



**ICA 2013 Montreal  
Montreal, Canada  
2 - 7 June 2013**

**Speech Communication**

**Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)**

**5aSCb53. One small step for (a) man: Function word reduction and acoustic ambiguity**

Laura Dilley, Melissa M. Baese-Berk\*, Stephanie Schmidt, Jesse Nagel, Tuuli Morrill and Mark Pitt

\*Corresponding author's address: Communicative Sciences and Disorders, Michigan State University, East Lansing, MI 48824, [mbaese@msu.edu](mailto:mbaese@msu.edu)

"That's one small step for man, one giant leap for mankind." Neil Armstrong insisted for years that his famous quote upon landing on the moon was misheard, and that he had said "one small step for a man." This controversy has continued, as examinations of the sound files of his transmission have yielded mixed opinions about whether he produced a. The disagreement stems partly from the fact that function words like a can be acoustically fleeting in casual speech, consist of just a few pitch periods, and be spectrally indistinguishable from the preceding context. As a result, they can be perceptually fragile, and easily disappear if the rate of surrounding speech varies (Dilley & Pitt, 2010, Psychological Science). Here, we examine naturally produced, reduced tokens of for (spoken as "fer") which were or were not followed by the word a from the Buckeye Speech Corpus, which consists of speakers from Mr. Armstrong's home state of Ohio. Comparison of the acoustic properties of the two sets of tokens will provide an indication of how similar they can be. Inclusion of Mr. Armstrong's production will assist in evaluating the likelihood of the function word being spoken. Work supported by NSF grant BCS-0847653.

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

## INTRODUCTION

“That’s one small step for man, one giant leap for mankind.” Neil Armstrong insisted for years that his famous quote upon landing on the moon was misheard, and that he had said “one small step for *a* man” (Crouch, 2012). This controversy has continued, as examinations of the waveforms and spectrograms generated from the sound files of his transmission have yielded mixed opinions about whether he produced *a* (BBC News, 2006). The disagreement stems partly from the fact that function words like *a* can be acoustically fleeting in casual speech, consist of just a few pitch periods, and be spectrally indistinguishable from the preceding context (e.g., Bell et al., 2003; Shockey, 2003). Therefore, it’s possible that an intended production of *for* and an intended production of *for a* may sometimes have identical, or nearly identical, acoustic signatures. Whether listeners hear an extra word (e.g., *a*) or not represents an example of the well-known problems in perception research of word segmentation and lexical access, i.e., how human listeners know how many word boundaries are in speech material, where word boundaries are located relative to acoustic cues, and how the boundary-delimited chunks of speech correspond to words in the lexicon (Cole & Jakimik, 1980; Lehiste, 1972). In the present research, we investigated the hypothesis that Neil Armstrong produced the word *a* in his famous quote, but that the nature of acoustic and contextual information for this quote tends to support this word’s being absent in perception.

Reduction is extremely frequent in casual speech (see Johnson, 2004). In studies of reduced speech, researchers have demonstrated that perception of reduced tokens in general is highly variable. For example, Ernestus, Baayen, and Schreuder (2002) demonstrated that strongly reduced forms are not well-recognized without substantial semantic and syntactic context. In the case of function words, Dilley and Pitt (2010) found that when prompted with a visual display to produce a sentence that contained a function word, talkers frequently produced the speech with heavy coarticulation (i.e., blending) between the function word and the preceding speech syllable(s). Under such conditions of reduction and coarticulation, listeners failed to report hearing the function word 21% of the time, even though talkers had spoken this word.

Subsequent research has shown that listeners hear function words less often when there is substantial coarticulation between a function word and the preceding syllable than when there is a less coarticulation (Heffner, Dilley, McAuley, & Pitt, 2012). The rate of context speech syllables also influences whether a coarticulated function word is heard (Dilley & Pitt, 2010; Heffner et al., 2012; Vinke, Dilley, Banzina, & Henry, 2009; Banzina & Dilley, 2010). Moreover, reduced syllables, including function words, are significantly less likely to be heard when they are short relative to their context, but even when such syllables are very short, they can often be made audible if the context speech rate is faster (Vinke et al., 2009; Banzina & Dilley, 2010). Furthermore, syntactic information in the speech context influences the segmentation of continuous speech (Mattys & Melhorn, 2007), including whether a heavily coarticulated function word is perceived as present or absent (Heffner & Dilley, 2011).

Our prior research thus suggests that the case of lexical ambiguity in the famous quote by Armstrong represents the “perfect storm” of conditions for a function word that was actually spoken to be missed perceptually in continuous speech. We theorize that Armstrong spoke the function word *a*, but that acoustic and syntactic factors converged to render the word difficult to hear. In the case of his quote, the syntactic context does not disambiguate whether the function word is present or not (since “man” can refer to “mankind” or else to a single, male individual). Previous work (Heffner & Dilley, 2011; Mattys & Melhorn, 2007), demonstrated that syntactic contextual information helps to disambiguate acoustic information regarding where word boundaries are present in speech. With no such information available, the differences between *for* and *for a* often may be difficult to disambiguate. Furthermore, the region of interest in the original quote [*for (a)*] is quite short and may be more consistent acoustically with typical productions of only the word *for* being spoken.

In the present study, we tested the hypothesis that in Armstrong’s dialect of American English, the temporal properties of naturally produced, reduced tokens of *for* (spoken as “fer”, [fɚ]) followed by *a* overlap with those of *for* (spoken as “fer”, [fɚ]) not followed by the word *a*. Our hypothesis was that in cases where *for* is pronounced [fɚ] the distributions of durational information in both sets of tokens would show substantial overlap for talkers in this region in Ohio, leading to significant ambiguity about the presence or absence of the word *a*. By examining these types of tokens, it will be possible to determine whether overlap in duration, in addition to perceptual fragility, may be responsible for the perceptually missing *a* in Armstrong’s original quote.

## METHODS

All data were extracted from the Buckeye Corpus for Conversational Speech (Pitt, et al. 2007), which contains spontaneous speech for 40 talkers from the Columbus, Ohio area, very near Armstrong's hometown of Wapakoneta, Ohio. Therefore, we anticipated that the dialects of the speakers in the Buckeye corpus should be similar to that of Armstrong, and could potentially show heavy reduction of function words, such as *a*. The speech is transcribed phonetically and orthographically, and transcriptions are time-aligned with the sound file. This alignment was hand-checked, allowing for search and extraction of durations of specific phrases. For this study, we relied on the transcription provided with the Buckeye Corpus, rather than determining whether there were spectral traces of *a* after *for*.

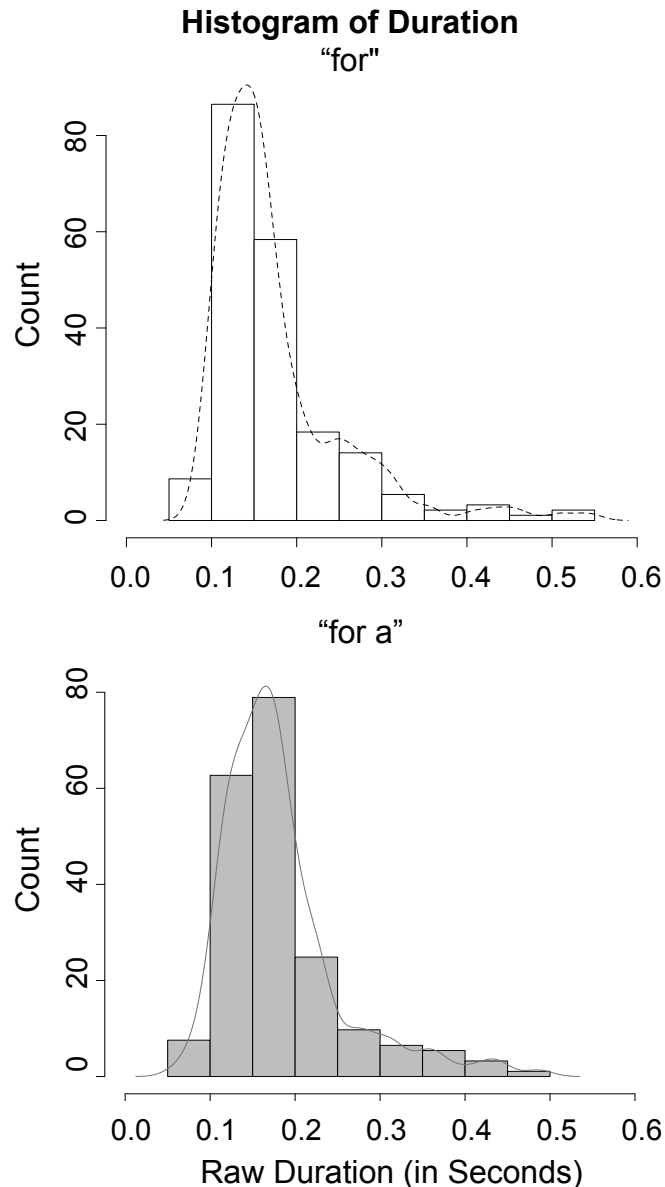
All instances of *for* followed by *a* ("*for a* tokens") were extracted from the corpus, for a total of 191 tokens. Each of these tokens was matched with a production of *for* followed by a noun ("*for* tokens"), spoken by the same speaker as each *for a* token. In all cases, *for* and *for a* tokens were spoken in a reduced style, i.e., *for* was transcribed in the Buckeye corpus as "fer", which corresponds to [fə] in the International Phonetic Alphabet (Kiesling et al., 2006). In many cases the presence of the word *a* is compatible with the surrounding, redundant syntactic and semantic information suggesting the same lexical interpretation, and this information may be in some cases the only information available for the inference of the presence of *a*. Previous work from our lab has suggested that people can use syntactic information to interpret an ambiguous signal relative to function word presence (Heffner & Dilley, 2011). In the case of the present study, the syntactic context largely supported the interpretation of inclusion or exclusion of the function word *a* for all selected tokens. The number of syntactically ambiguous contexts in the set of tokens was similar across the two categories (10 ambiguous contexts among *for* tokens; 11 among *for a* tokens).

The target region for durational measurement was defined as the words *for a* for *for a* tokens and as the word *for* for *for* tokens. Duration of the target region (*for* or *for a*) was calculated using the segmentation provided by the time-aligned transcription from the Buckeye corpus. We also examined the duration of [fə] in "for (a) [man]" from the original recording of the lunar transmission (NASA website, 2013).

## RESULTS

As shown in Fig. 1, distributions of the durations of the target regions for *for a* tokens and *for* tokens are very similar. Various measures of normalized duration (not shown), using the speech rate of the surrounding context, also yield similar patterns. A *t*-test reveals that, statistically, there is no significant difference between the duration of the target region in the *for* tokens and in the *for a* tokens ( $t < 1$ ;  $p > 0.5$ ). It is important to note that spectrally these two sets of tokens are also quite similar. In all cases reported here, *for* was reduced (to [fə]), allowing for substantial spectral acoustic overlap, even in cases when followed by a function word.

Moreover, the original recording of the lunar transmission revealed that the region of interest in Armstrong's original quote [*for (a)*] was 0.127 sec. Based on the distribution of duration information in the representative samples of *for* and *for a* tokens from talkers with a similar dialect in Ohio, the durational information of [fə] in Armstrong's original recording is highly compatible with either *for a* or *for* interpretations of the ambiguous stretch of speech. However, there is evidence that listeners internalize the statistical properties of the language to which they are exposed (e.g., Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). The histograms depicted in Fig. 1 suggest that if listeners were to rely predominantly or wholly on durational statistical information in their judgments of the lexical content of this portion of the phrase, then a duration of 0.127 sec would be more likely to be interpreted as *for* (with no following *a*) than as *for a*.



**FIGURE 1.** Histograms of duration for the target regions for *for* tokens (top) and *for a* tokens (bottom). The superimposed lines represent a smoothed function for each histogram. The x-axis represents the raw duration of the target region (in seconds). The y-axis represents the raw count for number of items of each duration bin.

## DISCUSSION

The data reported above suggest that productions of *for* tokens and *for a* tokens in this Ohio dialect of American English have substantial durational overlap. These findings therefore provide support for the hypothesis that Armstrong produced the word *a* in his famous quote upon the lunar landing, but that the acoustic and contextual cues which would typically support perception of the word *a* were ambiguous or missing (Banzina & Dilley, 2010; Heffner & Dilley, 2011; Heffner et al., 2012). These data also provide further evidence that function word reduction can lead to substantial acoustic ambiguity about the presence of a word, as shown by Bell et al. (2003) and Shockey (2003). Because the two categories of production (i.e., *for* and *for a*) substantially overlap in terms of duration, it is difficult to differentiate between the two possible productions based on durational information alone. The heavy coarticulation of *for* and *a* in the case of *for a* tokens furthermore suggests that spectral information is unlikely to

differentiate *for* and *for a* tokens in this American English dialect; recall that all instances of *for* in *for* tokens and *for a* tokens were pronounced as [fə] in the present analysis. Moreover, syntactic contextual information which sometimes influences perception of word segmentation (Mattys & Melhorn, 2007; Heffner & Dilley, 2011) is ambiguous in the case of Armstrong's quote. The present data suggests that durational information is likewise ambiguous between *for* and *for a* interpretations, based on a representative sample of comparable tokens from talkers with a similar dialect as Armstrong's. Previous research has shown that in the case of a heavily coarticulated function word, durational and speech rate information in the vicinity of the word can be critical for its successful perception (Banzina & Dilley, 2010; Heffner & Dilley, 2011; Vinke et al., 2009). The ambiguous syntactic, durational and spectral information about the presence of the function word *a*, combined with the substantial noise in the transmission from the moon to the Earth in the original recording of the lunar communication, probably created a situation in which the typical perceptual fragility of a coarticulated function word would be greatly heightened. By showing that the durational information in the ambiguous stretch of speech in Armstrong's quote is compatible with an interpretation of coarticulated productions of *for a* spoken by talkers in Ohio with a similar dialect, the present data serve to substantiate the longstanding claim by Armstrong that he actually said the word *a*. These data also lend support to the recent report by his brother that Armstrong showed him a scripted page with the word *a* in the quote months before his lunar landing (Gray, 2012). The present results, therefore, have not only theoretical implications for understanding spoken words, but also practical applications relative to the singular event of the first lunar landing, and how it has gone down in history.

## ACKNOWLEDGMENTS

This work was supported by NSF grant BCS-0847653 to Laura C. Dilley.

## REFERENCES

- BBC News. (2006). Armstrong 'got Moon quote right' [October 2, 2006]. Retrieved January 21, 2013, from <http://news.bbc.co.uk/2/hi/americas/5398560.stm>
- Banzina, E., & Dilley, L. (2010). Context speech rate and duration as cues to native and non-native perception of casually-spoken words in Russian. In Proceedings of the 5th International Conference on Speech Prosody, Chicago, IL. SProSIG.
- 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. *J. of the Acoust. Soc. of Am.*, 113(2), 1001. doi:10.1121/1.1534836
- Cole, R.A. & Jakimik, J. (1980). Segmenting speech into words. *J. of the Acoust. Soc. of Am.*, 64, 1323-1332.
- Crouch, Ian. (August 25, 2012). Neil Armstrong's voice. *The New Yorker: News Desk*. Retrieved January 19, 2013, from <http://www.newyorker.com/online/blogs/newsdesk/2012/08/neil-armstrongs-voice.html>.
- Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664-1670.
- Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1-3), 162-173. doi:10.1006/brln.2001.2514
- Heffner, C. & Dilley, L. (2011). Unlikely allies: Acoustic and syntactic cues in word segmentation. Presented at the Midwest Cognitive Science Meeting, Michigan State University, East Lansing, MI, April 16, 2011.
- Heffner, C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2012). When cues combine: how distal and proximal acoustic cues are integrated in word segmentation. *Language & Cognitive Processes*, 1-54.
- Gray, R. (2012). Neil Armstrong's family reveal origins of 'one small step' line [December 30, 2012]. Retrieved January 22, 2013, from <http://www.telegraph.co.uk/science/space/9770712/Neil-Armstrongs-family-reveal-origins-of-one-small-step-line.html>
- Johnson, K. (2004). Massive reduction in conversational American English. *Spontaneous Speech: Data and Analysis*. Proceedings of the 1st Session of the 10th International Symposium, 29-54.
- Kiesling, S., Dilley, L., & Raymond, W. (2006). The Variation in Conversation (ViC) Project: creation of the Buckeye Corpus of conversational speech. Unpublished manuscript, Columbus, OH: Department of Psychology, Ohio State University, available at [www.buckeyecorpus.osu.edu](http://www.buckeyecorpus.osu.edu).
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *J. of the Acoust. Soc. of Am.*, 51, 2018-2024.
- Mattys, S., & Melhorn, J. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *J. of the Acoust. Soc. of Am.*, 122(1), 554-567.
- NASA website. (2013). Retrieved January 22, 2013 from [http://www.nasa.gov/mission\\_pages/apollo/apollo11\\_audio.html](http://www.nasa.gov/mission_pages/apollo/apollo11_audio.html)
- Pitt, M., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). *Buckeye Corpus of Conversational Speech* (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

- Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101-105.
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Wiley-Blackwell.
- Vinke, L., Dilley, L., Banzina, E., & Henry, M. (2009). Lexical perception and segmentation of words beginning with reduced vowels: A role for timing cues. *Abstracts of the Psychonomic Society*, 14, 132.