Atten Percept Psychophys (2013) 75:1353–1358
DOI 10.3758/s13414-013-0542-x

Modulation sensitivity in the perceptual organization of speech

Robert E. Remez · Emily F. Thomas · Kathryn R. Dubowski · Stavroula M. Koinis · Natalie A. C. Porter · Nina U. Paddu · Marina Moskalenko · Yael S. Grossman

Published online: 13 September 2013

Abstract In a spoken utterance, a talker expresses linguistic constituents in serial order. A listener resolves these linguistic properties in the rapidly fading auditory sample. Classic measures agree that auditory integration occurs at a fine temporal grain. In contrast, recent studies have proposed that sensory integration of speech occurs at a coarser grain, approximate to the syllable, on the basis of indirect and relatively insensitive perceptual measures. Evidence from cognitive neuroscience and behavioral primatology has also been adduced to support the claim of sensory integration at the pace of syllables. In the present investigation, we used direct performance measures of integration, applying an acoustic technique to isolate the contribution of short-term acoustic properties to the assay of modulation sensitivity. In corroborating the classic finding of a fine temporal grain of integration, these functional measures can inform theory and speculation in accounts of speech perception.

Keywords Speech perception · Auditory sensory integration · Modulation sensitivity

R. E. Remez · E. F. Thomas · S. M. Koinis · N. A. C. Porter · N. U. Paddu
Department of Psychology and Program in Neuroscience & Behavior, Barnard College, Columbia University, New York, NY, USA

K. R. Dubowski
College of Physicians & Surgeons, Columbia University, New York, NY, USA

M. Moskalenko · Y. S. Grossman
Mount Sinai School of Medicine, New York, NY, USA

R. E. Remez (✉)
Department of Psychology, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027-6598, USA
e-mail: remez@columbia.edu

Speech exhibits nested linguistic properties: Clauses contain phrases, which contain words, which are composed of syllables, which comprise phonemic segments. The attributes at each scale are readily recognized, yet classic perceptual analyses of the information conveyed by speech have focused on the rapid rate of the production and perception of consonants and vowels, the elementary linguistic constituents that compose utterances (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Miller & Licklider, 1950). In ordinary circumstances, this rate might exceed a dozen segments per second. An acknowledgment of the rapidity of production underlies a foundational argument in cognitive science (Lashley, 1951), that utterances are planned and expressed, rather than triggered in chains of stimulus and response.

The projection of a fading auditory trace into durable linguistic form occurs with some urgency, according to classic estimates of auditory sensory decay: A trace fades in either 50 ms (Baddeley, 1986; Haggard, 1985; Lashley, 1951; Liberman et al., 1967), or <100 ms (Cudahy & Leshowitz, 1974; Elliot, 1967; Huggins, 1964; Miller & Licklider, 1950; Pisoni, 1973). Some more recent estimates converge on these measures: 50–100 ms (Fu & Galvin, 2001; Remez et al., 2010; Remez, Ferro, Wissig, & Landau, 2008). Nonetheless, a challenge to this estimate of fine-grained temporal sensitivity is posed by reports of sensory resolution at the far slower pace of syllables, between 3 and 8 Hz or 120 and 333 ms (Cherry, 1953; Drullman, Festen, & Plomp, 1994; Greenberg & Arai, 1998; Greenberg, Arai, & Grant, 2006; Saberi & Perrott, 1999).

One influential report noted perceptual sparing, despite temporal distortion approaching syllable duration (Saberi & Perrott, 1999; see also Steffen & Werani, 1994). In this method, a sample of speech was divided into equal intervals, each of which was reflected temporally, and the time-reversed segments were then sequenced in the original order, composing an utterance of veridically ordered time-reversed excerpts. The performance measures of the tolerance of temporal

distortion revealed that a reversed segment duration as great as 135 ms reduced the judged intelligibility merely by half. This was offered as evidence that neither a detailed sensory representation nor a perceptual analysis of the fine structure of the auditory stream is required for the recognition of linguistic properties. Yet, the acoustic technique used to estimate the effects of temporal distortion disrupts the acoustic modulation of speech, but not its short-term spectra. With time-reversed excerpts retaining the auditory quality of every vowel and of nasal, aspirate, and fricative consonants, this method presumably contaminates the assessment of time-critical modulation sensitivity with perceptual effects of timbre, which is unaltered by temporal distortion (see, e.g., Clarke, Becker, & Nixon, 1966; Van Lancker, Kreiman, & Emmorey, 1985). Accordingly, this confounding was likely to yield a falsely long estimate of the span of temporal integration. Moreover, the reliance on judged intelligibility with repeated exposure to test items, instead of a direct measure of intelligibility, was also likely to overestimate the tolerance for temporal distortion. A fairer test of modulation sensitivity might rely on a contingent task—that is, reports of the linguistic properties of unfamiliar utterances, not the subjective prominence of expected words—and would distinguish the effects of modulation from effects of the carrier spectrum.

Is auditory modulation sensitivity coupled to the rate of spoken syllables, 3–8 Hz? In a test of this claim, we report intelligibility measures of sentences exhibiting temporal distortion ranging from brief to moderate time spans. The findings corroborated both classic and recent reports that sensitivity to modulation in speech approximates the linguistic constituent of the phonetic segment, far briefer than the syllable. In extending the precedent, an assay was created to compare natural and sine-wave speech (Remez, 2008; Remez, Rubin, Pisoni, & Carrell, 1981), in order to estimate modulation sensitivity (Elliott & Theunissen, 2009; Greenberg & Arai, 2001) exclusive of the perceptual effects of short-term spectra. This empirical practice allows a test to use transcription accuracy, a contingent task that is simple for participants, and to measure modulation sensitivity independent of short-term auditory quality, a property unaffected by temporal reversal. Differing in short-term spectra only, the modulations of natural and sine-wave speech are matched. Indeed, the intelligibility difference reported here between natural speech and the sine-wave conditions is arguably due to the perceptual effect of short-term timbre, independent of modulation, and exposes the likely contribution of vocal quality in the precedent.

With no evident correspondence in the pace of syllables and the temporal grain of auditory integration, these new findings show that the syllable derives its cognitive importance from its linguistic function (see Peelle, Gross, & Davis, 2012), which weakens the claim (Ahissar et al., 2001; Kerlin, Shahin, & Miller, 2010; Luo & Poeppel, 2007) that cycles of brain activity at the approximate periodicity of syllables reflect a specifically sensory integrative function, or that a cortical cycle of this periodicity entrained a fundamental sensory function during primate evolution (Ghazanfar, Chandrasekaran, & Morrill, 2010; cf. MacNeilage, 1998).

Experiment

The method of the present project used the acoustic technique of Saberi and Perrott (1999), imposing temporal distortion on a speech waveform, but we sharpened the perceptual measures in two ways. First, a variety of sentences was used, in two acoustic forms, as natural samples and as sine-wave replicas. In addition to diversifying the variety of spoken items presented to listeners—the empirical precedent (Saberi & Perrott, 1999) had used a single natural utterance, and an extension had used nine (Kiss, Cristescu, Fink, & Wittmann, 2008)—these new tests also aimed to distinguish modulation sensitivity from the perceptual effects of short-term natural vocal timbre (e.g., Terasawa, Slaney, & Berger, 2005). Because some consonants and vowels briefly approximate stationary spectra, these impressions of familiar timbre are conserved despite temporal reversal, and arguably may retain their perceptual function whether a sample is temporally veridical or reversed. Sine-wave speech lacks the short-term spectral details of natural vocalization, and without familiar timbre the recognition of linguistic attributes rests largely on sensitivity to modulation, despite an unspeechlike subjective quality (Remez, 2008; Remez et al., 1981).

A second aspect of the procedure also improved the sensitivity of the test. Transcription accuracy was used here as a direct measure of intelligibility, in contrast to prior methods. Saberi and Perrott (1999) relied on indirect reports that a known sentence was spared subjective disruption by temporal distortion. An intelligibility measure was combined with the method of limits by Kiss et al. (2008), using ascending runs decreasing in the duration of reversed segments. In the present test, a listener was assigned to a single condition only, preventing a trial in one condition from influencing performance in another. As a control and replication, some listeners in the present study reported the extent to which a printed sentence shown during a trial remained intelligible, despite the imposition of temporal distortion on a natural sample. Adopting this method along with direct performance measures of intelligibility permitted a comparison of the present methods with the empirical precedent.

Method

Acoustic test materials

Twelve sentences (see the Appendix) spoken by one of the authors (K.R.D., an adult female) were sampled to disk at

44.1 kHz. The average syllable duration of these items was 277 ms (*SD* = 128 ms) excluding the final stressed syllables (average duration = 496 ms, *SD* = 124 ms), which lies within the range of 120–333 ms designated by the hypothetical syllable pace. Temporally distorted versions were created by reversing small portions of the waveform of each sample and assembling the reversed portions in veridical order (see Figs. 1 and 2). The reversal spans applied to the natural sentences were 0, 50, 75, 100, and 150 ms.

Unaltered natural samples of the 12 sentences were used as models for the creation of sine-wave speech. Estimates of the center frequency and amplitude of vocal resonances were created by hand and used as synthesis parameters for four time-varying sinusoids (see Remez et al., 2011). Temporally distorted versions were created by reversing a brief span of a waveform and composing a new waveform of reversed samples, preserving the original order. The reversal spans applied to sine-wave sentences were 0, 25, 50, and 75 ms.

Procedure

Each test session, we used 12 sentences of the same acoustic type, natural or sine-wave, the same temporal reversal, and the same response measure, transcription or the magnitude of subjective intelligibility. The design included 13 conditions overall, in three main tests: natural intelligibility, with reversal segments of 0, 50, 75, 100, and 150 ms; sine-wave intelligibility, with reversal segments of 0, 25, 50, and 75 ms; and judged subjective intelligibility, with reversal segments of 50, 75, 100, and 150 ms. These conditions were chosen to track performance in each of the three tests. In sessions testing intelligibility, a listener was instructed to transcribe each test sentence in a specially prepared booklet. In sessions replicating the method of reported subjective intelligibility, a listener read a printed version of a sentence before hearing it and indicated the apparent intelligibility by designating a magnitude ranging from 5 (*all words intelligible*) to 1 (*no words intelligible*). Each sentence was presented five times in succession, and all items presented within a test session shared the same temporal reversal.

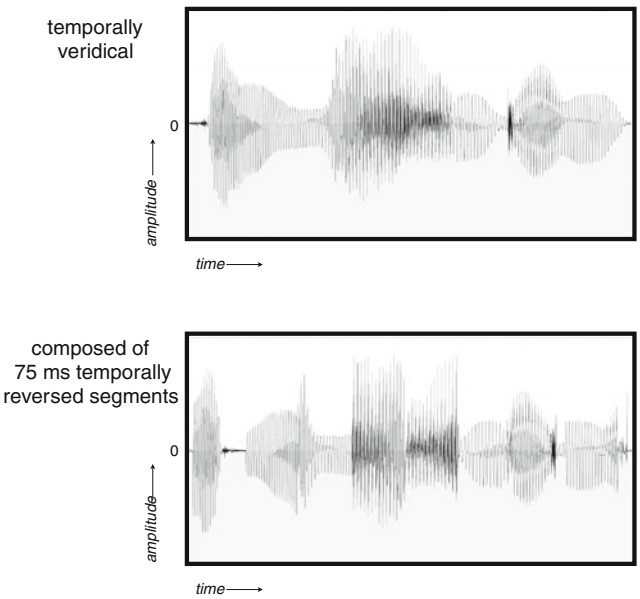

Fig. 1 Effect of temporal reversal of brief segments of a natural speech sample. (*Top*) A temporally veridical representation of the phrase "the winding," excerpted from the test item "Take the winding path to reach the lake." (*Bottom*) The waveform created by reversing 75-ms segments and recomposing the wave

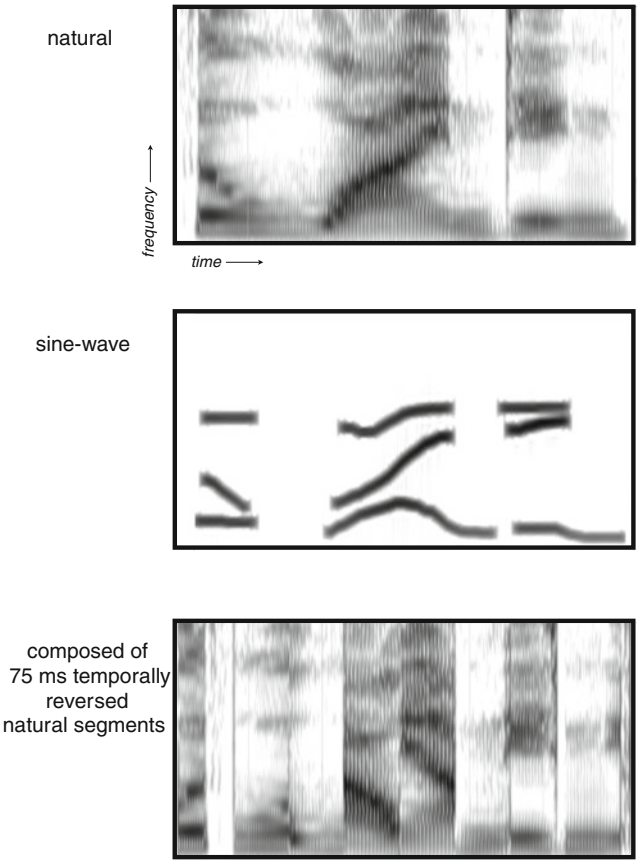

Fig. 2 Spectrographic comparison of natural speech (*top*), its sine-wave replica (*middle*), and the composite of temporally reversed 75-ms natural samples (*bottom*). The phrase is "the winding," excerpted from the test item "Take the winding path to reach the lake"

Participants

A total of 104 listeners were each assigned randomly to a test condition. Each participant was right-handed and reported no history of speech or hearing difficulty.

Results and discussion

Intelligibility performance was analyzed statistically using a one-way analysis of variance on the intelligibility parameter

for the natural and sine-wave sentences, and on the judged-intelligibility parameter for reports about natural sentences; each degree of temporal reversal that was tested was a treatment in the analysis. Performance differed significantly as a function of the duration of temporal reversal [natural judged, $F(3, 28) = 35.756$, $p = .0004$; natural intelligibility, $F(4, 35) = 333.197$, $p = .0004$; sine-wave intelligibility, $F(3, 28) = 37.289$, $p = .0004$]. The group performance is shown in Fig. 3; error bars portray 95 % confidence intervals. Significant differences between the individual treatment means may be seen directly in the figure.

The results of the three tests that we performed showed a clear pattern, with the first roughly replicating the finding of Saberi and Perrott (1999): When listeners knew the words composing the utterance in advance of the presentation and judgments of subjective intelligibility were used to estimate distortion tolerance, judged intelligibility declined by half when the reversal segment was 100 ms. If the sentences were not known in advance, transcription accuracy declined by half at a reversal segment of 75 ms, and at a reversal segment of 100 ms, the sentences were unintelligible. This difference as a consequence of the task is most likely due to the overestimation of distortion tolerance caused by the use of an indirect and subjective measure. Relying on transcription accuracy to estimate intelligibility, Kiss et al. (2008) used natural sentences, with each trial presenting a slightly less distorted sentence to the same listener. Ascending runs only occurred in this variant of the method of limits, and due to the cumulative effects of uncertainty across trials, it was likely to produce an underestimate of distortion tolerance. Indeed, they reported that intelligibility fell by half at a reversal segment of 50 ms, and sentences were unintelligible at 74 ms. Nonetheless, the present estimates and those reported by Kiss et al. are briefer than the hypothetical

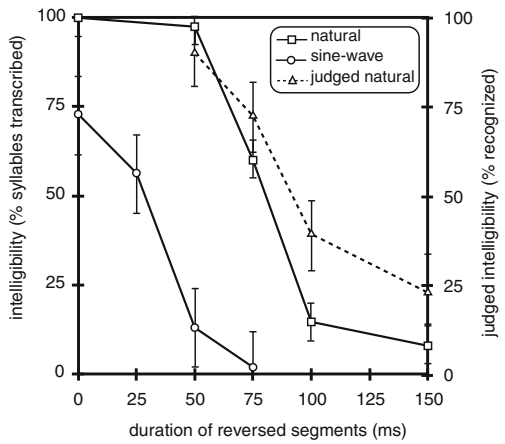

Fig. 3 Group performance on three perceptual tasks under differing grains of temporal reversal. Each *dot* represents the average performance of eight listeners and 12 sentences. *Error bars* show the 95 % confidence intervals estimated in the analyses of variance. Judged intelligibility was converted to an estimated percentage (scale on right) and plotted on the same frame with the direct measures of intelligibility

syllable range of 120–333 ms and are counterevidence to the claim that auditory integration of speech is intrinsically keyed to a syllabic rate.

The results of the sine-wave tests show that the intelligibility of intact sentences was good overall, but poorer than the natural items on which the synthesis was modeled. In the conditions with time-reversed segments, the intelligibility of sine-wave sentences was lost at a reversal segment as brief as 50 ms, which is evidence that sensitivity to modulation, independent of timbre, simply develops far more quickly than the syllable, arguably at the pace of the phonetic segment. These findings are approximate to independent measures of perceptual integration of speech signals with sparse acoustic spectra (Fu & Galvin, 2001; Remez et al., 2008; Silipo, Greenberg, & Arai, 1999) and provide a discriminating test of modulation sensitivity.

Psychoacoustically, a brief estimate of modulation sensitivity is plausible, although the performance-level difference between the natural and sine-wave conditions admittedly warrants caution. Performance was 25 % poorer with undistorted sine-wave items than with natural items, and the estimate of tolerable temporal reversal with sine-wave items was 50 % briefer than the estimate using natural items. One interpretation is that these differences reflect two consequences of the contribution to speech perception of short-term spectra and spectrotemporal modulation. When they combine, both aspects of performance are enhanced. When only modulation remains, intelligibility suffers a bit, and cognitive compensation for temporal distortion is hampered. Nonetheless, expressed in these measures might be a general relation between intelligibility and tolerance of temporal distortion. It must be conceded, however, that the technical literature has no precedent for this speculation. Moreover, it would be difficult to assess this conjecture parametrically—for instance, by titrating intelligibility in order to observe changes in distortion tolerance. Although some studies have used filtered, masked, reversed, or vocoded speech in order to preserve some acoustic properties of speech while reducing intelligibility, each of these manipulations disrupts the modulation characteristic of speech, and none is well suited for a direct investigation of modulation sensitivity. Because sine-wave synthesis and fine-grain acoustic chimeras (Smith, Delgutte, & Oxenham, 2002) retain the modulation characteristics of speech at fine frequency detail across 5 kHz, these methods are more appropriate. New tests that vary the distribution of phone classes systematically—fricatives, nasals, and liquids, for example—will also permit a parametric study of the independent effects of short-term timbre and sensitivity to modulation (see Remez et al., 2011).

In this conceptualization, the origin of modulation sensitivity is sensory, understood as an intrinsic function of an auditory system. Could this aspect of perceptual organization vary with the characteristics of an acoustic wave? Although it is customary to distinguish aspects of sensory function that are

fixed from those that are altered by attention, one recent study using a method similar to that of Saberi and Perrott (1999, and that of the present project) reported effects contingent on syllable rate (Stilp, Kiefte, Alexander, & Kluender, 2010). The method of their project imposed temporal reversals on brief segments of a waveform, in a test of the robustness of perception despite temporal distortion. Instead of natural speech, they used speech synthesized automatically from text, constructed to exhibit three different speech rates: slow, normal, and rapid. With intelligibility as the measure, the effect of temporal distortion appeared to vary with speech rate, and was very nearly a constant function of syllable rate, independent of absolute temporal characteristics. Stilp et al. concluded that the differences in distortion tolerance due to speech rate were attributable to a match between speech rate and the reversal segment that disrupted the syllables: Distortion of fast speech was relatively more harmed by brief reversal segments, slow speech by long reversal segments, and modal speech by reversal segments of intermediate duration. Although this report warrants caution in interpreting the present measures as the result of a fixed sensory function, the actual implications of the findings reported by Stilp et al. are less certain. To explain, synthetic speech was a surrogate for sampled speech in their test items, to make it feasible for them to vary speech rate with control. But, in assembling continuous speech from discrete segment-size samples, synthetic speech produced from text by unit selection (Hunt & Black, 1996) compromises the natural dynamics of speech acoustics, interpolating segments by algorithm rather than by the natural dynamics of coarticulation. Sine-wave speech is a form of copy synthesis that preserves the dynamics of the evolving utterances exactly. Moreover, in speech synthesized by unit selection, the compromise in the dynamics is great when the synthesis rate departs significantly from the original range of articulation rates at which the segmental templates were sampled. In the method of Stilp et al., the range of syllabic rates varied in the extreme, from 2.5 to 10 Hz (100–400 ms per syllable), incidentally exceeding the hypothetical range of modulation sensitivity proposed in this literature. No tests were reported of speech rates that varied in the natural range close to the modal rate. It will be useful to evaluate the effects of speech rate in new measures with realistic test materials. For now, to consider the condition in their report closest to the natural speech condition of the present test, the estimates of distortion tolerance coincide, despite a small difference in speech rates between the natural talker in this project and the synthetic talker in theirs.

Conclusion

Because the production of speech and its acoustic effects are structured in syllables, it has seemed reasonable at times for theorists to propose a reciprocal perceptual function exhibiting a grain of organization at the level of syllables, to a first approximation. Certainly, a widely influential view of the perception of speech is that perceptual integration occurs at the grain of syllables (Mehler, Dommergues, Frauenfelder, & Segui, 1981). One variant of this claim (Poeppel, 2003) describes the periodicity of cortical networks at roughly the same cycle rate that syllables are produced, and an extrapolation from this premise proposes that the phylogenetic age of this cortical pattern antedates speech and language (Ghazanfar et al., 2010). Syllables occur at 3–8 Hz, in this view, in order to coincide with the natural characteristics of a primate vocalization system exapted for speech (though see Fox & Cohen, 1977, for an equivalent in canid vocalization). However, direct performance estimates of the persistence of auditory sensory traces do not support the premise that the integration of sensory elements occurs at the slow pace of the syllable. It is far likelier that sensory samples are rapidly bound and resolved linguistically into aggregates approximate to syllables, a conceptualization consistent with measures that distinguish sensory and cognitive effects in the cortical accompaniment to speech (Peelle et al., 2012). Although a durable phonetic encoding persists after an auditory trace has decayed (e.g., Baddeley, 1986), tests with tones (Cudahy & Leshowitz, 1974; Elliott, 1967) and with speech (Pisoni, 1973) alike have noted the short span of an auditory trace, which fades so rapidly that very little remains after a tenth of a second. The findings reported here corroborate those psychoacoustic measures and can inform theory and speculation about fundamental functions in the perceptual neuroscience of speech.

Author note We thank Cecil Cornick for recording the natural speech samples, and David Pisoni, Philip Rubin, and Michael Studdert-Kennedy for advice on interpretation. This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (Grant No. DC000308).

Appendix: Sentences used in this study

Rake the rubbish up and then burn it.
On the islands the sea breeze is soft and mild.
A bath can cure a lot.
She called his name many times.
This is a grand season for hikes on the road.
Take the winding path to reach the lake.
The bill was paid every third week.
A pencil with black lead writes best.
Don't play dodge ball near the edge of the cliff.
The fear of flying caused goose bumps down his neck.
Glue the sheet to the dark blue background.
Look high in the sky at the hawk.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, 98,* 13367–13372.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Clarke, F. R., Becker, R. W., & Nixon, J. C. (1966). *Characteristics that determine speaker recognition (Electronic Systems Division, Air Force Systems Command Report ESDTR-66-638)*. Hanscom Field: Air Force Systems Command, Electronic Systems Division.

Cudahy, E., & Leshowitz, B. (1974). Effects of contralateral interference tone on auditory recognition. *Perception & Psychophysics, 15,* 16–20.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America, 95,* 2670–2680.

Elliot, L. L. (1967). Development of auditory narrow-band frequency contours. *Journal of the Acoustical Society of America, 42,* 143–153.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology, 5,* e1000302.

Fox, M. W., & Cohen, J. A. (1977). Canid communication. In T. A. Sebeok (Ed.), *How animals communicate* (pp. 728–748). Bloomington: Indiana University Press.

Fu, Q.-J., & Galvin, J. J., III. (2001). Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users. *Journal of the Acoustical Society of America, 109,* 1166–1172.

Ghazanfar, A. A., Chandrasekaran, C., & Morrill, R. J. (2010). Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech. *European Journal of Neuroscience, 31,* 1807–1817.

Greenberg, S., & Arai, T. (1998). Speech intelligibility is highly tolerant of cross-channel spectral asynchrony. In P. Kuhl & L. Crum (Eds.), *Proceedings of the Joint Meeting of the Acoustical Society of America and the International Congress on Acoustics* (pp. 2677–2678). Melville: Acoustical Society of America.

Greenberg, S, & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)* (pp. 473–476). Aalborg, Denmark: Aalborg University, Center for Personkommunikation.

Greenberg, S., Arai, T., & Grant, K. (2006). The role of temporal dynamics in understanding spoken language. In P. Divenyi, S. Greenberg, & G. Meyer (Eds.), *Dynamics of speech production and perception* (pp. 171–190). Amsterdam: IOS Press.

Haggard, M. (1985). Temporal patterning in speech: The Implications of temporal resolution and signal-processing. In A. Michelsen (Ed.), *Temporal resolution in auditory systems* (pp. 215–237). Berlin: Springer.

Huggins, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America, 36,* 1055–1064.

Hunt, A., & Black, A. W. (1996). Unit selection in a concatenative speech synthesis system using a large speech database. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-96* (pp. 373–376). Piscataway, NJ: IEEE.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party." *Journal of Neuroscience, 30,* 620–628.

Kiss, M., Cristescu, T., Fink, M., & Wittmann, M. (2008). Auditory language comprehension of temporally reversed speech signals in native and non-native speakers. *Acta Neurobiologiae Experimentalis, 68,* 204–213.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 421–461.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54,* 1001–1010.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21,* 499–511.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America, 22,* 167–173.

Peelle, J. E., Gross, J., & Davis, M. H. (2012). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23,* 1378–1387. doi:10.1093/cercor/bhs118

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260. doi:10.3758/BF03214136

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication, 41,* 245–255.

Remez, R. E. (2008). Sine-wave speech. In E. M. Izhikovitch (Ed.), *Encyclopedia of computational neuroscience* (p. 2394). San Diego: Scholarpedia.com.

Remez, R. E., Dubowski, K. R., Davids, M. L., Thomas, E. F., Paddu, N. U., Grossman, Y. S., & Moskalenko, M. (2011). Estimating speech spectra by algorithm and by hand for synthesis from natural models. *Journal of the Acoustical Society of America, 130,* 2173–2178.

Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S., & Davids, M. L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, & Psychophysics, 72,* 2054–2058. doi:10.3758/APP.72.8.2054

Remez, R. E., Ferro, D. F., Wissig, S. C., & Landau, C. A. (2008). Asynchrony tolerance in the perceptual organization of speech. *Psychonomic Bulletin & Review, 15,* 861–865. doi:10.3758/PBR.15.4.861

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. doi:10.1126/science.7233191

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature, 398,* 760. doi:10.1038/19652

Silipo, R., Greenberg, S., & Arai, T. (1999). Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. In *Eurospeech 1999* (pp. 2687–2690). Grenoble: ESCA.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416,* 87–90.

Steffen, A., & Werani, A. (1994). An experiment on temporal processing in language perception [In German]. In G. Kegel, T. Arnhold, K. Dahlmeier, G. Schmid, & B. Tischer (Eds.), *Sprechwissenschaft und Psycholinguistik 6. Beiträge aus Forschung und Praxis [Speech science and Psycholinguistics 6: Contributions from Research and Practice]* (pp. 189–205). Opladen: Westdeutscher Verlag.

Stilp, C. E., Kiefte, M., Alexander, J. M., & Kluender, K. R. (2010). Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of the Acoustical Society of America, 128,* 2112–2126.

Terasawa, H., Slaney, M., & Berger, J. (2005). A timbre space for speech. In *Proceedings of Interspeech 2005* (pp. 1729–1732). Lisbon: ISCA.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part 1: Recognition of backward voices. *Journal of Phonetics, 13,* 19–38.