Cognition 131 (2014) 69–74

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

# Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis

Tuuli H. Morrill [a,b,*], Laura C. Dilley [a,b], J. Devin McAuley [b], Mark A. Pitt [c]

[a] Department of Communicative Sciences and Disorders, Oyer Center, 1026 Red Cedar Road, East Lansing, MI 48824, United States
[b] Department of Psychology, Michigan State University, 316 Physics Road, East Lansing, MI 48824, United States
[c] Department of Psychology, The Ohio State University, 201 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210, United States

A R T I C L E  I N F O

*Article history:*
Received 24 April 2013
Revised 5 November 2013
Accepted 15 December 2013

*Keywords:*
Rhythm
Speech rate
Word segmentation
Word recognition
Perceptual grouping

A B S T R A C T

Due to extensive variability in the phonetic realizations of words, there may be few or no proximal spectro-temporal cues that identify a word's onset or even its presence. Dilley and Pitt (2010) showed that the rate of context speech, distal from a to-be-recognized word, can have a sizeable effect on whether or not a word is perceived. This investigation considered whether there is a distinct role for distal rhythm in the disappearing word effect. Listeners heard sentences that had a grammatical interpretation with or without a critical function word (FW) and transcribed what they heard (e.g., *are* in *Jill got quite mad when she heard* ***there are*** *birds* can be removed and *Jill got quite mad when she heard* ***their*** *birds* is still grammatical). Consistent with a perceptual grouping hypothesis, participants were more likely to report critical FWs when distal rhythm (repeating ternary or binary pitch patterns) matched the rhythm in the FW-containing region than when it did not. Notably, effects of distal rhythm and distal rate were additive. Results demonstrate a novel effect of distal rhythm on the amount of lexical material listeners hear, highlighting the importance of distal timing information and providing new constraints for models of spoken word recognition.



## 1. Introduction

Locating and identifying words in continuous speech is made difficult by substantial acoustic variability and a lack of consistent phonetic cues to word boundaries (e.g., Cole & Jakimik, 1980; Klatt, 1989). Variation in the realizations of words themselves due to coarticulation, and/or a casual speech style, can cause the spectral cues to phonemes within a word to blend with those of preceding phonemes across a word boundary; as a result, there may be few or no proximal spectrotemporal cues that clearly identify a word's onset or even its presence (Bell et al., 2003; Ernestus, Baayen, & Schreuder, 2002; Johnson, 2004). How do listeners manage to locate word boundaries and recognize words in continuous speech?

Prior research has established that listeners integrate multiple sources of information to segment and recognize words. This information includes semantic and/or syntactic context (Mattys & Melhorn, 2007), as well as acoustic and phonetic cues adjacent (i.e., proximal) to word boundaries. Among the proximal cues affecting word segmentation and recognition are (sub-)segmental and/or allophonic variation (e.g., Andruski et al., 1994; Byrd, Krivokapic, & Sungbok, 2006; Keating, Cho, Fougeron, & Hsu, 2003), phonotactic cues (e.g., McQueen, 1998; Vitevitch & Luce, 1999), and word-level prosodic information, such as lexical stress (e.g., Soto-Faraco, Sebastian-Galles, & Cutler, 2001; Van

* Corresponding author. Address: Michigan State University, Department of Communicative Sciences and Disorders, Oyer Center B9, 1026 Red Cedar Road, East Lansing, MI 48824, United States. Tel.: +1 517 432 7042.
*E-mail address:* tmorrill@msu.edu (T.H. Morrill).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.12.006

Donselaar, Koster, & Cutler, 2005), as well as durational and phrasal boundary information (Christophe, Gout, Peperkamp, & Morgan, 2003; Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003).

This investigation focuses on prosodic cues for spoken word recognition that are temporally removed (i.e., distal) from the to-be-recognized material, a topic that has only recently begun to receive attention (e.g., Dilley, Mattys, & Vinke, 2010; Dilley & McAuley, 2008; Dilley & Pitt, 2010; Reinisch, Jesse, & McQueen, 2011). The current research was motivated by two lines of work on distal prosody. The first line has revealed that the rate of distal speech can have a profound effect on whether or not listeners hear coarticulated function words (FWs) (Dilley & Pitt, 2010; Heffner, Dilley, McAuley, & Pitt, 2012). In Dilley and Pitt (2010), listeners heard phrases such as *Don must see the harbor or boats* in which the FW *or* was reduced, with spectrotemporal blending across the FW-containing region. When distal speech rate was slowed relative to the FW-containing region, listeners were substantially less likely to report the FW than when the context and FW-containing region were presented at the same rate. This finding of a *disappearing word effect* shows distal speech rate can induce listeners to perceive – or not perceive – a FW within a region of speech containing no obvious acoustic landmarks; c.f. Stevens (2002).

At this point, little is known about the nature of the disappearing word effect, which would not be readily accommodated by theories of spoken word recognition (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris & McQueen, 2008). As such, a thorough understanding of the effect would help guide development of a more expansive model. Here we consider the possibility that the underlying mechanism responsible for the disappearing word effect is sensitive to other temporal information, in addition to rate, and might be elicited by expectations generated by repeating (i.e., rhythmic) pitch patterns in the distal context.

This idea was motivated by our second line of work investigating distal prosodic effects in word segmentation. Dilley and colleagues (Dilley et al., 2010; Dilley & McAuley, 2008) have shown distal pitch patterns can influence how sequences of syllables with clear acoustic onsets are parsed into words. Consistent with a perceptual grouping account, binary pitch patterns (high-low vs. low–high) at the beginning of an utterance influenced the perceived organization of later-occurring syllables; depending on the initial pattern, the final syllables of the sequence (e.g., [krai.sis.tɛr.-nip]) were parsed differently (e.g., *crisis turnip* vs. *cry sister nip*).

To examine whether perceptual grouping can also affect the *amount* of lexical material listeners hear, we conducted an experiment that followed the methods of Dilley and Pitt (2010), but varied distal rhythm in addition to rate. Listeners heard naturally-produced sentences containing a critical FW that was not required for a grammatical interpretation of the sentence and transcribed what they heard. For example, if the word *are* in *Jill got quite mad when she heard* ***there are*** *birds* is not perceived, the listener could hear the grammatical sentence *Jill got quite mad when she heard* ***their*** birds. Distal rhythm consisted of distinct repeating pitch patterns (see Fig. 1). Binary rhythms consisted of a repeating disyllabic High–Low (H–L) pattern, while ternary rhythms consisted of a repeating trisyllabic H–L–L pattern. The final four syllables of each item always had a H–L–L–H pattern so that the FW and preceding syllable had a sustained low pitch (–L–L). Neutral rhythms, consisting of a monotone pitch pattern across the utterance, were also examined to provide a baseline measure of FW reports in the absence of rhythmic patterning.

Based on predictions from the perceptual grouping hypothesis, different pitch patterns in the portion of the sentence preceding the critical FW should alter expectations about the perceptual organization of the FW-containing region, including the perceived number of syllables. Consider that for the ternary rhythmic context, the repeating H–L–L pitch pattern matches the H–L–L pitch pattern in the FW-containing region; thus, if the ternary context influences the expected grouping of syllables in the FW-containing region, listeners are predicted to hear three distinct syllables, including the optional FW. Conversely, for the binary rhythmic context, the repeating binary (H–L) pattern does not match the identical H–L–L pitch pattern in the FW-containing region; thus, if the binary context influences the expected grouping of syllables, listeners are predicted to hear only two distinct syllables in the FW-containing region and be less likely to hear the critical FW.

We independently manipulated distal rate by slowing the context speech surrounding the critical FW for three distal rate levels (an unaltered rate, a slowed rate, and a slowest rate). As in Dilley and Pitt (2010), slower distal rates should reduce FW reports. Based on the non-speech auditory perception literature, the perceptual grouping of elements based on pitch patterns should also be stable across the range of rates typical of natural speech and those used here (Handel, 1989; Smith & Cuddy, 1989; Woodrow, 1909). Thus, according to a perceptual grouping hypothesis, we would expect the hypothesized effect of distal rhythm not to interact with the effect of distal rate (i.e., the effects of rhythm and rate should be additive).

## 2. Experiment

### 2.1. Participants and design

Seventy-three native speakers of American English with self-reported normal hearing from the Michigan State University community (Age: $M = 21.0$, $SD = 4.8$) participated for course credit or nominal financial compensation. Participants were randomly assigned to a binary rhythm ($n = 25$), ternary rhythm ($n = 24$), or neutral rhythm condition ($n = 24$), and heard all three distal rates: unaltered, slowed (by a factor of 1.4), and slowest (slowed by a factor of 1.8).

### 2.2. Materials

Stimuli consisted of 32 test and 32 filler sentences. Each test sentence consisted of 10 monosyllabic words. The penultimate syllable was always a FW (*are, or, our,* or *a*). Stimuli were recorded by a female native speaker of American English, speaking in a "casual" style at a conversational

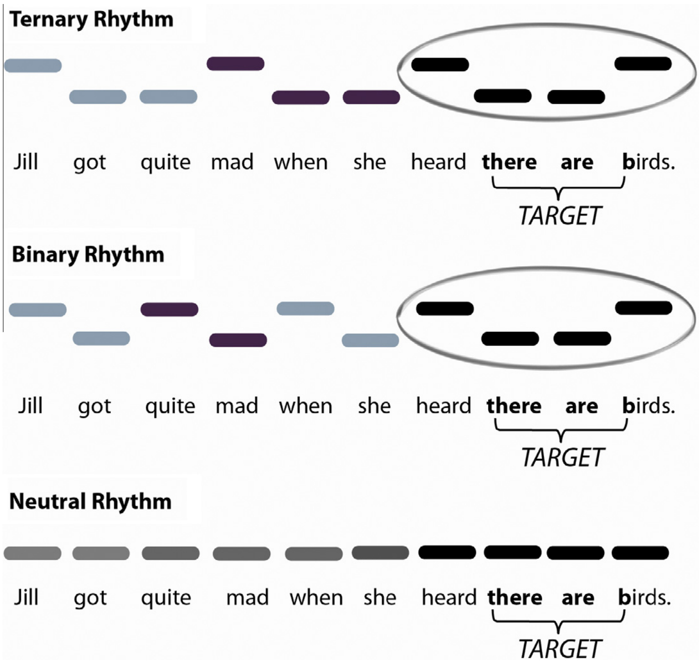


**Fig. 1.** Schematic of the pitch pattern in each Rhythm condition. The "target region" consisted of the FW, its preceding syllable, and the onset of the following syllable; the critical FW and preceding syllable were coarticulated (e.g., *there are* realized as [ðɛrː]). The preceding and following context was expanded for the Rate manipulations. The Ternary rhythm condition consisted of a H–L–L–H–L–L–H pattern in the distal context, while the Binary rhythm consisted of a H–L–H–L–H–L–H pattern in the distal context. The final four syllables always had a H–L–L–H pitch pattern (circled) so that the function word and the preceding syllable had a sustained low pitch (–L–L). The entire target region (e.g., *there are b-*) was acoustically identical across the Binary and Ternary conditions.

rate, while maintaining a monotonous F0. Fillers were 10-syllables long, but did not necessarily include a FW on the penultimate syllable (e.g., *I don't like juice, but I think tea tastes worse*), and underwent the same rhythm and rate manipulations as test stimuli.

*2.2.1. Rhythm manipulation*

F0 patterns were imposed over the distal context of each sentence (first seven syllables) by assigning a High (H) or Low (L) F0 to each syllable, creating repeating ternary (H–L–L) or binary (H–L) patterns (Fig. 1). The final four syllables always had a H(–L–L)–H pattern; the FW word and preceding syllable were realized with a sustained pitch (L–L). The pitch pattern across the FW-containing region was identical across conditions. The H and L F0 values were determined by adding 10 Hz to the median F0 of each sentence (H), and subtracting 10 Hz from the median (L) (prior to resynthesis, median F0 across stimuli was ~200 Hz: $M = 197.13$ Hz, $SD = 2.63$). F0 values were assigned using PSOLA resynthesis in Praat (Boersma & Weenink, 2002). The naturally produced articulation patterns and relatively small pitch range resulted in the resynthesized stimuli being readily interpreted as speech. In a third, neutral (monotone) rhythm condition, the pitch of the entire sentence was resynthesized to the median F0.

*2.2.2. Rate manipulation*

In one of the Rate conditions, the entire item was unaltered. In the two other conditions, the rate of the context speech was slowed by expansion factors of 1.4 and 1.8, while a target region was unaltered (Table 1). For each item, the target region, acoustically identical across conditions, consisted of the syllable preceding the FW, the FW itself, and the onset of the syllable following the FW.[1] For example, in *Jill got quite mad when she heard* *__there are b__irds*, [ðɛɹ#ɚ#b] remained unaltered, while the preceding and following context was slowed in the 1.4 and 1.8 conditions. Time expansion was performed using PSOLA resynthesis in Praat (Boersma & Weenink, 2002). For fillers in the slowed conditions, the entire sentence was slowed, as there was no target region. All soundfiles were normalized to 70 dB.

*2.3. Procedure*

Each spoken stimulus was presented with a display of the first 6 syllables (i.e., words); participants were instructed to type the remaining words. Responses were expected to be three or four syllables, depending on whether participants perceived the critical FW. Half of the fillers were presented with a display of the first 6 syllables, and half with 7 syllables so that responses would be three or four syllables long for fillers as well. There were 2 practice trials; stimuli and fillers were presented in random order (64 trials).

[1] If the onset contained a consonant cluster, the first consonant was included; vowel onsets were not included.

**Table 1**
Durations and speech rate in syllables/second for stimuli.

| Rate condition | Syllable duration in context region | Syllables/second in context region | Duration of target region in all conditions |
|---|---|---|---|
| 1.0 | $M$ = 176 ms, $SD$ = 17 ms | $M$ = 5.74, $SD$ = .56 | $M$ = 378 ms, $SD$ = 50 ms |
| 1.4 | $M$ = 246 ms, $SD$ = 24 ms | $M$ = 4.10, $SD$ = .40 | |
| 1.8 | $M$ = 316 ms, $SD$ = 30 ms | $M$ = 3.19, $SD$ = .31 | |

### *2.4. Data analysis*

Typed responses to stimulus items were coded for presence or absence of the FW. Responses containing inaccurate transcriptions of words adjacent to the FW were not coded (3.2% of responses). To examine the reliability of effects of Rhythm and Rate, a logit mixed-effects model analysis (e.g., Jaeger, 2008) was performed in R (Bates, Maechler, & Bolker, 2012). The model that best predicted the likelihood of FW reporting was fit with Rhythm and Rate as fixed effects, and subjects and items as random effects (Table 2).[2]

**Table 2**
Estimates of predictor variables and their reliability in the mixed models analysis. The model includes random slopes for repeated fixed effects – Rate for subjects, and Rhythm and Rate for items.

| | Estimate | Std. Error | $z$ Value | Pr(>\|$z$\|) |
|---|---|---|---|---|
| (Intercept) | 2.2550 | 0.7074 | 3.188 | $p < .005$ |
| Rhythm – Neutral | 0.7414 | 0.3140 | 2.361 | $p < .05$ |
| Rhythm – Ternary | 0.9125 | 0.2926 | 3.118 | $p < .005$ |
| Rate – 1.4 | −1.3945 | 0.2529 | −5.515 | $p < .001$ |
| Rate – 1.8 | −2.8718 | 0.2723 | −10.545 | $p < .001$ |

## 3. Results

Fig. 2 shows mean proportions of critical FWs transcribed for the binary, ternary, and neutral rhythm conditions at each rate. Consistent with previous research, FW reports were lower when the context speech rate was slower. Table 2 presents the model with coefficient estimates, standard errors, Wald's $z$ values, and the significance level for each predictor. Treatment coding with the binary rhythm and unaltered rate (1.0) as the baseline was used to examine the contrast between the binary and ternary and binary and neutral Rhythm conditions. The results show that the slowed 1.4 Rate condition exhibited significantly lower critical FW reports than the unaltered condition ($\beta = -1.39$, $p < .001$), as did the slowed 1.8 condition ($\beta = -2.87$, $p < .001$).

The results also reveal that Rhythm was a predictor of FW reports, with a significant difference between the binary and neutral rhythms, ($\beta = 0.74$, $p < .05$), and the binary and ternary rhythms ($\beta = 0.91$, $p < .005$). Treatment coding with ternary rhythm and unaltered rate (1.0) as the baseline was used to examine the contrast between ternary and neutral rhythms; these were not significantly different. However, as hypothesized, FW reports were lowest for binary rhythm, intermediate for neutral rhythm, and highest for ternary rhythm.

## 4. Discussion

Consistent with a perceptual grouping hypothesis, there were fewer FW reports with the binary rhythmic context than the ternary rhythmic context. That is, when listeners only expected two syllables in the FW-containing region based on an extrapolation of the perceptual grouping induced by the binary rhythm, they were less likely to report the optional (third syllable) function word. Reduced FW reports observed with the binary context were moreover lower than those observed in the neutral baseline condition. In line with previous work, slower distal rates also reduced function word reports, but notably the effects of rhythm and rate did not interact (i.e., they were additive).

Before concluding that speech rhythm independently affects FW reports, an alternative interpretation of the findings should be considered: perhaps the effect of distal rhythm could be the result of an indirect effect of distal rate. This possibility is suggested by work in music perception, showing that melodic pitch variations impact perceived rate (e.g., Boltz, 1998). To examine whether our distal rhythm manipulations might have impacted perceived rate, we conducted a follow up experiment in which 26 listeners rated the speed (from 1 to 6, "very slow" to "very fast") of the stimuli from all rhythm and rate conditions. While sentences with both binary and ternary rhythms were perceived as slightly faster than the neutral (monotone) stimuli, there was no difference in the perceived rates between the binary and ternary conditions. Thus, the effect of distal rhythm on lexical perception is unlikely to be an indirect effect of rate.

These findings contribute to a larger body of work demonstrating the importance of rhythmic cues in linguistic processing. Previous studies have shown that proximal rhythmic information, such as stress, affects word segmentation and lexical recognition (e.g., Mattys, White, & Melhorn, 2005; Soto-Faraco et al., 2001; Van Donselaar et al., 2005), and that phrase-level rhythmic patterns can facilitate processing (Cutler & Darwin, 1981; Martin, 1972; Quene & Port, 2005; Shields, McHugh, & Martin, 1974). However, this is the first demonstration that *distal* rhythmic cues to syllable organization can directly influence whether a word is heard at all.

The effect of rhythmic context on word recognition provides support for the perceptual grouping hypothesis initially proposed by Dilley and McAuley (2008). Previously, Dilley and colleagues (Dilley et al., 2010; Dilley & McAuley, 2008) have shown that distal pitch patterns can influence the parsing of syllable sequences in a manner consistent

[2] A model including an interaction term for Rhythm and Rate showed no interaction, and did not result in significant improvement in model fit ($\chi = 2.37$, $p = .67$).

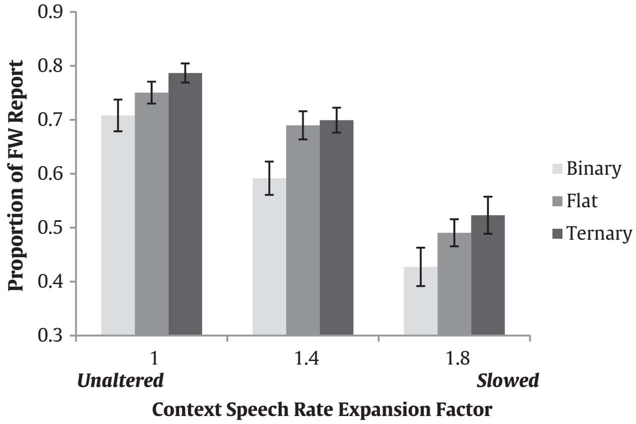


**Fig. 2.** Mean proportions of FW reports in each Rhythm and Rate condition (by subject). Error bars represent standard error of the mean.

with general principles of auditory perceptual organization, but this is the first demonstration that perceptual grouping can influence the amount of lexical material that is heard. Moreover, the additive effects of speech rhythm and rate suggest that perceptual grouping based on speech rhythm is stable within a range of speech rates, consistent with the range of rates previously examined in the grouping of tonal elements (Handel, 1989; Smith & Cuddy, 1989; Woodrow, 1909).

Taken together, the present findings show that models of word recognition must accommodate at least two kinds of distal acoustic information: speech rate and speech rhythm. Models to date have generally focused on the role of proximal acoustic information, with little consideration for the possible role of prosodic information and no mechanism by which acoustic information from the preceding context would affect word recognition (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris & McQueen, 2008).

One potentially promising approach to incorporating distal information in theories of spoken word recognition builds on Dilley and Pitt's (2010) proposal that the distal rate effect emerges as a consequence of entrainment. The rate of speech serves to pace the rate at which listeners process acoustic speech material, thus influencing the amount of material heard (e.g., one vs. two syllables). Peelle and Davis (2012) have recently considered how this perspective might be realized in terms of entrainment of neural oscillations; one possibility is that acoustic markers of syllable onsets (e.g., amplitude rise time) serve as an entraining cue (Ahissar et al., 2001; Giraud & Poeppel, 2012; Goswami, 2011; Kotz & Schwartze, 2010). The current work suggests that distal speech rhythm (e.g., comprised of pitch patterns) may also serve as an entraining cue (notably separate from speech rate) to influence the amount of speech material listeners hear and, more generally, speech understanding.

## 5. Conclusion

This article demonstrates a novel effect of distal rhythm on spoken word recognition that is consistent with a perceptual grouping hypothesis. Distal rhythms consisting of repeating pitch patterns influenced the *amount* of lexical material listeners perceived in the absence of any proximal cues. Moreover, as predicted, effects of distal rhythm and rate were additive. Findings demonstrate that models of spoken word recognition must incorporate at least two types of distal information, rate and rhythm; effects of rate and rhythm combine additively over a range of rates. A framework employing neural oscillations that are coupled across time scales may provide a means for encoding distal temporal information at multiple levels, thus suggesting a direction for future research.

## Acknowledgments

We would like to thank Albert Kim and two anonymous reviewers for their helpful suggestions, which greatly improved the manuscript. We are grateful to Claire Carpenter for help with stimulus creation, and Evamarie Cropseye, Mary Flynn, Elaine Foster, Mina Hirzel, Rose Merrill, Stephanie Schmidt, Kayla Tillman, Elizabeth Wieland, and other members of the MSU Speech Perception-Production Lab for their assistance with this study, and Daniel Ezra Johnson for assistance with statistical analysis. Partial support for this work was provided by NSF Award BCS 0874653 to Laura C. Dilley. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## Appendix A. Experimental items

That new store would sell us more <u>fur or c</u>oats.
Those thieves if they get caught would <u>flee a j</u>ail
We should tell them that the rats <u>are our p</u>ets.
John will try his best to catch <u>me a f</u>ish.
They went to some beach to find <u>there are n</u>ets.
Both cracked wheat and dark rye breads <u>are our b</u>est.
Eat good food and have less stress <u>are our g</u>oals.
Most of these old guys here know <u>there are b</u>eers
Zach plans he'll be here for an <u>hour or m</u>ore.
Mike liked sports camp so he learned <u>there are r</u>ules.
She cleaned off her fridge and saw <u>there are l</u>ists.
Todd asked if she wants to go <u>to a p</u>lay.

That young man played on his drums for our friends.
George was sure he knew that they are our aunts.
My mom will not park the car near our carts.
Some kid will yell when they see there are bugs.
They looked in the shed and found there are bikes.
Jill got quite mad when she heard there are birds.
Those bright green spots in the sketch are our trees.
That wine glass can be filled with more or juice.
Jim thought that street lamp must be far or off.
My dad will now go to the car or show.
Hank told him it's not good to see a moose.
Sue told me what's in the fire are our logs.
On the deck by that brick wall are our chairs.
By the door those small brown bags are our snacks.
Next to that desk and old bed are our shelves.
Those shoes she had on the desk are our mom's.
Leave some cake for those guests who are our friends.
We think the sounds from the woods are our dogs.
Jen will get dressed up and go to a dance.
At the beach near John's house he saw a deer.

## 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 (USA), 98*, 13367–13372.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163–187.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42.

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 the Acoustical Society of America, 113*(2), 1001–1024.

Boersma, P., & Weenink, D. (2002). Praat: Doing phonetics by computer [Computer program]. (Version 4.0.26): Software and manual available online at <http://www.praat.org>.

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics, 60*(8), 1357–1373.

Byrd, D., Krivokapic, J., & Sungbok, L. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America, 120*(3), 1589–1599.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics, 31*, 585–598.

Cole, R. A., & Jakimik, J. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America, 64*, 1323–1332.

Cutler, A., & Darwin, C. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29*, 217–224.

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., Mattys, S., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*, 274–294.

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.

Dilley, L. C., & Pitt, M. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664–1670.

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

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*(1), 3–10.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Heffner, C., Dilley, L., McAuley, J. D., & Pitt, M. (2012). When cues collide: How distal speech rate and proximal acoustic information jointly determine word perception. *Language and Cognitive Processes, iFirst*(1–28).

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*, 434–446.

Johnson, K. (2004). Massive reduction in conversational American English. Paper presented at the Spontaneous speech: Data and analysis. In *Proceedings of the 1st Session of the 10th International Symposium*, Tokyo, Japan.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation* (pp. 143–161). Cambridge: Cambridge University Press.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences, 14*(9), 392–399.

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

Martin, J. G. (1972). Rhythmic (heirarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*(6), 487–509.

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America, 122*(1), 554–567.

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.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*(1), 21–46.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*(320), 1–17.

Quene, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*, 1–13.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 978–996.

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.

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102*(2), 250–255.

Smith, K. C., & Cuddy, L. L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 457.

Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412–432.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111*(4), 1872–1891.

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology, 58A*(2), 251–273.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*(3), 374–408.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14*, 1–66.