darwin, D.S. Brungart, and B.D., Simpson, "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," in *The Journal of the Acoustical Society of America,* 114.5, pp. 2913–2912, 2003.

[2] P.A. Mesiano, J. Zaar, L. Bramsløw, H. Relaño-Iborra, T. Dau, "The role of average fundamental frequency difference on the intelligibility of real-life competing sentences," in *Journal of Speech, Language, and Hearing Research*, in press.

[3] L. Calandruccio, P.A. Wasiuk, E. Buss, L.J. Leibold, J. Cong, A. Holmes, and J. Oleson: "The effect of target/masker fundamental frequency contour similarity on masked-speech recognition," *The Journal of the Acoustical Society of America*, 146.2, pp. 1065–1076, 2019.

[4] J.B. Nielsen and T. Dau, "The Danish hearing in noise test," *International Journal of Audiology*, 50(3), pp. 202–208, 2011.

[5] L. Bramsløw, M. Vatti, R. Rossing, G. Naithani, and N.H. Pontoppidan, "A Competing Voices Test for Hearing-Impaired Listeners Applied to Spatial Separation and Ideal Time-Frequency Masks," *Trends in Hearing*, 23, pp. 1–12, 2019.