Speech rhythm and speech rate affect segmentation of reduced function words in continuous speech

Tuuli MorrillLaura DilleyJ. Devin McAuley and Mark Pitt

Citation: **19**, 060210 (2013); doi: 10.1121/1.4800122
View online: http://dx.doi.org/10.1121/1.4800122
View Table of Contents: http://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCb: Production and Perception I: Beyond the Speech Segment (Poster Session)

4pSCb29. Speech rhythm and speech rate affect segmentation of reduced function words in continuous speech

Tuuli Morrill*, Laura Dilley, J. Devin McAuley and Mark Pitt

***Corresponding author's address: Communication Sciences and Disorders, and Psychology, Michigan State University, Oyer Center, B9, East Lansing, MI 48824, tmorrill@msu.edu**

Recent work (Dilley & Pitt, 2010, Psychological Science) has demonstrated that reduced function words in speech can perceptually disappear if the rate of surrounding speech is slowed, even when the acoustic properties of the function word (FW) and its immediate phonetic environment are held constant. An experiment was therefore conducted to determine whether this disappearing word effect could be elicited through a manipulation involving speech rhythm, realized as binary and ternary alternations of high and low tones, as well as through manipulations to context speech rate. 74 participants transcribed 32 sentences containing a FW in which the preceding speech within the utterance was resynthesized with a binary or ternary speech rhythm presented at one of three context speech rates. A binary rhythm in the preceding speech context yielded lower FW report rates than the ternary rhythm. These results suggest that listeners' expectations about speech rhythm and/or syllable grouping affected the number of syllables and words perceived, indicating that such properties may play an important role in word segmentation and lexical access. Work supported by NSF grant BCS-0847653.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Segmenting continuous speech is a highly complex task in which listeners must integrate multiple cues to potential word boundaries in the acoustic speech signal [1-2]. In addition to using well-described local phonetic cues to word boundaries [e.g.,3, 4-6], recent research has shown that prosodic information from the broader speech context affects word segmentation [7-10]. In the present work, we extend recent research showing a role for context speech rate in word segmentation [10-11] to determine if there is a role in word segmentation for speech *rhythm* as well, not just speech rate.

In previous work examining the role of speech rate in word segmentation [10-11], participants heard phrases which contained a phonetically reduced function word (FW), such as *or* in the phrase *Don must see the harbor or boats,* where the sentence could be heard as grammatical even without the FW. When the speech context was slowed down relative to this acoustically ambiguous FW, listeners reported hearing the FW significantly less often than when the context was presented at the same speech rate as the FW.

In a related line of work, cues to prosodic phrasing in the speech context have been shown to affect the way in which listeners group syllables into words [7-9]. In these studies, an ambiguously segmentable syllable sequence such as [tɑɪ məɹ dəɹ bi], which can be segmented as *timer, derby* or as *tie, murder, bee,* were combined with a preceding context in which the tonal and temporal patterns created contexts which either supported a disyllabic final parse of *derby* or a monosyllabic parse of *bee*. The effects found in these studies demonstrate that the prosodic context associated with the beginning portion of an utterance can influence the processes of word segmentation and lexical recognition in subsequent speech material.

In the current study, we combined manipulations used in previous studies to examine a possible role for speech rhythm, in addition to speech rate, in lexical recognition and word segmentation. Here, we examine the detection of acoustically reduced FWs under combined speech rate and speech rhythm manipulations, where speech rhythm is realized as patterns of repeated intonation contours. For the current experiment, we employed a speech rhythm manipulation in which repeating high (H) and low (L) tones were used to create two distinct rhythmic contexts: one formed a binary rhythm (with repeated patterns of H-L), while the other formed a ternary rhythm (with repeated patterns of H-L-L). Speech rate was also manipulated. Each stimulus item contained an acoustically ambiguous FW region, as in previous studies [10-11].We hypothesized that hearing a prior binary or ternary rhythmic context would cause listeners to learn the repeated tonal patterns of the context and the associated patterns of syllable and tone alignment. We predicted that this would lead them to develop different expectations about word boundary locations when hearing the binary as opposed to the ternary rhythmic context, thereby affecting the perception of later occurring lexical material and the rate of FW reporting, as has been demonstrated with manipulations of speech rate [10-11].

EXPERIMENT

Participants and Design

Participants were 49 adult native speakers of American English from the Michigan State University community who received course credit or nominal financial compensation for their participation. All participants had self-reported normal hearing. The experimental design consisted of a 2 [Rhythm: binary, ternary] x 3 [Speech Rate: 1.0 (i.e., unaltered), 1.4 (i.e., slowed by a factor of 1.4), 1.8 (i.e., slowed by a factor of 1.8)] mixed factorial design. Rhythm was a between-subjects factor, and Speech Rate was a within-subjects factor. Participants were randomly assigned to either the ternary rhythm condition (n = 24) or the binary rhythm condition (n = 25).

Materials

Stimuli consisted of 32 ten-syllable test sentences; each syllable comprised a monosyllabic word. 32 filler sentences also consisted of 10 syllables, but did not did not necessarily include a function word on the penultimate syllable, and some disyllabic words also occurred in these items. A function word (*are, or, our,* or *a*) occurred in the penultimate syllable of each stimulus sentence, and each sentence was constructed in such a way that it could be grammatical either with or without the function word. For example, in the sentence *Jill got quite mad when she heard* ***there are*** *birds*, the sentence (presented auditorily) would still be grammatical without the word *are.*

For the speech Rhythm manipulation, High (H) or Low (L) tones were assigned to the first six syllables (the "distal context") in either a ternary or binary pitch pattern (i.e., HLLHLL for the ternary condition or HLHLHL for the binary condition, with one tone per each of the six syllables per condition). In both Rhythm conditions, the final four syllables of the sentence comprised a "target region" in which the speech material, including the tonal pattern, was acoustically identical across conditions, as in previous studies [10-11]. Thus, in both Rhythm conditions there was a H-L-L-H pattern across the final four syllables. In each test sentence, the H tone was 10 Hz higher than the median F0 and the L tone was 10 Hz lower than the median F0. F0 values were assigned using the PSOLA (pitch-synchronous overlap-and-add) algorithm resynthesis in Praat [12].

The three distinct Speech Rate conditions were created by slowing down the context surrounding the function word with time expansion by a factor of 1.4 in the "1.4" condition, by a factor of 1.8 in the "1.8" condition; the original, spoken rate was maintained in the "1.0" condition. The speech rate of the function word itself, the syllable preceding the function word, and the onset of the syllable following the function word were unaltered. For example, in the fragment "*...heard* ***there are b****irds,"* only that portion consisting of [ðɛɹɑɹb] was unaltered, while the preceding and following context was subject to rate manipulation. Time expansion of the context was performed using PSOLA in Praat [13-14] and all soundfiles were normalized to 70 dB.

Procedure

Participants heard two practice items, followed by the 32 test sentences and 32 fillers presented in random order. Each test or filler sentence was presented auditorily while paired with a concurrent visual display of the first 6 syllables in the sentence. Participants were instructed to complete the sentence by typing the remaining words that they heard. Half of the filler sentences were presented with a visual display of the text missing the final four syllables, while half were displayed with the text missing the final three syllables. Participants' typed responses were coded for presence or absence of the critical function word.

RESULTS

Figure 1 shows the mean proportion of function words in each Rhythm and Speech Rate condition. Consistent with previous findings [10-11], the overall proportion of function words transcribed by listeners decreased as the context speech rate slowed. Notably, for each Speech Rate condition considered separately, FW report rates were lower in the binary Rhythm condition than in the ternary Rhythm condition. A logit mixed-effects model analysis [e.g., 15] was performed in the lme4 package [16] for the R statistical programming language (version 2.12.1; the R foundation for statistical computing) to assess the reliability of these effects. Model fit was assessed with ANOVA using log likelihood ratio [e.g., 17].

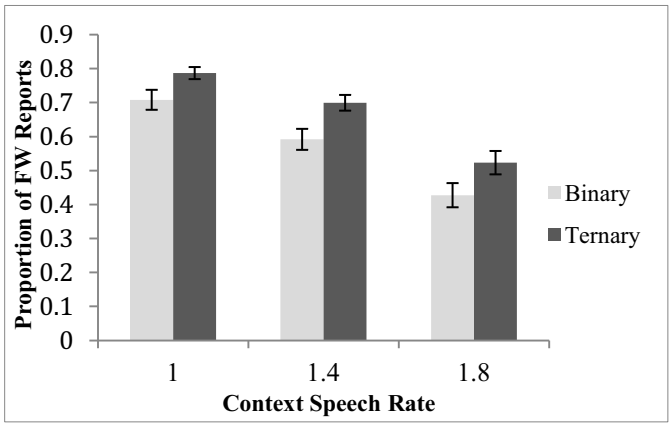

FIGURE 1. Mean proportions of function word reports in each Rhythm and Speech rate condition (by subject).

The full model was fitted with Rhythm and Speech Rate as fixed effects, as well as an interaction term for Rhythm and Speech Rate, with Subjects and Items as random effects. Both Rhythm and Speech Rate emerged as significant predictors of function word report rates (see Table 1), with higher FW report rates in the ternary condition than in the binary condition (p <.05), and significantly lower FW report rates in each of the slowed Speech

Rate conditions ($p < .001$). There were no significant interactions between Rhythm and Speech Rate ($p > .59$). The fit of the full model (with an AIC score of 1156.4) was significantly better than the fit for a model without speech Rhythm ($\chi = 10.229$, $p < .05$), or for a model without Speech Rate ($\chi = 157.94$, $p < .001$), indicating that both Rhythm and Speech Rate are needed to best account for the FW report rates.

TABLE 1. Logit mixed effects model with Rhythm and Speech Rate as fixed effects, and Subjects and Items as random effects.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.94194	0.61321	3.167	***p* < .005**
Rhythm - *ternary*	0.84547	0.38787	2.180	***p* < .05**
Rate – *1.4*	-1.28561	0.28020	-4.588	***p* < .001**
Rate – *1.8*	-2.52857	0.28559	-8.854	***p* < .001**
Rhythm * Rate (*1.4*)	0.22269	0.41675	0.534	$p = 0.594$
Rhythm * Rate (*1.8*)	-0.02081	0.40584	-0.051	$p = 0.959$

DISCUSSION AND CONCLUSION

These results indicate that speech rhythm, as signaled in the current study by binary or ternary patterned groupings of high and low tones, affected word segmentation and lexical recognition, in addition to the context speech rate. In particular, when the rhythm in the preceding context consisted of a binary rhythmic pattern, FW report rates were lower than when the preceding context consisted of a ternary rhythmic pattern. Because the ternary rhythmic pattern exhibited the same alignment of tones and syllables in both the context and target regions, this suggests that listeners developed expectations about speech rhythm and syllable and word boundary locations from the context speech, where these expectations affected the number of syllables and words perceived in the target region. Consistent with previous studies [10-11], temporal information in the speech context affects word segmentation, as shown by the fact that lower function word report rates occurred when the context speech rate was slowed. Moreover, patterns of linguistic prominence and rhythm play an important role in speech perception and lexical access, regardless of variability in speech rate. These results also suggest that tonal patterns alone can function as correlates of rhythm in speech and may serve as effective cues to prosodic and linguistic structure.

ACKNOWLEDGMENTS

We are grateful to Claire Carpenter for help with stimulus creation, and members of the MSU Speech Perception-Production Lab for help with running the experiment. This work was supported by Grant BCS-0847653 to Laura C. Dilley from the National Science Foundation.

REFERENCES

1. Cole, R.A. and J. Jakimik, *Segmenting speech into words.* Journal of the Acoustical Society of America, 1980. **64**: p. 1323-1332.
2. Lehiste, I., *The timing of utterances and linguistic boundaries.* Journal of the Acoustical Society of America, 1972. **51**(6 (Part 2)): p. 2018-2024.
3. Fougeron, C. and P.A. Keating, *Articulatory strengthening at edges of prosodic domains.* Journal of the Acoustical Society of America, 1997. **101**(6): p. 3728-3740.
4. McQueen, J.M., *Segmentation of continuous speech using phonotactics.* Journal of Memory and Language, 1998. **39**(1): p. 21-46.
5. Van Donselaar, W., M. Koster, and A. Cutler, *Exploring the role of lexical stress in lexical recognition.* The Quarterly Journal of Experimental Psychology, 2005. **58A**(2): p. 251-273.
6. Byrd, D., J. Krivokapic, and L. Sungbok, *How far, how long: On the temporal scope of prosodic boundary effects.* Journal of the Acoustical Society of America, 2006. **120, No. 3**: p. 1589-1599.
7. Dilley, L.C., S. Mattys, and L. Vinke, *Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation.* Journal of Memory and Language, 2010. **63**: p. 274-294.

8. Dilley, L.C. and J.D. McAuley, *Distal prosodic context affects word segmentation and lexical processing.* Journal of Memory and Language, 2008. **59**(3): p. 294-311.
9. Brown, M., et al., *Expectations from preceding prosody influence segmentation in online sentence processing.* Psychonomic Bulletin and Review, 2011. **18**(6): p. 1189-1196.
10. Dilley, L.C. and M. Pitt, *Altering context speech rate can cause words to appear or disappear.* Psychological Science, 2010. **21**(11): p. 1664-1670.
11. Heffner, C., et al., *When cues collide: How distal speech rate and proximal acoustic information jointly determine word perception.* Language and Cognitive Processes, 2012. **iFirst**(1-28).
12. Boersma, P. and D. Weenink, *Praat: doing phonetics by computer [Computer program].* 2002, Software and manual available online at http://www.praat.org.
13. Moulines, E. and F. Charpentier, *Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones.* Speech Communication, 1990. **9**(5-6): p. 453-467.
14. Moulines, E. and W. Verhelst, *Time-domain and frequency-domain techniques for prosodic modification of speech*, in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliural, Editors. 1995.
15. Jaeger, T.F., *Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models.* Journal of Memory and Language, 2008. **59**: p. 434-446.
16. Bates, D. and M. Maechler. *Ime4: Linear mixed effects models using S4 classes*. 2009.
17. Raudenbusch, S.W. and A.S. Byrk, *Hierarchical linear models: Applications and data analysis methods*. 2002, Newbury Park, CA: Sage.