RUNNING HEAD: Altering speech rate

Altering context speech rate can cause words to (dis)appear

Laura C. Dilley[1,2,3,4], and Mark A. Pitt[5]

[1]Department of Communicative Sciences and Disorders, Michigan State University

[2]Department of Psychology, Michigan State University

[3]Department of Psychology, Bowling Green State University

[4]Department of Communication Sciences and Disorders, Bowling Green State University

[5]Department of Psychology, Ohio State University

Address correspondence to:

Dr. Laura C. Dilley

Michigan State University

Department of Communicative Sciences and Disorders

Department of Psychology

116 Oyer Building

East Lansing, MI 43403

Email: ldilley@msu.edu

Phone: 517-884-2255

Fax: 517-353-3176

Abstract

Speech is produced over time, making sensitivity to timing between speech events crucial for understanding language. Two experiments investigated whether perception of function words (e.g., *or*, *are*) is rate-dependent in casual speech, which is often highly reduced. In Experiment 1, talkers spoke sentences containing a target function word; slowing speech rate around this word caused listeners to perceive sentences as lacking the word (e.g., *leisure or time* perceived as *leisure time*). In Experiment 2, talkers spoke matched sentences lacking the word; speeding speech rate around the analogous region caused listeners to perceive a function word that was never spoken (e.g., *leisure time* perceived as *leisure or time*). The results suggest that listeners formed expectancies based on speech rate which influenced the number of words and word boundaries perceived. Findings may help explain the robustness of speech recognition under conditions of a distorted signal, e.g., due to a casual speaking style.

The perception of spoken words is thought to depend largely on recovery of phonemic cues from frequency-specific (spectral) information (e.g., Marslen-Wilson & Welsh, 1978). Yet, recognition of spoken words can be remarkably accurate when spectral cues are missing or severely distorted, i.e., in sinewave speech (Remez, Rubin, Pisoni, & Carrell, 1981), phase-vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) or auditory chimeras (Smith, Delgutte, & Oxenham, 2002), suggesting that temporal information is crucial for accurate spoken word recognition. Despite compelling demonstrations, progress has been slow in understanding the role of timing (cf. speech rate, duration, hierarchical rhythmic structure, etc.) in word recognition (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). Here we show using natural, undegraded speech that entire words can disappear or appear perceptually as a function of the speaking rate of their context. The results have implications for understanding how spoken words are recognized and segmented from speech by adults and infants, issues which are longstanding puzzles in the literature (e.g., Klatt, 1980).

We hypothesized that speech timing plays a decisive role in perceiving a word when its spectrum shows substantial overlap or blending with adjacent words, a pervasive phenomenon known as *coarticulation*. Coarticulation of adjacent words can sometimes be so severe that spectral information is insufficient to identify whether a given word is present in the speech stream, let alone where it begins. This is especially true for short, high-frequency words, e.g., function words like *or* and *and* (Bell et al., 2003; Shockey, 2003).

We reasoned that when coarticulation of words is severe, the presence of a word could be conveyed by the duration of the blended phonemes relative to context speech rate. A typical case of heavy coarticulation is shown in Figure 1. The spectrum for the word *or*, spoken in its reduced form as *er*, blends almost totally with that of the preceding syllable -*sure* in the phrase *leisure or*

time. Thus, there is a relatively homogeneous span of spectral material for most of the two syllables: -(*s*)*ure or*. We hypothesized that such a span is heard as two syllables because it is too long relative to the context speech rate to contain a single syllable. On this view, slowing down the context speech rate should make the span sound relatively shorter, like a single syllable, causing the function word to disappear and the phrase to be heard as *leisure time* instead of *leisure or time*.

One reason for thinking that speech rate might alter perception of a word's presence is that context speech rate affects the boundary between spectrally-related phonemes, e.g., /p/ vs. /b/ (e.g., Liberman, Delattre, Gerstman, & Cooper, 1956; Miller & Liberman, 1979) and singleton and geminate segments (Fujisaki, Nakamura, & Imoto, 1975; Pickett & Decker, 1960). We hypothesized that context speech rate could also affect the perceived presence of larger morphophonological units (i.e., words or syllables). This possibility stems from proposals concerning entrainment to temporal sequences, according to which an auditory event (e.g., a tone or syllable) of a given duration can be heard as corresponding to different rhythms (i.e., different numbers of "beats" or onsets), depending on the rate or rhythm of surrounding events (e.g., Large & Jones, 1999; McAuley, 1995; Port, 2003; Povel & Essens, 1985; Saltzman & Byrd, 2000). Rate normalization has been proposed as one mechanism behind speech rate effects on phoneme boundaries (Miller & Liberman, 1979; Pisoni, Carrell, & Gans, 1983; Sawusch & Newman, 2000). Generalizing this account based on entrainment, we hypothesized that listeners entrain to context speech rate, which thereby affects the number of morphophonological units (words, syllables, segments) perceived in a given stretch of speech. According to this view, the lexical content (number of words) in a spectrally ambiguous stretch of speech depends on its duration relative to speech rate, as well as other information, such as grammatical context. When

a coarticulated stretch of speech is long relative to its surroundings, the listener should perceive a function word, because doing so is plausible based on rate cues as well as higher-level information (e.g., semantic and syntactic context).

In two experiments, we tested whether the number of morphophonological units – here, the number of function words – is dependent on the duration of a given stretch of speech relative to context speech rate, given grammatically viable contexts. For Experiment 1 we predicted that if context events are made slow relative to a stretch of speech containing a function word – either by slowing down the context speech rate or speeding up the stretch of speech itself – then that stretch should be perceived as short and as containing fewer phonological units (i.e., fewer words). For Experiment 2 we predicted that if context events are made fast relative to a stretch of speech not containing a function word – either by speeding up context speech rate or slowing down the stretch of speech itself – then the stretch of speech would be perceived as relatively long and thus as containing an additional phonological unit (i.e., a function word that was never spoken).

Experiment 1

Method

Participants

Participants (n = 41) were young, American English talkers from the Midwest US with self-reported normal hearing.

Materials

Fifty sentences were constructed containing a critical function word (see Appendix) embedded in a phonetic context expected to show heavy coarticulation with the function word. Each sentence had a "grammatically acceptable beginning" whether or not the critical function

word was present, i.e., the span from the beginning of the sentence until just after the critical function word was grammatical, even if the function word was not present. For example, *Deena doesn't have any leisure or time…* is a grammatically acceptable beginning for a sentence, even if the word *or* is missing.

Recordings of experimental sentences were elicited from 29 speakers of American English from the Midwest US. All but the last word of a sentence was presented on a computer screen in front of the participant. The last word was presented only after the sentence had been erased for 1.5 seconds, at which point the participant had to speak the sentence into a head-mounted microphone. Instructions stressed accuracy in repeating the sentence verbatim, which conveyed the experiment was investigating memory. Because no mention was made of speech clarity, participants spoke naturally after adjusting to the task. An additional 70 sentences served as fillers intended to increase sentence variety (e.g., length and structure). Presentation software (Version 12.1, www.neurobs.com) controlled visual sentence presentation and audio recording.

We identified twelve speakers who produced the fewest speech errors and disfluencies, as well as the fewest glottal onsets in critical function words, since continuous formant transitions across this function word were desired; multiple talkers were used to increased the generalizability of the results across speakers. A single token was selected for each item (i) in which *are, or, our,* and *her* were spoken as [ɚ] and *a* was spoken as [ə], (ii) which showed continuous formant transitions across the critical function word plus preceding syllable rhyme, and (iii) which contained no hesitations or disfluencies.[1] Sentences were then divided into target and context portions using spectrogram and waveform displays. The target corresponded to the critical function word plus the preceding syllable and following phoneme (e.g., -*sure or t*-). The

context corresponded to all speech preceding and following the target (e.g., *Deena doesn't have any lei-… -ime*)

Target and context regions were spliced out at zero crossings and subjected to time-manipulation using the PSOLA algorithm implemented in Praat software (Boersma & Weenink, 2002), and then recombined to create four conditions (Figure 2); this method kept intact the spectral detail of the speech while altering only speech rate. In the normal rate condition, the entire fragment was presented at the spoken rate. In the slowed context condition, the context was slowed through time-expansion while the target was presented at the spoken rate (and was acoustically identical to the target in the normal rate condition). In the speeded target condition, the target was speeded through time-compression while the context was presented at the spoken rate. Finally, in the speeded target+context condition, both target and context portions were speeded through time-compression to the same degree. The time-compression and time-expansion factors were 0.6 and 1.9, respectively. Each filler item was likewise speeded, slowed, or left unaltered in rate, with approximately equal numbers at each rate. All stimuli were then amplitude-normalized to 70 dB SPL.

Design and Procedure

The experiment consisted of 120 trials. Fifty were experimental fragments of interest and the remainder were fillers. Each participant heard each experimental fragment only once, in one of the four rate conditions. Each list contained 12 items in each of three rate conditions and 14 items in the fourth, with the pairing of items with conditions counterbalanced across four lists. Each participant was randomly assigned to one of the lists, with approximately an equal number of participants in each list.

The experiment began with 20 filler trials. The remaining trials were presented in a single random ordering. Participants were instructed to listen carefully to each sentence and to play it back as often as necessary to produce a veridical transcription of what was heard, typing the response using a computer keyboard. Stimuli were presented over studio-quality headphones at a comfortable listening level.

Results and Discussion

The frequency of transcribing a function word in the target region was scored. Responses which did not minimally include a transcription of the target region plus the following syllable were discarded (6% of trials).[2] For remaining trials, function word presence vs. absence was coded as 1 or 0, respectively.

Figure 3a shows that reports of the critical function word depended on the relative rate of the target and context. In the normal-rate condition, function word reports were quite high; the fact that they were not at ceiling is expected given that the speech was casually spoken. Critically, a comparison of reports in the normal-rate and slowed context conditions shows that merely slowing the context surrounding a function word caused the rate of function word reports to drop by more than half, from 79% to 33%, even though the target regions containing the function word were acoustically identical. An equally dramatic reduction in function word reports, relative to the normal rate condition, is found in the speeded target condition when the target region is instead speeded and the context is unaltered. Replicating the basic effect, when both the target and context are speeded (i.e., the speeded target+context condition), function word reports rebound to close to their original levels (cf. normal rate condition). That this mean did not reach the same level as the normal-rate condition is likely due to an overall drop in recognition accuracy associated with the significant compression factor (the compressed

fragment was 60% of its original duration). A repeated measures one-way ANOVA by-subjects (F_1) and by-items (F_2) was significant, $F_1(3, 120) = 48.34$, $p < 0.001$, $\eta^2 = 0.55$, $F_2(3, 147) = 54.99$, $p < 0.001$. $\eta^2 = 0.53$. Post-hoc two-tailed paired samples *t*-tests with Bonferroni correction showed that all conditions differed significantly from one another in both by-subjects and by-items analyses ($p < 0.01$) except for the speeded target vs. slowed context conditions.

These results support the predictions of the generalized rate normalization account: making the duration of a stretch of speech containing a function word slow relative to its context affected the number of morphophonological units perceived. This perceptual change was accomplished by either slowing down the context speech rate or speeding up the stretch of speech itself. That listeners could be induced to hear fewer morphophonological units entails that manipulating context speech rate also could induce listeners to hear fewer phonemes and fewer word boundaries than were actually spoken, a finding that has implications for word segmentation.

Experiment 2

A further test of the generalized rate normalization account is whether listeners can be made to hear *more* morphophonological units than were actually produced. This possibility was tested using fragments like those in Experiment 1 except for one minor (but crucial) change: the critical function word was never spoken. We predicted based on the generalized rate normalization account that if context events are made fast relative to a stretch of speech that does not contain a function word – either by speeding up context speech rate or slowing down the stretch of speech itself – then the stretch of speech would be perceived as relatively long and as containing more morphophonological units, even though those words were never spoken.

Participants

Characteristics of participants (n = 69) were identical to Experiment 1.

Materials

Sentences were constructed which had the same grammatical beginnings as in Experiment 1, but lacked the critical function word, e.g., *Deena doesn't have any leisure time*… compared with *Deena doesn't have any leisure or time* in Experiment 1. Fillers were the same as in Experiment 1.

Recordings of experimental and filler sentences were obtained from 23 speakers using the elicitation task described in Experiment 1. From these, we identified fifteen speakers producing the fewest speech errors and disfluencies, and a single token was selected for each item.[3] Each fragment corresponding to the grammatically acceptable beginning was spliced out of its context and the rest of the sentence discarded; the fragment was then divided into target and context portions. The target region was bounded by the same phoneme string as in Experiment 1, only the function word was not present (e.g., *-sure t-* in *Deena doesn't have any leisure time…*). The context portion corresponded to all speech material preceding and following the target.

Four speech rate conditions were created from each fragment. In the normal rate condition, the entire fragment occurred at the spoken rate. In the speeded context condition, the context was speeded while the target was at the normal rate. In the slowed target condition, the target was slowed while the context was at the normal rate. Finally, in the slowed target+context condition, both target and context portions were slowed to the same degree. The time-compression and time-expansion factors were 0.6 and 1.9, respectively. After alteration, portions were concatenated in the proper order, and stimuli were amplitude-normalized to 70 dB SPL. The design and procedure were identical to Experiment 1.

Results and Discussion

The frequency of transcribing a function word in the target region was scored. Responses which did not minimally include a transcription of the target region plus the following syllable were discarded from analysis (7% of trials).[4]

Clear evidence was found that listeners can be induced to hear a function word by altering only speech rate (Figure 3b). In the normal (baseline) rate condition, participants seldom (3% of the time) reported a function word in the target region - an expected finding since critical function words were never spoken. However, speeding the context surrounding the target caused an eight-fold increase in the rate of reporting a function word, even though the target was identical in the two conditions. Slowing the target similarly increased five-fold the rate of hearing a function word that was never spoken. As in Experiment 1, when the context and target were time-altered together (in this case slowed), reports of the function word returned to the level found in the normal-rate condition. A repeated measures one-way ANOVA with rate condition as the factor was significant, $F_1(3, 204) = 60.82$, $p < 0.001$, $\eta^2 = 0.47$, $F_2(3, 147) = 25.50$, $p < 0.001$, $\eta^2 = 0.34$. Post-hoc tests showed all conditions differed significantly from one another in both by-subjects and by-items analyses ($p < 0.01$) except for normal rate vs. slowed target+context conditions.

These results provide even stronger evidence for the generalized rate normalization hypothesis that context speech rate affects whether listeners perceive a word. Here, a function word was made to appear perceptually, even though it was never spoken. By implication, context speech rate affected the number of phonemes and word boundaries perceived by listeners, thus replicating and extending the findings of Experiment 1.

General Discussion

The current studies provide new insight into how timing information is used in speech perception. In Experiment 1, sentence fragments containing a critical function word were heard as having fewer such words when context speech rate was slowed. In Experiment 2, matched sentences in which the critical function word was never spoken were heard as containing function words when context speech rate was speeded. These experiments were based on a generalized rate normalization hypothesis, according to which the number of perceived morphophonological units depends on the duration of a stretch of speech relative to context speech rate. These experiments indicate that listeners used context speech rate to help decode spectrally ambiguous portions of the signal, thereby aiding in perceiving spoken words and segmenting them from the speech stream.[5]

These studies are the first to show that context speech rate can modulate whether an entire word is perceived. The duration of a stretch of speech relative to context speech rate also modulated the number of phonemes and implied word onsets perceived as present. These findings have implications for how infants and adults identify word onsets in connected speech, an important and unsolved problem (Cutler, Mehler, Norris, & Segui, 1983; Mattys, White, & Melhorn, 2005; Thiessen, Hill, & Saffran, 2005). Note that our rate manipulations were several phonemes distant from the variably-perceived function word; this contrasts with previous work in which rate manipulations were imposed immediately adjacent to a to-be-perceived phoneme, with little evidence of more distant manipulations having effects (e.g., Sawusch & Newman, 2000).

These findings suggest that relative speech rate information aids in interpreting ambiguous spectral cues, helping listeners to identify and segment spoken words. How words and word boundaries are so robustly perceived when spectral cues are unclear remains poorly

understood (Ernestus, Baayen, & Schreuder, 2002; Pitt, 2009). Our experiments suggest that word recognition depends in part on relative rate cues provided by speech context, adding to a growing body of work showing prosodic properties of speech context influence lexical recognition and word segmentation (Dilley & McAuley, 2008; Gout, Christophe, & Morgan, 2004; Salverda et al., 2003).

More generally, the results demonstrate the rapid and seamless integration of signal-based (spectral-temporal) and knowledge-based (syntax, semantics) cues during spoken word recognition. In this regard, our speech rate phenomenon, particularly the results of Experiment 2, can be viewed as a temporal version of phonemic restoration, in which listeners readily restore phonemes in words whose acoustic evidence was replaced by noise (Samuel, 2001). In phonemic restoration, sentential context biases perception, and such higher-level biases are likely at work here; they may be a pre-condition for the effect, for example.

Compared with spectral cues, studying how timing information is used in speech perception has proven challenging. The present results provide one answer to the puzzle of how reduced and/or spectrally attenuated speech is recognized and segmented from the signal, with rate normalization via temporal entrainment as a possible explanation. In the absence of clear spectral information, timing information becomes increasingly important in conveying the message intended by the talker.

Appendix.

Grammatically acceptable beginnings of sentences constituting experimental stimuli are shown. Parentheses indicate the word was present in the Experiment 1 fragment, but not the Experiment 2 fragment, while square brackets indicate the reverse.

Taylor knew the principal and teacher (are) from Ohio

Conor knew that bread and butter (are) both

Frank thinks that sadness and anger (are) both

Claire said that sour and bitter (are) both

Chris said his mother and father (are) both

Zach knew that there (are) things

George thought my father and brother (are) like [good]

Glenn thought his friend and neighbor (are) like plenty

Ruth saw the maid and butler (are) at the top

Rose knew that there (are) lamps

The company moved to (a) different

Trent might get to (a) certain

Clay thinks that would be (a) good

The Smiths wouldn't buy (a) Butterball

Anne wanted to see (a) very funny

It makes no sense to obey (a) petty

It takes a lot of work to review (a) personal

It costs a lot to tattoo (a) pink

The boy wanted to glue (a) broken

Dave asked how long it takes to repay (a) large

Aspirin and other painkillers are (our) drugs

The Murrays are (our) favorite

The callers are (our) French contacts

Mom said these are (our) gray gloves

The accountants are (our) wise advisors

Phil and Mary are (our) young cousins

The leaves fell after (her) green

The manager hid the candy before (her) six kids

The sign was replaced after (her) black

The message was clear after (her) blank

Chris was very quick after (her) sharp

The Perrys thought carefully after (her) wise advice

The value went up after (her) rich neighbors

People were offended after (her) rude

The Smiths were shocked after (her) weird

Deena doesn't have any leisure (or) time

Anyone must be a minor (or) child

Marty gave him a dollar (or) twenty last week

George turned left at the river (or) bank

Sally sold all her silver (or) jewelry last month

Don must see the harbor (or) boats

Fred would rather have a summer (or) lake

Steve pitched the ball to center (or) left

They promised him the future (or) aid

Susan said those are (our) black socks

Jake didn't vote for the member (or) constituent

Jack reported trouble before (her) two children

These documents are (our) fake

Those tickets are (our) late entries

These houses are (our) best

Footnotes

[1]To select similar speaking rates across talkers, we first determined the grand mean duration of the two syllables preceding the critical function word across all items for these speakers. After other selection criteria had been applied, a token of a given item was selected from the talker who produced the two syllables preceding the function word with a duration which was minimally different from the grand mean duration. This resulted in varied numbers of experimental items from each talker (mean: 4.2; range: 0-13, with 11 talkers represented in the final experimental stimulus set).

[2] Results are identical if these responses are included.

[3]The selection procedure for controlling speaking rate across items was identical to Experiment 1. This resulted in varied numbers of experimental items from each talker (mean: 3.3; range: 0-8, with 13 talkers represented in the final set).

[4] Results are identical if these responses are included.

[5] A competing hypothesis that speech rate mismatches *per se* were responsible for effects on function word perception is untenable, for several reasons. According to one version of this hypothesis, function words failed to be perceived when the rate across stimuli mismatched. However, in Experiment 2, more, not fewer, function words were perceived when the rate mismatched than when it did not, indicating this hypothesis cannot account for data across both experiments. A weaker version of the hypothesis also is not supported, namely, that speech rate mismatches cause reductions in general intelligibility associated with different illusory lexical percepts, depending on the veridical grammatical properties of fragments. To test this, we conducted additional analyses of transcription accuracy of phonemes in words preceding the critical function word in matching (Experiment 1: normal rate, speeded target+context;

Experiment 2: normal rate, slowed target+context) and mismatching (Experiment 1: slowed context, speeded target; Experiment 2: speeded context, slowed target) conditions. Results revealed no difference in transcription accuracy in Experiment 1 ($\overline{X}_{match}$ = 94%; $\overline{X}_{mismatch}$ = 95%; paired-samples $t(49) = 1.13$, $p = 0.27$). Although a significant difference was found in Experiment 2 ($\overline{X}_{match}$ = 97%; $\overline{X}_{mismatch}$ = 93%; paired-samples $t(49) = 6.21$, $p < 0.001$), the size of the change (4%) is much smaller than the rise in function word report rates (13-21%), further suggesting that such a rate effect cannot account for differences in function word perception.

References

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of Acoustical Society of America, 113*(2), 1001-1024.

Boersma, P., & Weenink, D. (2002). Praat, a system for doing phonetics by computer (Version 4.0.26): Software and manual available online at http://www.praat.org.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language-specific comprehension strategy. *Nature, 304*, 159-160.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218-244.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*(3), 294-311.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162-173.

Fujisaki, H., Nakamura, K., & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In G. F. M. A. A. Tatham (Ed.), *Auditory analysis and perception of speech*. London: Academic Press.

Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language, 51*, 547-567.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119-159.

Liberman, A. M., Delattre, P., Gerstman, L., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52*(2), 127-137.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognition, 10*, 487-509.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

McAuley, J. D. (1995). *Perception of time as phase: toward an adaptive-oscillator model of rhythmic pattern processing.* Unpublished Ph.D. Dissertation, Indiana University.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occuring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(6), 457-465.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language & Speech, 3*, 11-17.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception and Psychophysics, 34*(4), 314-322.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61*, 19-36.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics, 31*, 599-611.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*(4), 411-440.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-949.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science, 19*, 499-526.

Salverda, A. P., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51-89.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II. Effects of signal discontinuities. *Perception and Psychophysics, 62*(2), 285-300.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Shockey, L. (2003). *Sound Patterns of Spoken English*. Cambridge: Blackwell.

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

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1), 53-71.

Acknowledgments

We thank Delphine Dahan, Sven Mattys, Arthur Samuel and an anonymous reviewer for useful feedback on the manuscript. Also, we thank Victoria Hoover, Michael Tat, Andrea Hulme, Chris Heffner, and Claire Carpenter for help with data acquisition and analysis. This work was supported by grants NIH-NIDCD DC004330 to MAP and NSF BCS-0847653 to LCD.

Figure Legends

Figure 1. **Spectrogram illustrating heavy coarticulation of a function word in the phrase *leisure or time***. Phonemic content is shown as time-aligned International Phonetic Alphabet (IPA) symbols at the top of the figure. The arrow on the x-axis indicates the approximate start of the function word *or*; note the utter absence of discontinuity marking the start of this word and lack of clear cues differentiating the function word spectrally from the preceding syllable.

Figure 2. **Waveforms of one time-altered stimulus across the four conditions of Experiment 1**. The sections of the waveform shown in light grey boxes correspond to context, while the remainder of the waveform is the target region.

Figure 3. **Effects of changes in speech rate on hearing function words**. **a)** Mean percent (± s.e.m) reports of function words across the four conditions of Experiment 1, in which the function word was spoken. **b)** Mean percent (± s.e.m) reports of function words across the four conditions of Experiment 2, in which the function word was not spoken.

Figure 1.

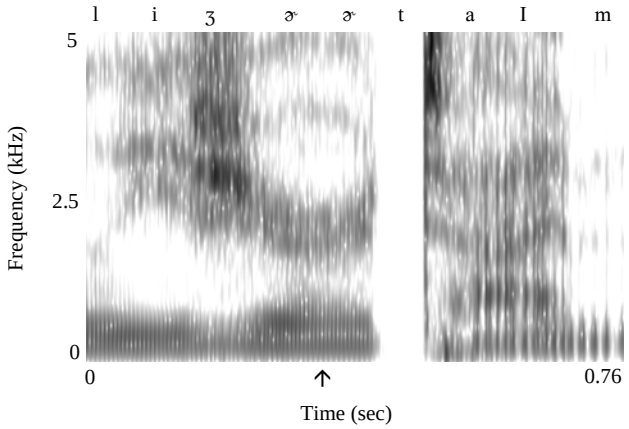
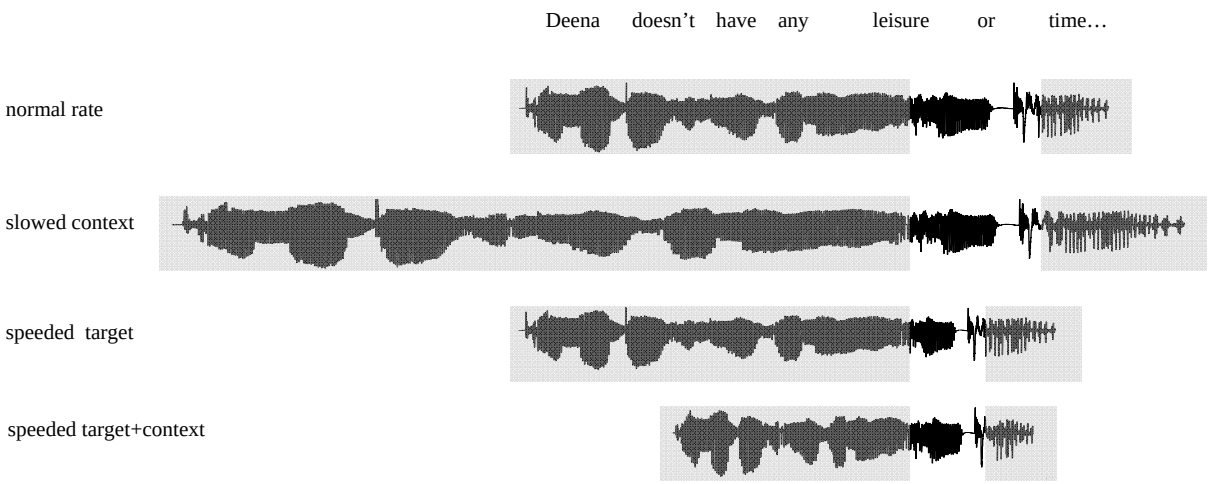
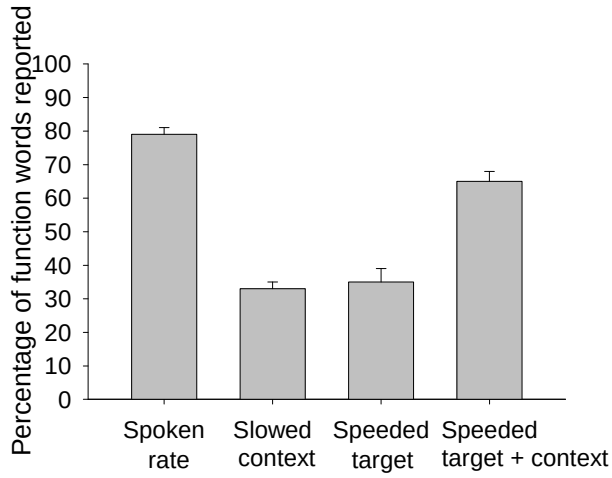


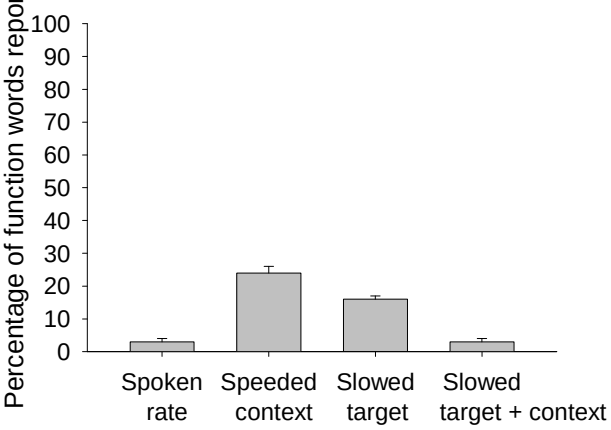
Figure 2.


Deena doesn't have any leisure or time…
normal rate
slowed context
speeded target
speeded target+context

a


Percentage of function words reported
100
90
80
70
60
50
40
30
20
10
0
Spoken rate
Slowed context
Speeded target
Speeded target + context

b


Percentage of function words reported
100
90
80
70
60
50
40
30
20
10
0
Spoken rate
Speeded context
Slowed target
Slowed target + context