Effect of *F*0 contours on top-down repair of interrupted speech

Jeanne Clarke,[a] Deniz Kazanoğlu,[b] Deniz Başkent,[c] and Etienne Gaudrain[d]

Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, P.O. Box 30.001, BB21, 9700 RB Groningen, The Netherlands

j.n.clarke@umcg.nl, dkazanoglu@anadolu.edu.tr, d.baskent@umcg.nl, etienne.gaudrain@cnrs.fr

Abstract: Top-down repair of interrupted speech can be influenced by bottom-up acoustic cues such as voice pitch (*F*0). This study aims to investigate the role of the dynamic information of pitch, i.e., *F*0 contours, in top-down repair of speech. Intelligibility of sentences interrupted with silence or noise was measured in five *F*0 contour conditions (inverted, flat, original, exaggerated with a factor of 1.5 and 1.75). The main hypothesis was that manipulating *F*0 contours would impair linking successive segments of interrupted speech and thus negatively affect top-down repair. Intelligibility of interrupted speech was impaired only by misleading dynamic information (inverted *F*0 contours). The top-down repair of interrupted speech was not affected by any *F*0 contours manipulation.

[DDO]

Date Received: November 14, 2016 **Date Accepted:** June 14, 2017

1. Introduction

The brain is able to reconstruct speech partially inaudible, for example, due to background noise. Top-down repair of speech is affected by both linguistic knowledge, expectations, and context (e.g., Bashford *et al.*, 1992), as well as the availability of acoustic bottom-up cues (Bhargava *et al.*, 2014; Clarke *et al.*, 2016). Voice pitch, the perceptual correlate of the fundamental frequency (*F*0), has been identified as an important bottom-up cue that helps speech in noise perception. It is a strong across-frequency grouping cue, as (i) pitch information brings coherence to speech sounds by fusing together different parts of the spectrum (e.g., formants), which helps phoneme identification, and (ii) average *F*0 and/or dynamic patterns (*F*0 contours) help to segregate different sound sources, which is useful to attend to and better understand the target speech in the presence of maskers. The present study investigates whether voice pitch is also used for sequentially linking successive speech segments across interruptions.

Specifically, the goal of this study was to investigate the effect of the magnitude and the direction of *F*0 contours on intelligibility of interrupted speech (with silence and noise) and on phonemic restoration, quantifying top-down repair of speech. The modifications of the original *F*0 contours (original *F*0) consisted of (i) inverting the *F*0 contours within the same magnitude to misrepresent the direction of the *F*0 dynamic information (inverted *F*0), (ii) compressing the magnitude of *F*0 around its median value, thus removing the dynamic information of the *F*0 contours (flat *F*0), (iii) expanding the magnitude of *F*0 by exaggerating the *F*0 contours with a factor 1.5 (exaggerated 1.5), and (iv) with a factor 1.75 (exaggerated 1.75). These *F*0 contour manipulations were specifically chosen as they consistently showed an effect on speech perception in background noise (Meister *et al.*, 2011; Miller *et al.*, 2010) or with single-talker interferer (Binns and Culling, 2007).

1.1 Effects of F0 contour manipulations on speech perception and top-down restoration

First, the *F*0 contour manipulations can affect intelligibility of interrupted speech at the speech segment level. Phoneme and coarticulation identification depend on

[a]Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands.

[b]Present address: Anadolu University, Faculty of Health Science, Department of Language and Speech Therapy, Eskişehir, Turkey.

[c]Author to whom correspondance should be addressed. Also at: University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands.

[d]Also at: Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS UMR 5292, INSERM U1028, University Lyon 1, Lyon, France.

previous language experience, from learned covariation patterns of *F*0 and formants in speech (Peterson and Barney, 1952). Thus, *F*0 manipulations not coordinated with unmanipulated formants, such as changing the direction of *F*0 contours (inverted *F*0 contour) and reducing *F*0 contour's magnitude (flat *F*0 contour) might produce more identification errors, thus impairing the intelligibility of the speech segments. However, with exaggerated *F*0 contours, phonemes are more contrasted, and provided they still correspond to proper categories, perception of the independent speech segments could be facilitated (such as in infant-directed speech).

Second, *F*0 seems to contribute to a robust linking of successive segments of interrupted speech. However, it has been shown that the average *F*0 of a voice does not partake to the phonemic restoration effect (Clarke *et al.*, 2014). Therefore, here we hypothesize that it is, instead, the dynamic fluctuations of *F*0 (the contour over time) that are used to link successive speech segments, and thus to restore missing ones. The predictable nature of *F*0 contours supports that the *F*0 contour direction would guide the listener to successfully link successive segments of interrupted speech, even more so when the interruptions are filled with noise (Dannenbring, 1976). Inverted contours might impair this linking because of the misleading dynamic information it provides, possibly resulting in a reduced restoration effect, whereas flat contours, without dynamic *F*0 information, would not help nor impair linking successive speech segments. We were not expecting an exaggerated *F*0 contour to impair linking successive speech segments (in line with *F*0 alternation of one octave between successive speech segments not hindering interrupted speech intelligibility and phonemic restoration, Clarke *et al.*, 2014).

In addition, *F*0 contours, as a primary feature contributing to prosody, also give information on the intonation of an utterance at the sentential level. Some linguistic functions that can be associated with *F*0 contours are segmentation (word boundaries) and lexical stress (used for segmentation in English and in Dutch), accentuation (focus on important words in a sentence), and types of utterance (statement or question), as well as lexical meaning in tonal languages (e.g., Wang *et al.*, 2013). In our study, we expected the inverted *F*0 contour to have a detrimental effect on intelligibility at the sentential level because lexical stress and accentuation would be misleading (attenuation of important information instead of being highlighted). Moreover, we expected no effect of a flat *F*0 contour on intelligibility at the sentential level, as the lack of dynamic information could be compensated by linguistic context (in line with Chatterjee *et al.*, 2010). Furthermore, it is possible that an exaggerated *F*0 contour might strengthen lexical stress and accentuation, with important information even more highlighted (proportionally with the expansion ratio). We can thus expect a positive effect of exaggerated *F*0 contour on intelligibility at the sentential level.

Overall, taking into account the expectations at the different levels, i.e., individual speech segments intelligibility, linking of successive speech segments, and intelligibility of the whole sentence, we expected to have better to worse performance for the exaggerated contour, the flat contour, and finally for the inverted contours.

2. Methods

Sixteen native Dutch speakers with normal hearing (20 dB hearing level or less pure-tone thresholds at audiometric frequencies of 250–6000 Hz in both ears), aged between 20 and 40 yrs (mean = 25.5, standard deviation = 6.2), and with no hearing or speech-related problems (self-reported), participated in this study. The study was approved by the Medisch Etische Toetsingscommissie (Medical Ethical Review Committee) of the University Medical Center Groningen. All participants were informed about the procedure and signed a consent form. Participants were paid for their participation.

The speech stimuli of this study were Dutch everyday sentences with "high sentential-context" (i.e., common vocabulary and semantic context), spoken by a male talker (for more details see Versfeld *et al.*, 2000). The fundamental frequency (*F*0) contours of the voiced segments of the sentences were manipulated offline using TANDEM-STRAIGHT in MATLAB (Kawahara and Morise, 2011). The five *F*0 contour conditions, namely inverted *F*0, flat *F*0, original *F*0, exaggerated *F*0 by a factor of 1.5, and exaggerated *F*0 by a factor of 1.75 (see Fig. 1(A), were implemented using the same formula described in Binns and Culling (2007), Miller *et al.* (2010), and Meister *et al.* (2011). For the inverted *F*0 condition, the symmetry about the median *F*0 value was used in the logarithmic scale (modification ratio $\alpha = -1$). For the flat *F*0 condition, the median *F*0 replaced the original *F*0 ($\alpha = 0$). For the exaggerated *F*0 by a factor 1.5 and 1.75, the *F*0 contours were expanded in the logarithmic scale by a ratio 1.5 and 1.75. The resynthesized sentences were modulated with a square wave to obtain an

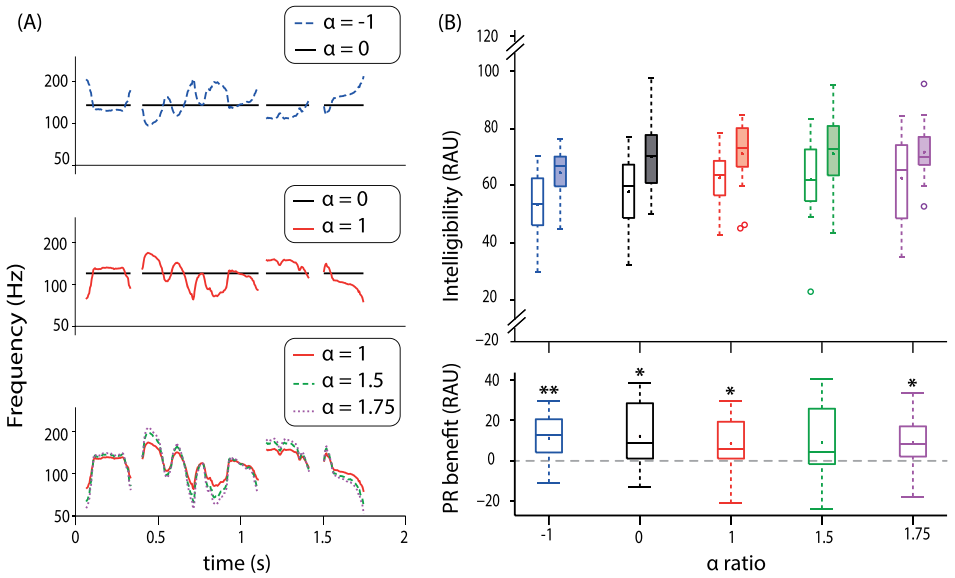

Fig. 1. (Color online) (A) *F*0 contour manipulations on an example sentence "Buiten is het donker en koud" (Outside it is dark and cold). (B) Intelligibility results (top panel) for silent interruptions (empty boxes) and noise interruptions (filled boxes) and phonemic restoration effect (lower panel) for the manipulated *F*0 contours, from left to right: inverted ($\alpha = -1$), flat ($\alpha = 0$), original ($\alpha = 1$), exaggerated with a factor 1.5 ($\alpha = 1.5$), and 1.75 ($\alpha = 1.75$). The horizontal line indicates the median, the box indicates the 25th and 75th quartiles, and the dashed whiskers indicate the 1.5 interquartile range. The circles indicate the outliers. The dots indicate the mean.

interrupted version, with an interruption rate of 2.2 Hz, and a duty cycle of 50%, resulting in 227-ms speech segments followed by 227-ms interruptions. Silent and speech shaped noise (signal-to-noise ratio of −5 dB) were used for interruption (as done in Clarke *et al.*, 2014).

The participants were seated in a sound-attenuated booth during the experiment. The apparatus was the same as used in Clarke and colleagues (2014, 2016). The responses from the participants were recorded for offline scoring of sentence intelligibility by native Dutch student assistants.

Participants came for a single session which included written informed consents, the audiometric test, the baseline measurement, familiarization and data collection, the debriefing, and occasional breaks. To measure the baseline, the first five lists of the stimulus corpus, 13 sentences each, were used without interruption. Each list was presented with a different *F*0 contour condition, and the order of all lists and conditions were randomized for each participant. Participants then familiarized with the procedure listening to five conditions selected randomly out of the ten experimental conditions. The familiarization phase was similar to the experiment expect that written and auditory feedback were provided.

The data collection consisted of 10 conditions: 5 *F*0 contour conditions (inverted, flat, original, and exaggerated with 1.5 and 1.75 ratio) × 2 interruption conditions (silent and noise). Both the sentence lists and the conditions were presented in random order. At the beginning of each set the participants heard the same introduction sentence processed with the experimental condition to come, to prepare them for the trial condition. A tone preceded each sentence to alert the participant. After hearing the sentence, the participants were asked to repeat what they could understand from the sentence stimuli, and were additionally encouraged to guess as much as possible. Each word in the sentence was scored according to the participants' correct response. Total rationalized arcsine transformed unit (RAU) scores were computed for each condition.

3. Results

The upper panel of Fig. 1(B) displays the intelligibility scores in each of the ten conditions (5 *F*0 contours × interruptions). A repeated-measure two-way analysis of variance (ANOVA) was performed on the RAU scores with *F*0 contours (5 levels) and interruption (2 levels) as the within-subject factors. The effect size is indicated by eta square, η^2. A significant effect of *F*0 contours [$F(4,60) = 4.20$, $p = 0.0046$, $\eta^2 = 0.063$] indicated that *F*0 contours had an overall effect on intelligibility of interrupted speech. A significant effect of interruption [$F(1,15) = 71.63$, $p < 0.001$, $\eta^2 = 0.15$] indicated the

presence of phonemic restoration effect. However, there was no interaction between the two factors [$F(4,60) = 0.16$, $p = 0.95$, $\eta^2 = 0.0036$] which indicated that the *F*0 contour manipulations had the same effect on intelligibility with silent and noise interruptions, suggesting that the manipulations of the *F*0 contours did not affect the difference between the two interruptions (noise and silent) that is a measure of phonemic restoration. The phonemic restoration effect was computed by subtracting the scores in the silent condition from those in the noise condition for each *F*0 contour condition. The results from the phonemic restoration for the manipulated *F*0 contours are displayed in the lower panel of Fig. 1.

The overall effect of *F*0 contours on intelligibility was small, as shown by the small effect size ($\eta^2 = 0.063$) that indicated that only 6% of the variance of intelligibility was explained by the *F*0 contour manipulations. Differences were observed between only some pairs of conditions. A *post hoc* analysis using pairwise *t*-tests with False Discovery Rate (FDR) control was conducted to compare performance of *F*0 contours between each other (averaged on noise and silent conditions, because of the lack of interaction between the two factors). The results showed that intelligibility performance with inverted *F*0 contour was significantly poorer from that of other contours [$t(15) = 4.023$, $p_{FDR} = 0.0034$; $t(15) = 2.84$, $p_{FDR} = 0.026$; $t(15) = 3.60$, $p_{FDR} = 0.0054$, for original, exaggerated *F*0 contour with ratio 1.5 and 1.75, respectively] except that of the flat *F*0 contour [$t(15) = 2.0075$, $p = 0.13$]. All other comparisons were not significantly different. Moreover, the baseline scores (of uninterrupted sentences) were at ceiling and were not affected by the *F*0 contour manipulations [$F(4,18) = 0.481$, $p = 0.749$, $\eta^2 = 0.013$], suggesting that our *F*0 contour manipulations did not impair intelligibility of uninterrupted speech.

There was no effect of *F*0 contours on the phonemic restoration benefit as indicated by the lack of interaction in the ANOVA on the intelligibility scores. However, as an *a priori* variable, we tested whether the phonemic restoration scores were significantly different from 0 with a one-sample *t*-test for each *F*0 contour condition (indicated by a black star on the lower panel of Fig. 1). A significant phonemic restoration benefit was observed in all *F*0 contour conditions [$t(15) = 3.87$, $p = 0.0015$ for inverted; $t(15) = 2.90$, $p = 0.011$, for flat; $t(15) = 2.53$, $p = 0.023$, for original; $t(15) = 2.88$, $p = 0.012$, for exaggerated *F*0 contour with ratio 1.75], except one, the exaggerated *F*0 contour with ratio 1.5 [$t(15) = 2.022$, $p = 0.061$]. This indicates that *F*0 contour manipulations did not affect top-down repair mechanisms of interrupted speech.

4. Discussion

We were interested in investigating the effects of the magnitude and direction of *F*0 contours on intelligibility and top-down repair of interrupted speech. We showed that modifying the magnitude of *F*0 contours (all conditions except the inverted contours) did not have any effect on interrupted speech intelligibility, contrary to other studies with continuous background interferer (Binns and Culling, 2007; Miller *et al.*, 2010; Wang *et al.*, 2013). However, as already pointed out for interrupted speech with silence, Chatterjee *et al.* (2010) did not show reduced intelligibility with flat *F*0 contour at a faster interruption rate, suggesting again the prevalence of linguistic cues. The results from the two exaggerated *F*0 contours did not confirm our hypothesis on a better phonetic categorization improving speech perception with wider *F*0 variations, but confirmed our hypothesis that wider *F*0 variations do not weaken linking successive speech segments (in line with Clarke *et al.*, 2014). On the other hand, the direction of the *F*0 contours seems to be a cue listeners relied on for intelligibility of interrupted speech. Indeed, partially validating our hypothesis, having misleading dynamic information of *F*0 (inverted contours) impaired interrupted speech intelligibility. This confirms that wrong *F*0 dynamic information can lead to lower intelligibility of interrupted speech, as was found for speech with continuous background interferer (Binns and Culling, 2007; Meister *et al.*, 2011; Miller *et al.*, 2010). An explanation can be that original *F*0 contour helps to define clause boundaries whereas inverted *F*0 contour distorts those boundaries (Wingfield *et al.*, 1984), impairing sentence intelligibility. Taken all together, these results suggest that the *F*0 variations' magnitude may not be used as a linking cue for interrupted speech perception, but that direction of *F*0 contours may. Moreover, only when all three aspects of interrupted speech perception (individual speech segments intelligibility, linking successive speech segments, and sentential intelligibility) were affected by the *F*0 contour manipulation, i.e., for the inverted *F*0 contour, did overall intelligibility decrease. This suggests that participants seem to fail

to compensate for the inverted *F*0 contour manipulation that impairs more aspects of interrupted speech perception than our other *F*0 contour manipulations.

Nevertheless, even the misleading dynamic information of *F*0 did not impair the top-down repair of interrupted speech. This result suggests that participants may have compensated for the atypical *F*0 cues with the linguistic context (as suggested by Chatterjee *et al.*, 2010). In this study, it seems that top-down repair of speech may rely more on linguistic cues than on *F*0 cues (Clarke *et al.*, 2014). For *F*0 contours' magnitude (flat and exaggerated *F*0 contour conditions), the restoration benefit remained, suggesting that sequentially linking successive speech segments was not affected by *F*0 contours' magnitude, likely because of the prevalence of the linguistic contextual cues in the present task. However, *F*0 variations may be necessary for other tasks, such as recognizing emotions, and might thus be more difficult to compensate for in such a case. We can thus speculate that with "low-context" sentences (no semantic context but syntactically correct, for example), the task becomes harder, and *F*0 contour manipulations might further impair intelligibility and top-down-repair of interrupted speech. Furthermore, when *F*0 contours were exaggerated, the noise bursts filling the silent interruptions still acted as a masker with wider *F*0 variations across the noise bursts (in line with Clarke *et al.*, 2014). The present study suggests that the *F*0 contour manipulations, presumably weakening the sequential linking of successive speech segments (for inverted *F*0 contour), did not affect phonemic restoration benefit. Thus, for interrupted speech, it seems that speech segments with *F*0 contour manipulations can still clearly be discriminated from the filler noise, a mechanism involved in top-down repair of speech. In contrast, in the case of a speech masker, discriminating competing talkers (target and maskers) relies on *F*0 contours. Presumably, discriminating two sources of same nature (speech-on-speech) is more affected by *F*0 contour manipulations than discriminating two sources different in nature (speech and noise) as suggested by the difference of results observed between the present study and previous studies (Binns and Culling, 2007; Meister *et al.*, 2011; Miller *et al.*, 2010).

Even if interrupted speech intelligibility did not significantly differ between *F*0 contour manipulations (except for the inverted contours), it is still possible that our participants did require more effort to perform the task with the atypical *F*0 contours. Moreover, speech redundancy is present at different layers of speech processing, and other cues, instead of *F*0 contours, might be used for prosody processing, such as duration and intensity. Depending on the task difficulty (affected by the amount of information in the speech stimuli), listeners may rely differently on prosodic information. For example, the redundancy added to speech from normal prosody may be relevant when the task becomes harder by reducing the processing time (i.e., increasing the cognitive load), especially using duration cues in prosody (a deficit of flat *F*0 over normal *F*0 being observed at normal speech rate and not for time-compressed speech in Wingfield *et al.*, 1984). This is in line with the fact that other cues, instead of *F*0 contours, might be used for prosody processing. Indeed, intensity and duration are good indicators of prosodic information as they co-vary with *F*0 contours and provide redundant information for prosody processing. As a result, even when *F*0 contours are manipulated, intensity and duration can be used for stress perception, segmentation, and intonation recognition (e.g., Peng *et al.*, 2012). In the present study, the inverted *F*0 contour, which provided misleading and distorted cues, did not complement speech redundancy, which might explain the lower performance in speech intelligibility for this condition.

To summarize, the present study shows a relatively small effect of *F*0 contour manipulations on intelligibility of interrupted speech and no effect on phonemic restoration. Confirming the findings of Clarke *et al.* (2014), these results indicate that top-down repair of speech could be robust to atypical voice cues, suggesting that listeners may partly compensate for the degraded voice cues. It is possible that linguistic information, such as the sentential context, which plays an important role in the restoration mechanisms, helped overcome the negative effects of manipulated *F*0 contours. Another possibility is that participants relied on other co-varying prosodic cues, such as intensity and duration. Presumably, a combination of both mechanisms may occur to achieve the best possible performance.

Acknowledgments

The authors would like to thank Floor Burgerhof and Wilke Bosma for transcribing participant responses, as well as the participants. This study was supported by a VIDI grant from the Netherlands Organization for Scientific Research, NWO, and Netherlands Organization for Health Research and Development, ZonMw (Grant No. 016.096.397).

Further support came from a Rosalind Franklin Fellowship from the University of Groningen, University Medical Center Groningen, and funds from Heinsius Houbolt Foundation. The study is part of the "Healthy Aging and Communication" research program of the Otorhinolaryngology Department of University Medical Center Groningen and was also conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique," ANR-10-LABX-0060/ANR-11-IDEX-0007) operated by the French National Research Agency.

References and links

Bashford, J. A., Riener, K. R., and Warren, R. M. (**1992**). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. **51**, 211–217.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top-down restoration of speech in cochlear-implant users," Hear. Res. **309**, 113–123.

Binns, C., and Culling, J. F. (**2007**). "The role of fundamental frequency contours in the perception of speech against interfering speech," J. Acoust. Soc. Am. **122**, 1765–1776.

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). "Recognition of interrupted sentences under conditions of spectral degradation," J. Acoust. Soc. Am. **127**, EL37–EL41.

Clarke, J., Gaudrain, E., Chatterjee, M., and Başkent, D. (**2014**). "T'ain't the way you say it, it's what you say—Perceptual continuity of voice and top-down restoration of speech," Hear. Res. **315**, 80–87.

Clarke, J. N., Başkent, D., and Gaudrain, E. (**2016**). "Pitch and spectral resolution: A systematic comparison of bottom-up cues for top-down repair of degraded speech," J. Acoust. Soc. Am. **139**, 395–405.

Dannenbring, G. L. (**1976**). "Perceived auditory continuity with alternately rising and falling frequency transitions," Can. J. Psychol. **30**, 99–114.

Kawahara, H., and Morise, M. (**2011**). "Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework," Sadhana **36**, 713–722.

Meister, H., Landwehr, M., Pyschny, V., Grugel, L., and Walger, M. (**2011**). "Use of intonation contours for speech recognition in noise by cochlear implant recipients," J. Acoust. Soc. Am. **129**, EL204–EL209.

Miller, S. E., Schlauch, R. S., and Watson, P. J. (**2010**). "The effects of fundamental frequency contour manipulations on speech intelligibility in background noise," J. Acoust. Soc. Am. **128**, 435–443.

Peng, S.-C., Chatterjee, M., and Lu, N. (**2012**). "Acoustic cue integration in speech intonation recognition with cochlear implants," Trends Amplif. **16**, 67–82.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Wang, J., Shu, H., Zhang, L., Liu, Z., and Zhang, Y. (**2013**). "The roles of fundamental frequency contours and sentence context in Mandarin Chinese speech intelligibility," J. Acoust. Soc. Am. **134**, EL91–EL97.

Wingfield, A., Lombardi, L., and Sokol, S. (**1984**). "Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation," J. Speech Lang. Hear. Res. **27**, 128–134.