Published in final edited form as:

Exp Aging Res. 2016 ; 42(1): 4–17. doi:10.1080/0361073X.2016.1108741.

Constraints on Sensitivity to Auditory Modulation in the Perceptual Organization of Speech

Robert E. Remez, Emily F. Thomas, Andrea M. Wycoff, Rebecca E. Giglio, Aislinn T. Crank, Chloe B. Cheimets, and Stavroula M. Koinis

Department of Psychology and Program in Neuroscience & Behavior, Barnard College, Columbia University, New York, New York U. S. A

Abstract

Background—The perception of speech requires the integration of sensory details from a rapidly fading trace of a time-varying spectrum. This effortful cognitive function has been difficult to assess. New tests measuring intelligibility of sine-wave replicas of speech provided an assay of this critical function in normal hearing young adults.

Method—Four time-varying sinusoids replicated the frequency and amplitude variation of the natural resonances of spoken sentences. The temporal tolerance of perceptual integration of speech was measured by determining the effect on intelligibility of desynchronizing a single sine-wave component in each sentence. This method was applied in tests in which the sentences were temporally compressed or expanded over a 40% range.

Results—Desynchrony was harmful to perceptual integration over a narrow temporal range, indicating that modulation sensitivity is keyed to a rate of 20 Hz. No effect of variation in speech rate was observed on the intelligibility measure, whether rate was accelerated or decelerated relative to the natural rate.

Conclusion—Performance measures of desynchrony tolerance did not vary when speech rate was accelerated or decelerated, revealing constraints on integration that are arguably primitive, sensory, auditory, and fixed. Because these are not adaptable, they limit the potential for perceptual learning in this aspect of perceptual organization. Implications for describing the elderly listener are noted.

Keywords

speech perception; modulation sensitivity; auditory integration

A listener who resolves the linguistic properties of a talker's speech has succeeded in two perceptual acts, one contingent on the other. First, the diverse acoustic properties issuing from a single vocal source, and their consequent auditory samples, have been recognized as a single stream segregated perceptually from concurrent sounds. In this function, the whistles, clicks, hisses buzzes and hums composing speech cohere perceptually despite their

acoustic variety and discontinuity. By meeting this fundamental challenge, a second perceptual function becomes possible, namely, to analyze a coherent sensory stream for its linguistic and personal attributes (Remez, 2005; Remez, Rubin, Berns, Pardo & Lang, 1994). Auditory sensory samples of speech fade rapidly (Cowan, 1984), yet perceivers are exquisitely sensitive to the spectrotemporal pattern of a speech spectrum. In turn, the rapid fading of an auditory trace imposes a distinct urgency on functions that serve speech perception because, were a mistaken organization or analysis to occur, little of a raw sensory sample persists to allow a second attempt (Pisoni, 1973, 1975).

Nonetheless, estimates of the time course of auditory sensory integration in the perceptual organization of speech vary. The nesting of linguistic constituents of speech invites this variation, inherently, for the symbolic properties produced by perceptual analysis appear at differing timescales: sentences at 2 s, clauses at 1 s, phrases at 750 ms, words at 500 ms, syllables at 200 ms, and segments at 50 ms. At a span as great as a sentence, the resources of working memory obviously supply the power to integrative functions which accumulate the sensory details before they decay, project them into a durable symbolic form which can be committed to memory, short-term or long-term, and compose an impression of a series of clauses of the present utterance (Baddeley, 1986). But, at the brief end of this linguistic gradient, estimates of sensory integration have differed, ranging from segmental grain to suprasyllabic grain. In addition, some reports describe a sensory integration window of fixed temporal span (Remez, Ferro, Dubowski, Meer, Broder & Davids, 2010). Yet, because the rate of speech is variable, the modulation characteristic of the auditory samples of speech also varies. Accordingly, some projects have sought evidence of the adaptability of auditory integration, manifest as a variable temporal span of integration (Stilp, Kieft, Alexander, & Kluender, 2010). In the present investigation, direct tests were applied to calibrate this aspect of speech perception, and the measures indicate clearly that the rate of integration is rapid — 50 ms or so — and fixed, arguably set by limits of the physiology and biophysics of the auditory periphery.

Segments or syllables? Fixed or adaptable?

Several methods have been applied to calibrate the time course of auditory integration. All have assumed that this perceptual function is limited by the brief persistence of the auditory sensory trace of speech. Among these are: the greatest rate of interruption of a speech stream that spares intelligibility (Miller & Licklider, 1950); the greatest rate of ear switching of a speech signal that preserves intelligibility (Huggins, 1964; Stewart, Yetton & Wingfield, 2008); the subphonemic discriminability of syllables presented successively over a variable interval (Howell & Darwin, 1977; Pisoni, 1973, 1974); the greatest bearable span of temporally reversed sections of speech conserving order (Remez, Thomas, Dubowski, Koinis, Porter, Padu, Moskalenko & Grossman, 2013; Saberi & Perrott, 1999); and, the tolerable lag in some components of a speech signal rendered desynchronous by digital techniques (Fu & Galvin, 2001; Remez, Ferro, Wissig & Landau, 2008). Converging evidence has also been offered based on an observed 3–8 Hz period of cyclical cortical activity accompanying auditory exposure to utterances (Doelling, Amal, Ghitza & Poeppel, 2014; Luo & Poeppel, 2007; Peelle, Gross & Davis, 2013). Claims about integration have also been theoretically motivated based on the estimated rate of decay of an auditory trace

(Haggard, 1985), or the intrinsic temporal properties of segmental phonetic contrasts (Lashley, 1951; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). Overall, the estimates of auditory perceptual integration encourage a conclusion that this time-critical aspect of perception is faster than the procession of syllables.

It remains to be determined whether the window of integration is fixed or adaptable. Were measures to show that integration is adaptable, varying in timespan, it would be plausible to claim that a single function, varying between segmental and syllabic grain, is responsible in the auditory integration of speech. Alternatively, were measures to show that the window of integration is fixed, conceivably imposed by a constant rate of auditory decay, this would establish a clear constraint on descriptions of the earliest stage of speech perception.

A new project described in this report estimated the temporal span of the window of auditory integration. To create a sensitive indicator, the empirical method combined a measure of sensitivity to spectral modulation independent of vocal timbre, and a measure of tolerance of perceptual organization to temporal distortion. By using sentence-length test items that were sine-wave replicas of speech, the measure targeted modulation sensitivity directly. This acoustic technique models the spectrotemporal properties of speech with a small set of time-varying sinusoids, each defined by the estimated center frequency and amplitude of a natural vocal resonance (Remez, Rubin, Pisoni & Carrell, 1981). Three oral resonances, or formants, vary critically for intelligibility, and these are modeled with a corresponding set of three sine-waves. Other momentary effects — release bursts, aspirations, frications, murmurs — are also modeled sinusoidally with a fourth sine-wave that follows the center-frequency of a transitory acoustic property. Acoustic materials of this kind permit a perceptual assay of sensitivity to coherent variation and prevent the momentary effects of familiar timbre from affecting a straightforward measure of intelligibility. Indeed, in order to group time-varying sinusoids into a coherent signal, a listener must tolerate the absence of natural acoustic products of vocalization and the familiar timbres these spectral elements evoke. Moreover, the listener must disregard the nonvocal quality of the contrapuntal tones in resolving the coherence among the auditory constituents of an evolving utterance (Remez, 2005).

To assess the timespan of the window of auditory sensory integration, a set of test sentences was composed which parametrically varied temporal distortion in two dimensions. A temporally veridical set of sine-wave sentences was first composed which replicated the spectrotemporal properties of the resonant peaks of the original natural utterances. Then, relying on digital synthesis techniques, a set of variants was composed in which the second formant tone was desynchronized relative to the remainder of the tone pattern, the extent of desynchrony ranging in steps of 50 ms from ± 100 ms (*cf.* Remez et al., 2008). When a listener heard a test item with a desynchronized tone, the ability to integrate the tone as a coherent part of the sine-wave utterance indicated that the sensory trace of the tone remained available for integration. To assess the adaptability of this span of integration, two sets of sentences were derived from the originals in which the rate of synthesis was varied over a range of 40%. Specifically, the test items were accelerated by 20% or slowed by 20% relative to the veridical spectrotemporal pattern. If the span of auditory sensory integration is moderated with the rate of auditory variation, then this should have affected the

intelligibility of desynchronized items as a function of auditory rate, as Stilp et al. (2010) concluded based on an indirect measure using unit-selection synthesis and extremes of rate variation spanning 200% (see Remez et al., 2013). In their procedure, the span of integration was briefer when speech-rate was speeded, and the span was longer when speech-rate was slowed. Instead, in the present tests, the tolerability of desynchrony did not vary with the synthesis rate, indicating that the window of auditory sensory integration is apparently brief and fixed.

Method

Acoustic test materials

To compose a sensitive test for calibrating the perceptual tolerance of temporal desynchrony applied to a component of a speech spectrum, the acoustic materials must be intelligible at a high level when integration occurs, but unintelligible when temporal distortion is sufficient to block integration, even if a desynchronized component remains fully audible. To meet this empirical aim, the items used in these tests were derived from materials originally created for Remez et al. (2008; 2010) which sought to sharpen the indices of Greenberg and Arai (1998) and Silipo, Greenberg and Arai (1999). Fifteen sentences spoken by an adult male provided the natural samples from which the test items were produced. The natural items were spoken by a male talker (R.E.R.) seated in a sound-attenuating chamber, sampled at 22.05 kHz, equated for amplitude, and analyzed to compose parameters for a sine-wave synthesizer. Detailed description of sine-wave synthesis is provided by Remez et al. (2011). In this form of copy synthesis, the acoustic resonances, bursts, frictions, and murmurs are estimated by hand-tracing a spectrographic display of a natural utterance, compiling a synthesis table. This table represents the frequency and amplitude parameters of four time-varying sinusoids at a grain of 10 ms throughout each utterance. To perform the synthesis, waveforms were calculated with 16-bit amplitude resolution at a rate of 44.1 kHz and were stored in sampled-data format (Rubin, 1980).

The fifteen sine-wave sentences used in these tests allowed a sensitive assay of perceptual integration. Specifically, pilot testing revealed that each sentence was highly intelligible when all of the tone components were present, yet was unintelligible, or nearly so, when the tone analog of the second formant was removed. Intact, the sentences selected for this procedure were identified at an average of 98% correct; when the second formant analogue was absent, identification fell to an average of 6% correct (Remez et al., 2008). It may be useful to note that not all utterances meet this standard. A strict criterion was used to eliminate any sentence which remained partly or largely intelligible when lacking a second formant, due to phonemic composition or phonetic expression. As a consequence, the test provided a wide range in performance level within which to observe effects of temporal perturbation on perceptual integration. Moreover, when intelligibility was non-zero, this could be taken as evidence that all of the tone components were grouped into a single perceptual stream despite any temporal desynchrony. Each sentence that met this standard came from the well-used phonemically balanced IEEE set (Egan, 1948) or the Speech Perception in Noise set (Kalikow, Stevens, & Elliott, 1977). Test sentences are listed in an Appendix.

Accelerated and decelerated versions of each temporally veridical sine-wave sentence were created by altering the temporal parameters in the synthesis tables, producing a uniform strain which speeded or slowed each acoustic pattern. The original synthesis table had a 10 ms grain. By synthesizing a sine-wave pattern from these values at a temporal grain of 8 ms, a set of variants was created that exhibited the same frequency excursions and amplitude changes at a rate accelerated by 20%. In complementary fashion, by synthesizing the parameters at a 12 ms grain, the spectrotemporal variation in the tone pattern was decelerated by 20%.

Once the accelerated and decelerated versions of the intact sentences were created, it was a simple matter to create temporally desynchronized versions by offsetting the tone analog of the second formant relative to the other tone components. By altering the synthesis table, the second formant tone was made to lead or to lag the other three tones by 50 ms or 100 ms, creating five variants of each of fifteen sine-wave sentences. Figure 1 portrays variation in synthesis rate and desynchronized acoustic constituents schematically.

Test items were stored in sampled-data format, and were played for listeners at the time of testing. All test items were presented binaurally at a nominal level of 68 dB SPL via Beyerdynamic DT770 headphones to listeners seated in a sound attenuating chamber.

Procedures

Ten test sessions were conducted. Five different desynchronies were applied to the tone analog of the second formant at two rates of synthesis. The five desynchronies were 100 ms lead, 50 ms lead, synchronized, 50 ms lag and 100 ms lag. The two rates of synthesis were 20% accelerated and 20% decelerated, relative to the temporally veridical rate. Tests were blocked by synthesis rate and desynchrony. Within a session, 15 sine-wave sentences were used, each exhibiting a consistent synthesis rate and degree of desynchrony in the second formant tone. The order of presentation of sentences was random. A single trial consisted of five repetitions of a sine-wave sentence, during which a listener transcribed it by writing in a specially prepared booklet.

Participants

Eighty young adult listeners were tested (64 female, mean age = 19;6, SD = 1;8, range 18–29; 16 male, mean = 21;0, SD = 1;6, range 18–23), assigned randomly to one of the ten test conditions. Each participant was a self-reported normal hearing native speaker of English recruited from the undergraduate population of Columbia University. By participating, a listener satisfied a course requirement. To assess the susceptibility of each listener to sine-wave speech, a brief transcription pretest of six intact, temporally veridical novel sine-wave sentences was used, and inclusion in the study was based on performance on this test. The experimental protocol was reviewed and approved by the Barnard College Institutional Review Board for the Protection of Human Subjects from the Risks of Research Participation.

Results and Discussion

Each listener contributed fifteen measures, the percent of syllables in each test sentence transcribed correctly. The performance level at each of the five desynchronies within a condition of synthesis rate was the average of these transcriptions. Group averages of each desynchrony condition in the two rate conditions are shown in Figure 2, with each point representing the mean of 120 measures. A two-way repeated measures ANOVA on the transcription measures for the factors SYNTHESIS RATE and DESYNCHRONY revealed a significant effect of desynchrony on performance level [$F(4,70) = 94.132, p < .0005$, partial $\eta^2 = .843$]. There was no main effect of synthesis rate [$F(1,70) = 0.594, p = .443$], nor interaction of synthesis rate and desynchrony [$F(4,70) = 1.373, p = .252$].

Figure 2 shows the performance measures of the tests. Error bars represent the 95% confidence interval for differences in desynchrony collapsing over synthesis rate. For comparison, a solid line with round bullets shows the performance reported by Remez et al. (2010) on tests with the same desynchronies of the tone analog of the second formant of temporally veridical sine-wave sentences.

A comparison of performance level among the three plots shows that the perceptual organization of desynchronized sine-wave components was unaffected over a wide range of rate variation. Although a portion of a desynchronized second formant analog remained available for combination with the rest of the tones at a lead or lag of 50 ms, performance declined precipitously at desynchronies greater than this, notwithstanding variation in the rate of auditory modulation. To a first approximation, it appears that the window of sensory integration or perceptual organization is slightly wider than 50 ms, a hypothetical point of desynchrony at which an audible but desynchronized second formant analog is too discrepant, temporally, to integrate with the rest of the tone pattern, and, in consequence, intelligibility is lost. The congruence of the three functions shows that this point is neither determined nor influenced by the rate of variation characteristic of an auditory sensory sample of speech.

The measures reported here agree with psychoacoustic findings that auditory sensory integration is possible only over a brief time-span (Fu & Galvin, 2001; Haggard, 1985; Miller & Licklider, 1950; Pisoni, 1975; Remez et al., 2008). The fading of auditory sensory samples might impose this limit intrinsically if two conditions constrain speech perception. The first condition is that perceptual organization might warrant an initial integration of concurrently available sensory samples differing greatly in pitch, across as much as seven octaves. A requirement to integrate concurrent sensory effects also entails broadband sensitivity to coherent variation, in contrast to piecemeal assessment of isolated auditory elements, or “speech cues” in familiar parlance. The second principle is that the fast rate at which volatile auditory samples fade is set by the intrinsic temporal resolving power of the auditory pathway. If an auditory sensory sample of speech decays in little more than a twentieth of a second, this would limit its availability for integration in the best scenario, because once a portion of a sensory sample of speech has faded, it is lost from further opportunity for inclusion in an organized auditory stream.

In this analysis, the persistence of auditory sensory samples over a brief span is sufficient nonetheless to allow the perceptual resolution of the coherent modulation within a speech spectrum. The presence of noise in the auditory channel, whether external in origin or due to the physiology itself, is likely to impair the detection of coherent variation over such short intervals. Likewise, loss of temporal acuity is likely to narrow the interval during which a clear sample of an incident spectrum is available. Seen from this perspective, the function of perceptual organization entails the composition of a broadband stream of diverse acoustic composition by detecting the coordinate variation of its components in the short-term. The mere apprehension of the acoustic constituents of speech in an incident spectrum is insufficient.

With respect to the perceptual organization of speech across the lifespan, although this function is primitive, appearing by the fourteenth week of life in infancy (Eimas & Miller, 1992), little is known about the ability of older listeners to find and to follow the acoustically heterogeneous constituents of an evolving speech stream. The function of perceptual organization obliges a perceiver, young or old, to resolve the time-varying pattern of a speech spectrum independent of the specific physical characteristics of the carrier, as studies of sine-wave speech (Remez, 2005), noiseband vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) and chimerical speech (Smith, Delgutte & Oxenham, 2002) have shown. Further, this sensitivity to pattern irrespective of the spectral elements which compose it is arguably a key feature of the robustness of speech to distortion. Although it is reasonable to expect that age-related changes in auditory temporal acuity hamper the faithfulness of auditory samples of the speech spectrum, the present measures in the technical literature do not permit a distinction between evidence of integration failure – in which a critical constituent of a speech spectrum was missed – or of temporal blur in the sampling of an integrated component.

Measures of normal hearing elderly listeners have shown that decreases in temporal sensitivity impair receptive language (Gordon-Salant & Fitzgibbons, 1993). Nonetheless, the gold-standard psychoacoustic measures of temporal acuity do not offer a straightforward parallel to the temporally constrained functions of integration of a broadband speech spectrum described here. To calibrate auditory temporal resolving power, measures of gap detection or duration discrimination have been used, among others (for instance, Bergman, Blumenfeld, Cascardo, Dash, Levitt & Margulies, 1976; Gordon-Salant, Fitzgibbons & Yeni-Komshian, 2011). Such measures of temporal resolving power appear at first to be related to the temporal characteristics of auditory integration, though the sensitivity to the presence or absence of a brief interruption of a steady tone or band-limited noise – essentially a detection of discontinuity in an acoustic form – differs from an organizational function which rapidly amalgamates the intrinsically discontinuous constituents of speech spectrum into a single perceptual stream. Likewise, the detection of the smallest duration difference distinguishing two auditory forms offers no direct premise for estimating how a perceiver collects a rapidly fading diversity of simultaneous and successive auditory properties of speech differing in both frequency and spectrum.

It remains for new studies to determine whether aging changes the perceptual integration of speech directly, or whether indirect effects on auditory sensitivity coarsen frequency

resolution, add noise, affect sensory persistence directly or impair the focus or suppleness of auditory attention. An understanding of the conditions required for this first step in spoken communication can sharpen the account of language use, generally, and offers the potential to identify prospects for intervention when these systems are weakened. Inasmuch as slowing with age is well demonstrated for cognitive functions in spoken communication (Schneider, Daneman & Murphy, 2005; Wingfield, Tun, Koh & Rosen, 1999), temporal fidelity in the auditory neural pathway might also be affected by aging and, in consequence, might alter the sensory acuity required for the perceptual organization of speech.

References

- Baddeley, AD. Working Memory. Oxford: Clarendon Press; 1986.
- Bergman M, Blumenfeld VG, Cascardo D, Dash D, Levitt H, Margulies MK. Age-related decrement in hearing for speech: Sampling and longitudinal studies. *Journal of Gerontology*. 1976; 31:533–538.10.1093/geronj/31.5.533 [PubMed: 950447]
- Cowan N. On short and long auditory stores. *Psychological Bulletin*. 1984; 96:341–370.10.1037/0033-2909.96.2.341 [PubMed: 6385047]
- Doelling KB, Amal LH, Ghitza O, Poeppel D. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*. 2014; 85:761–768.10.1016/j.neuroimage.2013.06.035 [PubMed: 23791839]
- Egan JP. Articulation testing methods. *Laryngoscope*. 1948; 58:955–991.10.1288/00005537-194809000-00002 [PubMed: 18887435]
- Eimas PD, Miller JL. Organization in the perception of speech by young infants. *Psychological Science*. 1992; 3:340–345.10.1111/j.1467-9280.1992.tb00043.x
- Fu QJ, Galvin JJ III. Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users. *Journal of the Acoustical Society of America*. 2001; 109:1166–1172.10.1121/1.1344158 [PubMed: 11303930]
- Gordon-Salant S, Fitzgibbons PJ. Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research*. 1993; 36:1276–1285.10.1044/jshr.3606.1276 [PubMed: 8114494]
- Gordon-Salant S, Fitzgibbons P, Yeni-Komshian GH. Auditory temporal processing and aging: Implications for speech understanding of older people. *Audiology Research*. 2011; 1:9–15.10.4081/audiores.2011.e4
- Greenberg, S.; Arai, T. Speech intelligibility is highly tolerant of cross-channel spectral asynchrony. In: Kuhl, P.; Crum, L., editors. *Proceedings of the Joint Meeting of the Acoustical Society of America and the International Congress on Acoustics*. Melville, NY: Acoustical Society of America; 1998. p. 2677–2678.
- Haggard, M. Temporal patterning in speech: The implications of temporal resolution and signal-processing. In: Michelson, A., editor. *Time Resolution in Auditory Systems*. Heidelberg: Springer Verlag; 1985. p. 1–23.
- Howell P, Darwin CJ. Some properties of auditory memory for rapid formant transitions. *Memory & Cognition*. 1977; 5:700–708.10.3758/BF03197419 [PubMed: 24203289]
- Huggins AWF. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*. 1964; 36:1055–1064.10.1121/1.1919151
- Kalikow DN, Stevens KN, Elliot LL. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*. 1977; 61:1337–1351.10.1121/1.381436 [PubMed: 881487]
- Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. *Cerebral Mechanisms in Behavior*. New York: Wiley; 1951. p. 112–131.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:421–461.10.1037/h0020279

- Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 2007; 54:1001–1010.10.1016/j.neuron.2007.06.004 [PubMed: 17582338]
- Miller GA, Licklider JCR. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*. 1950; 22:167–173.10.1121/1.1906584
- Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*. 2013; 23:1378–1387.10.1093/cercor/bhs118 [PubMed: 22610394]
- Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*. 1973; 13:253–260.10.3758/BF03214136 [PubMed: 23226880]
- Pisoni DB. Auditory short-term memory and vowel perception. *Memory & Cognition*. 1975; 3:7–18.10.3758/BF03198202 [PubMed: 24203819]
- Remez RE. Perceptual organization of speech. In: Pisoni, DB.; Remez, RE., editors. *The Handbook of Speech Perception*. Oxford: Blackwell; 2005. p. 28-50.
- Remez RE, Dubowski KR, Davids ML, Thomas EF, Paddu NU, Grossman YS, Moskalenko M. Estimating speech spectra for copy synthesis by linear prediction and by hand. *Journal of the Acoustical Society of America*. 2011; 130:2173–2178.10.1121/1.3631667 [PubMed: 21973371]
- Remez RE, Ferro DF, Dubowski KR, Meer J, Broder RS, Davids ML. Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception & Psychophysics*. 2010; 72:2054–2058.10.3758/BF03196682
- Remez RE, Ferro DF, Wissig SC, Landau CA. Asynchrony tolerance in the perceptual organization of speech. *Psychonomic Bulletin & Review*. 2008; 15:861–865.10.3758/PBR.15.4.861 [PubMed: 18792517]
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychological Review*. 1994; 101:129–156.10.1037/0033-295X.101.1.129 [PubMed: 8121955]
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981; 212:947–950.10.1126/science.7233191 [PubMed: 7233191]
- Remez RE, Thomas EF, Dubowski KR, Koinis SM, Porter NAC, Paddu NU, Moskalenko M, Grossman YS. Modulation sensitivity in the perceptual organization of speech. *Attention, Perception & Psychophysics*. 2013; 75:1353–1358.10.3758/s13414-013-0542-x
- Rubin, PE. Internal memorandum. Haskins Laboratories; New Haven, Connecticut: 1980. Sinewave synthesis.
- Saberi K, Perrott DR. Cognitive restoration of reversed speech. *Nature*. 1999; 398:760. [PubMed: 10235257]
- Schneider BA, Daneman M, Murphy DR. Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging*. 2005; 20:261–271.10.1037/0882-7974.20.2.261 [PubMed: 16029090]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304.10.1126/science.270.5234.303 [PubMed: 7569981]
- Silipo, R.; Greenberg, S.; Arai, T. Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. *Proceedings of Eurospeech*; 1999; Grenoble: European Speech Communication Association; 1999. p. 2687-2690.
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. 2002; 416:87–90.10.1038/416087a [PubMed: 11882898]
- Stewart R, Yetton E, Wingfield A. Perception of alternated speech operates similarly in young and older adults with age-normal hearing. *Perception & Psychophysics*. 2008; 70:337–345.10.3758/PP.70.2.337 [PubMed: 18372754]
- Stilp CE, Kieft M, Alexander JM, Kluender KR. Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences. *Journal of the Acoustical Society of America*. 2010; 128:2112–2126.10.1121/1.3483719 [PubMed: 20968382]
- Wingfield A, Tun PA, Koh CK, Rosen MJ. Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech. *Psychology and Aging*. 1999; 14:380–389.10.1037/0882-7974.14.3.380 [PubMed: 10509694]

APPENDIX

Sentences used in the tests

A pencil with black lead writes best.

Cut the meat into small chunks.

Football is a dangerous sport.

He ran halfway to the hardware store.

Her purse was full of useless trash.

His boss made him work like a slave.

Press the pants, and sew a button on the vest.

The bark of the pine tree was shiny and dark.

The beauty of the view stunned the young boy.

The bill was paid every third week.

The drowning man let out a yell.

The sandal has a broken strap.

The steady drip is worse than a drenching rain.

The watchdog gave a warning growl.

Two blue fish swam in the tank.

Figure 1.

Schematic representation of the acoustic variations used in the present tests. Each panel depicts the phrase, “sew a button,” from the sentence, “Press the pants and sew a button on the vest.” Time ticks on the abscissa are 100 ms apart; frequency ticks on the ordinate mark kHz. A temporally veridical sine-wave version is shown (top center panel), in the next row the left panel depicts synthesis decelerated by 20%, and the right panel shows synthesis accelerated by 20%. In the bottom row, the left panel shows the pattern produced by desynchronizing the tone analog of the second formant, to make it lead the remaining tones by 50 ms. The 20% decelerated pattern is shown. The bottom right panel shows the pattern produced by desynchronizing the tone analog of the second formant to make it lag the remaining tones by 50 ms. The 20% accelerated pattern is shown.

Figure 2.

The results of ten tests of the effects of synthesis rate on the perceptual tolerance of desynchrony. The ordinate is intelligibility (transcription performance). The abscissa indicates the temporal offset of the tone analog of the second formant relative to the other components of the sine-wave sentence. Intelligibility measures of the effects of a desynchronized tone component in temporally veridical sentences (from Remez et al., 2010) are shown in the round bullets.