Intelligibility of chimeric locally time-reversed speech: Relative contribution of four frequency bands

Ueda, Kazuo
Department of Human Science, Faculty of Design, Kyusyu University

Matsuo, Ikuo
Department of Information Science, Tohoku Gakuin University

出版情報：JASA Express Letters. 1 (6), pp.065201-1-065201-6, 2021-06-23. Acoustical Society of America
バージョン：

JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Intelligibility of chimeric locally time-reversed speech: Relative contribution of four frequency bands

Kazuo Ueda[1,a] and Ikuo Matsuo[2]
[1]Department of Human Science, Faculty of Design/Research Center for Applied Perceptual Science/Research and Development Center for Five-Sense Devices, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
[2]Department of Information Science, Tohoku Gakuin University, 2-1-1 Tenjinzawa, Izumi-ku, Sendai 981-3193, Japan
ueda@design.kyushu-u.ac.jp, matsuo@mail.tohoku-gakuin.ac.jp

Abstract: Intelligibility of four-band speech stimuli was investigated ($n = 18$), such that only one of the frequency bands was preserved, whereas other bands were locally time-reversed (segment duration: 75–300 ms), or vice versa. Intelligibility was best retained (82% at 75 ms) when the second lowest band (540–1700 Hz) was preserved. When the same band was degraded, the largest drop (10% at 300 ms) occurred. The lowest and second highest bands contributed similarly less strongly to intelligibility. The highest frequency band contributed least. A close connection between the second lowest frequency band and *sonority* was suggested.

[Editor: Douglas D O'Shaughnessy] https://doi.org/10.1121/10.0005439

Received: 29 March 2021 **Accepted:** 4 June 2021 **Published Online:** 23 June 2021

1. Introduction

Sonority can be defined as "a unique type of relative, *n*-ary (non-binary) feature-like phonological element that potentially categorizes all speech sounds into a hierarchical scale," according to Parker (2012a). Since the concept of sonority was proposed by several authors [for a review, see Clements (1988)], e.g., as *aperture* by de Saussure (1916), the concept has been studied for many years, and the framework of sonority has been supported as well. One of the most influential theories in linguistics today, *optimality theory* (Prince and Smolensky, 2004), is based fully on the *sonority sequencing principle* (Clements, 1988), which states that a phoneme at the position of a syllable nucleus has the peak in sonority; in contrast, the phonemes before and after the nucleus have increasingly lower sonority as they are further from the nucleus. Nevertheless, controversy regarding the concept of sonority continues (e.g., Parker, 2012b; Rahilly, 2016). For example, the consonant /s/ shows conspicuous behavior that does not fit into sonority scales in Indo-European languages as well as in other languages (Goad, 2016), the sonority scales proposed by phoneticians are slightly different from each other (e.g., de Saussure, 1916; Harris, 1994; Selkirk, 1984; Spencer, 1996), and the arguments about the concept of sonority tend to be circular (Rahilly, 2016).

Nakajima and colleagues (Nakajima *et al.*, 2017; Ueda and Nakajima, 2017; Zhang *et al.*, 2020) took an approach to the issue entirely different from the previous approach in linguistics. Based on factor scores extracted from correlation coefficients between power fluctuations in critical-band filtered speech (Ueda and Nakajima, 2017), they examined relationships between the factor scores and phonemic labels assigned to speech sentences. Nakajima *et al.* (2017) found high correlation coefficients (0.82–0.87) between factor scores in one of the factors and the order of phonemic categories in the previously proposed sonority scales (de Saussure, 1916; Harris, 1994; Selkirk, 1984; Spencer, 1996). Specifically, the factor that was closely related to the second lowest frequency band (540–1700 Hz), i.e., the *mid-low factor* (Nakajima *et al.*, 2017), produced factor scores that were correlated best with the sonority scales. If this is indeed the case, this frequency band (540–1700 Hz) should affect the intelligibility of speech more profoundly than other frequency bands. The present investigation aims to further clarify the relationship between the four frequency bands extracted from speech (Ueda and Nakajima, 2017) and speech intelligibility.

The four frequency bands (50–540, 540–1700, 1700–3300, and 3300–7000 Hz), which provided the basis for the arguments by Nakajima *et al.* (2017), were determined using factor analyses on 58–200 spoken sentences in eight languages/dialects by 10–20 talkers (Ueda and Nakajima, 2017). The spoken sentences were bandpass filtered with a critical-band filter bank. Each filter output was squared to obtain power fluctuations. Correlation coefficients were calculated between every possible combination of the power fluctuations. The correlation coefficient matrix was submitted to a principal component analysis, and extracted components were varimax rotated to obtain factors. The cutoff frequencies for

[a]Author to whom correspondence should be addressed, ORCID: 0000-0002-1885-0463.

the four frequency bands came from the crossover frequencies of the factor loading curves. Therefore, the frequency bands were obtained without referring to any linguistic processing (e.g., identifying phonemes or segmenting syllables). Rather, they originated from one of the most basic and common functions of the auditory system, namely, critical-band filtering. Putting the matter in a broader context, the frequency bands relate to other specific studies as well. One example is the investigation on how informational masking (e.g., Durlach *et al.*, 2003; Shinn-Cunningham, 2008) on each formant of the first three formants affects intelligibility (Roberts and Summers, 2018), suggesting that the second formant is most important for the intelligibility of three-formant analogs of speech. Because the range of frequency variation for the second formant largely overlaps with the range of the second lowest frequency band, i.e., 540–1700 Hz, the study by Roberts and Summers also predicts that the second lowest frequency band should be the most influential frequency band for intelligibility.

Matsuo *et al.* (2020) provided evidence suggesting that the second lowest frequency band may be the key maintaining high speech intelligibility. They used chimeric locally time-reversed speech, in which original speech samples were divided into two frequency bands, i.e., an upper band and a lower band. One of the bands was locally time-reversed, but the other band was preserved as in the original. By *chimeric*, they meant a composite stimulus comprising both the locally time-reversed part and the original part from the same speech sample (we will adopt the same meaning for this term in the present paper). *Locally time-reversed speech* is a kind of degraded speech in which original speech samples are periodically segmented, individually reversed in time, and then concatenated in the original order (e.g., Greenberg and Arai, 2004; Ishida *et al.*, 2016; Saberi and Perrott, 1999; Steffen and Werani, 1994; Stilp *et al.*, 2010; Teng *et al.*, 2019; Ueda *et al.*, 2017). Typically, intelligibility of locally time-reversed speech is almost perfect at about 40-ms segment duration under normalized speech rates. Intelligibility goes down to 50% at about 65-ms segment duration and then becomes less than 10% at about 100-ms segment duration, irrespective of language (Ueda *et al.*, 2017). In the study of Matsuo *et al.* (2020), a low-pass filter and a high-pass filter of the same cutoff frequency were employed. The cutoff frequencies were 570, 840, 1170, 1600, 2150, and 2900 Hz [five steps of two critical-bandwidth intervals (Ueda and Nakajima, 2017; Zwicker and Terhardt, 1980)]. The results of the experiment, in which intelligibility of the chimeric stimuli was examined, showed that intelligibility started to decline when the degradation included the frequency range of 840–1600 Hz.

However, the relationship between that particular frequency band and intelligibility was inferred indirectly, because their chimeric speech stimuli always included other frequency bands being processed, except for the conditions in which bands at both ends of the frequency axis were degraded. Therefore, we planned to perform an experiment to directly address the relationship between each frequency band and intelligibility. Specifically, we adopted the four frequency bands that Ueda and Nakajima (2017) determined, targeting each band for preservation or degradation one by one (Fig. 1). We examined intelligibility of the stimuli, in addition to that of control stimuli in which all bands were preserved as in the original or degraded. The frequency range highlighted by Matsuo *et al.* (2020), i.e., 840–1600 Hz, was included in the second lowest frequency band (540–1700 Hz) determined by Ueda and Nakajima (2017). Thus, the purpose of this investigation was to examine whether any particular band was more influential than others on intelligibility when each individual band was preserved or degraded.

2. Method

2.1 Listeners

A total of 18 unpaid listeners (ages 19–23) participated in the experiment. They were all Japanese native listeners with normal hearing. Their hearing levels were tested with an audiometer (RION AA-56, RION, Kokubunji, Japan) within the frequency range of 250–8000 Hz. The research was conducted with prior approval of the Ethics Committee of Kyushu University (approval ID: 70).

2.2 Stimuli and conditions

A total of 150 Japanese sentences spoken by a female talker were extracted from the Multilingual Speech Database 2002 (NTT Advanced Technology Corp., Kawasaki, Japan; 16 000-Hz sampling, 16-bit linear quantization). The speech rate of this particular talker (95%) was very close to the average calculated with ten talkers (five females and five males) included in the speech database. It has been shown that the primary determinant of intelligibility for locally time-reversed speech, except for segment duration, is speech rate (Stilp *et al.*, 2010; Ueda *et al.*, 2017). Ueda *et al.* (2017) employed both a female and a male talker in four languages and found that the intelligibility curves along segment duration for the two talkers in each language were almost identical or only marginally different. Therefore, we employed only one female talker for the stimuli in this experiment. The sentences in the database were based on articles published in newspapers and magazines. The average number of morae (a mora is a syllable-like unit in Japanese) per sentence was 18 [standard deviation (SD) = 2.9]. Extracted spoken sentences were edited to eliminate unnecessary blanks and noises. Edited speech samples were converted into 44 100-Hz sampling, with 16-bit linear quantization with Praat (Boersma and Weenink, 2020) before further processing.

Three variables were manipulated: segment duration (75, 150, and 300 ms), stimulus types [ORG-*n*: a target band (*n*) was preserved as in the original except for filtering and segmenting, and the other bands were locally

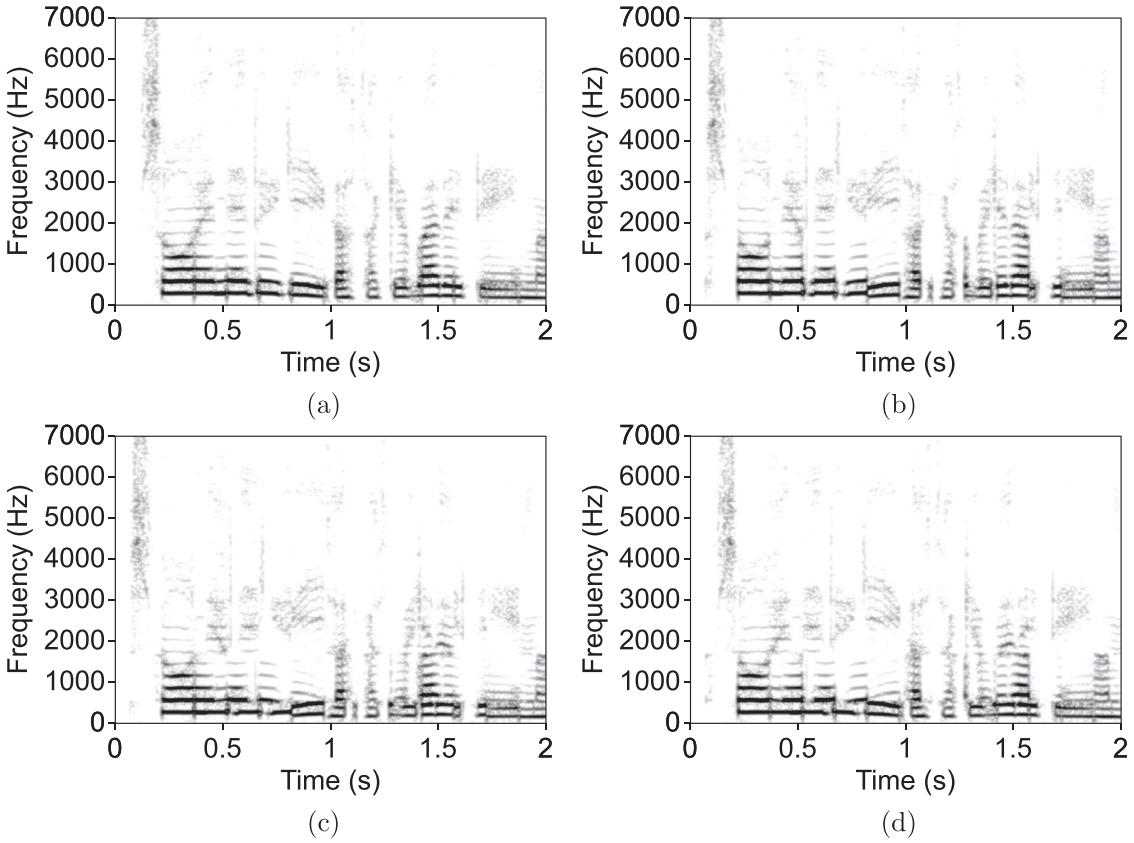

Fig. 1. Examples of stimuli in Japanese represented in narrowband spectrograms, made from a speech sample spoken by a female talker. Segment duration was 150 ms. (a) Original speech (reconstructed after filtering and segmentation), (b) locally time-reversed speech, (c) chimeric locally time-reversed speech with band 2 (540–1700 Hz) being preserved as in the original (ORG-2), and (d) chimeric locally time-reversed speech with band 2 being locally time-reversed (LTR-2). The original speech sample was extracted from the Multilingual Speech Database 2002 (NTT Advanced Technology Corp., Kawasaki, Japan).

time-reversed; LTR-*n*: a target band (*n*) was locally time-reversed, and the other bands were preserved as in the original], and a target band (none, 1, 2, 3, or 4). Target bands were numbered from 1 (the lowest) to 4 (the highest). The target band "none" refers to the conditions in which no band was preserved or degraded. Thus, "ORG-none" represents the condition in which all frequency bands were locally time-reversed, and "LTR-none" represents the condition in which all frequency bands were preserved as in the original.

The speech samples were passed through a bank of bandpass filters, dividing the frequency range from 50 to 7000 Hz into four frequency bands: 50–540, 540–1700, 1700–3300, and 3300–7000 Hz. Filtered speech samples were segmented with the three steps of segment duration, including 5-ms cosine ramps. Depending on the conditions, each segment was locally time-reversed or preserved as it was. The segments were then concatenated in the original order, and all frequency bands were summed up across frequency. The signal processing was performed with a custom software written in the J language (J Software, 2020).

2.3 Procedures

The stimuli were presented to participants diotically through headphones (DT 990 PRO, Beyerdynamic GmbH, Heilbronn, Germany) in a sound-attenuated booth (Music cabin SC3, Takahashi Kensetsu, Kawasaki, Japan). Custom software written with the LiveCode package (LiveCode Community, 2018) was used to present the stimuli. The headphones were driven with an optical interface (USB interface, Roland UA-4FX, Roland Corp., Shizuoka, Japan) and a headphone amplifier with a built-in digital-to-analog (D/A) converter (AT-DHA 3000, Audiotechnica, Machida, Japan). The sound pressure level of speech was adjusted to 72 dB (A), using a 1000-Hz calibration tone provided with the speech database. The sound pressure level was measured with an artificial ear (Brüel & Kjær type 4153, Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark), a condenser microphone (Brüel & Kjær type 4192), and a sound level meter (Brüel & Kjær type 2250).

Participants were instructed to write down exactly what they heard with hiragana or katakana (sets of symbols that are used to represent Japanese morae) without guessing. Each mora was examined for whether it was correct or

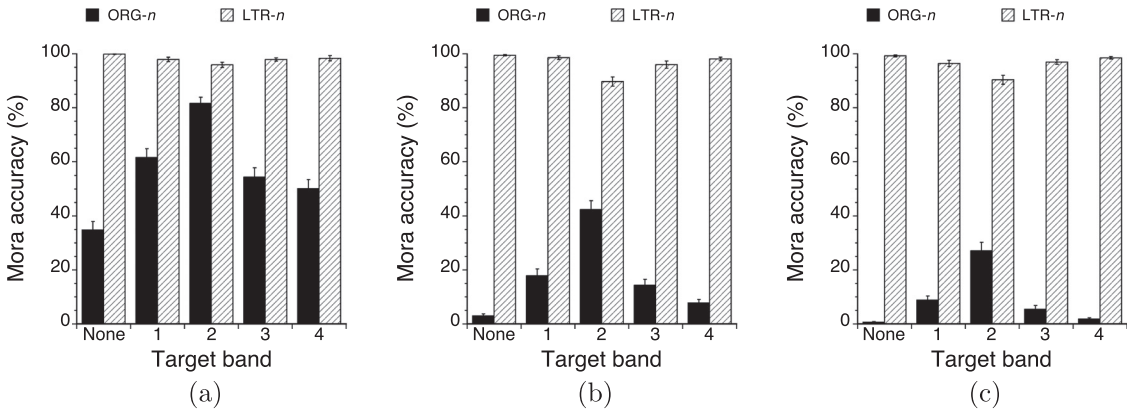

Fig. 2. Mean percentage of mora accuracy ($n = 18$) as a function of segment duration, target band, and stimulus types. Segment duration: (a) 75 ms, (b) 150 ms, and (c) 300 ms. The numbers 1–4 on the horizontal axis identify the four frequency bands from the lowest to the highest. In the legend, "ORG-n" identifies the stimulus type in which a target band (n) was kept intact and the other bands were locally time-reversed, and "LTR-n" identifies the stimulus type in which a target band (n) was locally time-reversed and the other bands were kept intact. Error bars, standard error of the mean (SEM).

incorrect. The number of correct morae in each sentence was counted. A blank response was counted as incorrect, and homophone errors were permitted. The percentage of correct morae was calculated for summarizing and presenting the results in a figure. Statistical analysis was based on the binomial results, i.e., correct or incorrect.

3. Results

Percentages of mora accuracy are represented in Fig. 2. The shape of the distribution remained roughly the same as the segment duration was extended from 75 over 150 to 300 ms. Among ORG-n stimuli, intelligibility was best preserved for the ORG-2 stimuli. Specifically, when the segment duration was 75 ms, intelligibility was 82% for the ORG-2 stimuli, whereas for the control stimuli (ORG-none), intelligibility dropped to 35%. When the segment duration was 150 ms, intelligibility was 42% for the ORG-2 stimuli, whereas for the ORG-none stimuli, intelligibility was just 3%. When the segment duration was 300 ms, intelligibility was still 27% for the ORG-2 stimuli, but only 1% for the ORG-none stimuli. For ORG-1, 3, and 4 stimuli, intelligibility was always lower than the intelligibility for the ORG-2 stimuli. The intelligibility for ORG-1 and -3 was comparable. The intelligibility for the ORG-4 stimuli (50%, 8%, and 2% at 75-, 150-, and 300-ms segment duration, respectively) was slightly better than the intelligibility for the control stimuli (ORG-none; 35%, 3%, and 1%, respectively).

For LTR-n stimuli, the intelligibility dropped mostly for the LTR-2 stimuli, for which intelligibility went down from 96% to 90% as segment duration increased from 75 to 300 ms. Intelligibility for the control stimuli (LTR-none), however, was invariably high, that is, more than 99%. The intelligibility for LTR-1 and -3 stimuli was comparable. The intelligibility for LTR-4 stimuli was indistinguishable from the intelligibility for the control stimuli (LTR-none).

The observations above were confirmed by the analyses using a generalized linear mixed model (GLMM), with a logistic linking function as implemented in an add-in for JMP (SAS Institute Inc., 2018), applied to the results of the two stimulus types separately. The data were analyzed for fixed effects of segment duration (continuous predictor), target band (categorical predictor), and their interaction and for a random effect of listener. For the ORG-n stimuli, the model revealed p values smaller than 0.001 in all effects: segment duration [$F(1, 1326) = 260.98$], target band [$F(4, 1324) = 55.72$], and their interaction [$F(4, 1323) = 11.29$]. For the LTR-n stimuli, this model revealed p values smaller than 0.05 in segment duration [$F(1, 1322) = 3.91$, $p < 0.048$] and target band [$F(4, 1322) = 25.02$, $p < 0.001$]. The p value for the interaction was 0.6. To examine whether or not the differences between target bands were reliable, Tukey–Kramer honestly significant difference (HSD) tests were applied. For the ORG-n stimuli, p values were smaller than 0.05 for the differences between all combinations of target bands except for the difference between 1 and 3 ($p = 0.21$). For the LTR-n stimuli, p values were smaller than 0.01 for the differences between 1 and 2, 1 and none, 2 and 3, 2 and 4, 2 and none, and 3 and none. Other p values (for 1 and 3, 1 and 4, 3 and 4, and 4 and none) exceeded 0.05.

4. Discussion

Summarizing the results, among the ORG-n stimuli, intelligibility was highest for the ORG-2 stimuli. Among the LTR-n stimuli, intelligibility was lowest for the LTR-2 stimuli. These features were observed irrespective of segment duration. Thus, band 2 (540–1700 Hz) mostly influenced intelligibility of the chimeric locally time-reversed speech stimuli. At the same time, intelligibility was comparable for either the pair of ORG-1 and -3 or the pair of LTR-1 and -3. Thus, the contribution to intelligibility by bands 1 and 3 was comparable too. The contribution by band 4 was the smallest—but distinct

in ORG-*n* conditions—whereas in LTR-*n* conditions, the contribution by band 4 was not apparent, probably because of a ceiling effect. The intelligibility of ORG-none control stimuli for which all frequency bands were locally time-reversed was 35%, 3%, and 1% at 75-, 150-, and 300-ms segment duration, respectively, which corresponded well with the previous results obtained with the same speech database (Ueda *et al.*, 2017).

This pattern of results cannot be attributed to the average power level differences between the bands, because an analysis of the average power levels observed in the same speech database and comparable frequency bands showed that the levels were highest in the lowest band (corresponding to band 1 in the current study), followed by the levels of the second lowest band (corresponding to band 2) and gradually going down in the third and fourth bands in this order (Ueda *et al.*, 2018). Thus, the order of the average power levels is at odds with the current results. In addition, an experiment in which the average power levels were exchanged between every possible pair of bands showed that the manipulation did not affect the intelligibility of four-band noise-vocoded speech, when the levels of bands 1 and 2, 1 and 3, and 3 and 4 were exchanged. The reduction in intelligibility caused by the level exchange between bands 1 and 4, 2 and 3, and 2 and 4 was marginal and 20% at most (Ueda *et al.*, 2018).

Thus, band 2 is most informative for speech intelligibility. When the band was degraded, the drop in intelligibility was the largest, whereas when the band was preserved, the intelligibility was preserved best. The present results are in accord with our previous findings obtained with the combination of a lower frequency band and a higher frequency band, in which either one of the bands was degraded and the other band was preserved (Matsuo *et al.*, 2020). Also, the current results are consistent with the results of the study on informational masking of three-formant speech analogs by extraneous formants (Roberts and Summers, 2018). Furthermore, the results are in line with the previous findings by Nakajima *et al.* (2017), suggesting that band 2 correlated mostly with sonority. This band corresponds to a factor with a single peak (Ueda and Nakajima, 2017), the *mid-low factor* (Nakajima *et al.*, 2017). It has been confirmed that vowels and sonorants dominate the factor (Nakajima *et al.*, 2017; Zhang *et al.*, 2020); thus, it is natural that band 2 has the closest connection with sonority and, hence, syllable formation. Bands 1 and 3 contributed also to intelligibility to some extent, although the contribution was less prominent compared to that of band 2. Nakajima *et al.* (2017), in fact, found moderate correlation coefficients (0.30–0.54) between sonority scales and the *low and mid-high factor*, which is bimodal and corresponds to bands 1 and 3 (Grange and Culling, 2018; Ueda and Nakajima, 2017). It is therefore plausible that the contributions by bands 1 and 3 to intelligibility were comparable, because the two bands are connected to the bimodal factor. The negative correlation coefficients (–0.45 to –0.28) observed by Nakajima *et al.* (2017) between sonority scales and the *high factor*, corresponding to band 4 in the current study, were not confirmed in the present results. This may be the limitation of the current experimental paradigm. Further investigations are warranted to clarify the issue.

Acknowledgments

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. JP19H00630. The authors would like to thank Yoshitaka Nakajima for providing J language software routines, Hikaru Eguchi for programming in LiveCode, Daiki Higuchi for running the experiment, Gerard B. Remijn for providing helpful comments on the draft, and Yoshitaka Nakajima and Hiroshige Takeichi for valuable discussion.

References and links

Boersma, P., and Weenink, D. (**2020**). "Praat: Doing phonetics by computer (version 6.0.21) [computer program]," http://www.praat.org/ (Last viewed 27 February 2021).

Clements, G. N. (**1988**). "The role of the sonority cycle in core syllabification," in *Working Papers of the Cornell Phonetics Laboratory*, Cornell University, Ithaca, NY. [Reprinted in J. Kingston and M. E. Beckman (1990). Papers in Laboratory Phonology 1: Between the Grammar and the Physics of Speech (Cambridge University, Cambridge, UK), pp. 283–333].

de Saussure, F. (**1916**). *Cours de linguistique générale* (*Course in General Linguistics*) (Éditions Payot & Rivages, Paris).

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**(6), 2984–2987.

Goad, H. (**2016**). "Sonority and the unusual behaviour of /s/," in *Challenging Sonority: Cross-Linguistic Evidence*, edited by M. J. Ball and N. Müller (Equinox, Sheffield, UK), pp. 21–44.

Grange, J., and Culling, J. (**2018**). "The factor analysis of speech: Limitations and opportunities for cochlear implants," Acta Acust. united Acust. **104**, 835–838.

Greenberg, S., and Arai, T. (**2004**). "What are the essential cues for understanding spoken language?," IEICE Trans. Inf. Syst. **E87-D**(5), 1059–1070.

Harris, J. (**1994**). *English Sound Structure* (Blackwell, Oxford).

Ishida, M., Samuel, A. G., and Arai, T. (**2016**). "Some people are 'more lexical' than others," Cognition **151**, 68–75.

J Software (**2020**). "The J programming language (version J901) [computer language]," http://www.jsoftware.com/ (Last viewed 20 March 2021).

LiveCode Community (**2018**). "LiveCode (version 9.0) [computer language]," https://livecode.org/ (Last viewed 17 August 2019).

Matsuo, I., Ueda, K., and Nakajima, Y. (**2020**). "Intelligibility of chimeric locally time-reversed speech," J. Acoust. Soc. Am. **147**(6), EL523–EL528.

Nakajima, Y., Ueda, K., Fujimaru, S., Motomura, H., and Ohsaka, Y. (**2017**). "English phonology and an acoustic language universal," Sci. Rep. **7**(46049), 1–6.

Parker, S. (**2012a**). "Introduction," in *The Sonority Controversy*, edited by S. Parker (De Gruyter Mouton, Berlin), pp. xi–xvi.
Parker, S. (ed) (**2012b**). *The Sonority Controversy* (De Gruyter Mouton, Berlin).
Prince, A., and Smolensky, P. (**2004**). *Optimality Theory: Constraint Interaction in Generative Grammar* (Blackwell, Malden, MA).
Rahilly, J. (**2016**). "Sonority in natural language: A review," in *Challenging Sonority: Cross-Linguistic Evidence*, edited by M. J. Ball and N. Müller (Equinox, Sheffield, UK), pp. 5–20.
Roberts, B., and Summers, R. J. (**2018**). "Informational masking of speech by time-varying competitors: Effects of frequency region and number of interfering formants," J. Acoust. Soc. Am. **143**(2), 891–900.
Saberi, K., and Perrott, D. R. (**1999**). "Cognitive restoration of reversed speech," Nature **398**(6730), 760–760.
SAS Institute Inc. (**2018**). "JMP (version 14.3.0) [computer program]," SAS Institute Inc., Cary, NC.
Selkirk, E. (**1984**). "On the major class features and syllable theory," in *Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students*, edited by M. Aronoff and R. T. Oehrle (MIT, Cambridge, MA), pp. 107–136.
Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cogn. Sci. **12**(5), 182–186.
Spencer, A. (**1996**). *Phonology: Theory and Description* (Blackwell, Oxford).
Steffen, A., and Werani, A. (**1994**). "Ein experiment zur zeitverarbeitung bei der sprachwahrnehmung" ("An experiment on temporal processing in speech perception"), in *Sprechwissenschaft & Psycholinguistik (Speech Science and Psycholinguistics)*, edited by G. Kegel, T. Arnhold, K. Dahlmeier, G. Schmid, and B. Tischer (Westdeutscher Verlag, Opladen, Germany), pp. 189–205.
Stilp, C. E., Kiefte, M., Alexander, J. M., and Kluender, K. R. (**2010**). "Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences," J. Acoust. Soc. Am. **128**, 2112–2126.
Teng, X., Cogan, G. B., and Poeppel, D. (**2019**). "Speech fine structure contains critical temporal cues to support speech segmentation," NeuroImage **202**, 116152.
Ueda, K., Araki, T., and Nakajima, Y. (**2018**). "Frequency specificity of amplitude envelope patterns in noise-vocoded speech," Hear. Res. **367**, 169–181.
Ueda, K., and Nakajima, Y. (**2017**). "An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech," Sci. Rep. **7**(42468), 1–4.
Ueda, K., Nakajima, Y., Ellermeier, W., and Kattner, F. (**2017**). "Intelligibility of locally time-reversed speech: A multilingual comparison," Sci. Rep. **7**(1782), 1–8.
Zhang, Y., Nakajima, Y., Ueda, K., Kishida, T., and Remijn, G. B. (**2020**). "Comparison of multivariate analysis methods as applied to English speech," Appl. Sci. **10**, 7076.
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.