ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

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

4pSCb33. Calibrating the detection of spontaneous speech: From sentences to noun phrases
Sara Parker and Jennifer Pardo*

*Corresponding author's address: Psychology Department, Montclair State University, 1 Normal Avenue, Montclair, New Jersey 10590, pardo@optonline.net

Many studies have examined the differences between speech that is produced spontaneously as opposed to read from a prepared script. Most of these studies have focused on prosodic measures taken from clauses, sentences, or connected discourse. Furthermore, studies have shown that listeners are able to identify the context of production when presented with sentence-length utterances. The current study examined whether a listener can identify the context for utterances that are briefer than a sentence. A set of 20 talkers (10 male) produced spontaneous descriptions of maps that they then read aloud in a separate session at least one week later. Pairs of sentences that matched in fluency across both contexts were selected, and listeners judged which member of a pair was produced spontaneously. In separate blocks, listeners heard either full sentences, sentence beginnings, sentence endings, or two-word noun phrases excised from sentences. Overall, listeners could identify the spontaneously produced utterances, but only for excerpts longer than two-word noun phrases. These findings indicate that the information present in two-word noun phrases is not sufficient to support perception of spontaneous versus read speaking style.
Talkers vary their speaking style across different settings of language use. Many studies have demonstrated that various characteristics of speech produced spontaneously or casually differ from that of speech that is either read from a prepared script or produced with the intent to speak clearly (e.g., Blaauw, 1994; Bradlow, 2002; Laan, 1997; Picheny, Durlach, & Braida, 1986). Moreover, some studies have also found that spontaneous/causal speech is less intelligible to listeners than clear/read speech (Picheny, Durlach, & Braida, 1985; Smiljanic & Bradlow, 2008), and that listeners can identify the speaking mode to some extent (Blaauw, 1994; Laan, 1997). Although the impact of speaking mode on intelligibility has been established for the full range of utterance types (from words to discourse), the ability to identify speaking mode has only been reported for utterances of sentence length or longer. Thus, it is not clear to what extent talkers rely on segmental and/or suprasegmental attributes when identifying speaking mode. The current study attempts to calibrate the ability of a listener to detect spontaneous speech mode for briefer utterances. If listeners are unable to identify speaking mode in briefer utterances, then phonetic attributes are unlikely to play a role in spontaneous speech detection.

Research on speaking mode falls into two parallel veins, one that emphasizes the productive factors that lead to the intelligibility benefits of clear speech style (mainly segmental/acoustic-phonetic), and another that emphasizes those contributing to the ability to identify spontaneous versus read speech style (mainly suprasegmental/prosodic). The bulk of the effort has gone to studies of intelligibility, probably because improving intelligibility is a more pressing issue than improving the quality of speech synthesis. Early on, Lieberman (1963) demonstrated that words produced in a predictable context were more reduced (in duration and amplitude) and less intelligible than the same words produced in an unpredicatable sentence (A stitch in time saves nine vs. The number that you will hear is nine). In later studies, talkers were instructed to produce speech that would be highly intelligible, so-called clear speech. In doing so, their speech contained less reduced vowel formants and increased duration of overall sentences, individual phonemes, and even features of phonemes such as VOT of voiceless consonants (Picheny et al., 1986). Furthermore, these utterances were more intelligible than so-called conversational-style speech (Picheny et al., 1985).

More recent research has confirmed the hyperarticulation of clear versus casual speech production and extended the analysis of intelligibility to utterances in longer stretches of discourse (Bradlow, 2002; Smiljanic & Bradlow, 2008a, 2008b). Talkers maintained a clear speech style in paragraph-length utterances, and temporal attributes appeared to be most responsible for the distinction. While the authors of these studies often contrast clear speech style with the more natural spontaneously produced style that a listener typically encounters, none of these studies actually employed spontaneous speech. The materials were always read from a prepared script with instructions to produce the utterances in two different styles, clear and casual. Thus, some of the acoustic-phonetic factors that have been found to contrast the production of clear versus casual speech, leading to enhanced intelligibility, might not align with the sorts of variation that occur incidentally when a talker switches from spontaneous to prepared speech modes. Likewise, identification of speech mode might not rely on the sorts of acoustic-phonetic attributes that have been found to contribute to the intelligibility of clear or prepared speech.

In research that focuses on identification of speaking mode, the methods have contrasted speech that was actually produced spontaneously with speech that has been read from a transcript of the spontaneous utterances. Across multiple studies, listeners have been found to identify utterances as spontaneous or read better than chance, with performance reaching as high as 84% correct identification of speech mode. The highest estimates of performance are found for utterance pairs that preserve the typical fluency differences between the speech modes, and identification performance was still at around 70% correct even when sentence pairs are matched for fluency (see Laan, 1997). Most of the observed differences in these studies centered on prosodic attributes, but some, like duration or speaking rate, are relevant for both segmental and suprasegmental levels of analysis. Both Blaauw (1994) and Laan (1997) modeled the relation between acoustic features and perceptual identification judgments in regression analyses, finding that the most important predictors were prosodic attributes such as F0 declination, speaking rate, and various correlates of prosodic boundaries. In particular, F0 declination was more consistent in read than in spontaneous speech (Lieberman, Katz, Johgman, Zimmerman, & Miller, 1985).

To date, there are no published empirical reports that examined spontaneous speech mode identification for utterances that are briefer than sentence length. Blaauw (1994) mentions in passing that utterances of 4-6 syllables in length were classifiable above chance, but at a much lower rate than for entire utterances. It is clear that the prosodic differences between spontaneous and read speech are related to a listener’s ability to identify speech mode for utterances of sentence length. For briefer utterances, listeners would need to rely on attributes more closely related to the acoustic-phonetic scale. If listeners can identify speech mode for briefer utterances, then the same attributes that support increased intelligibility for clear speech likely support speech mode identification. If these attributes do not support speech mode identification in briefer utterances, then it is likely that intelligibility of clear speech mode and identification of speech style arise from independent processes in perception.

The current study contributes to these findings by examining speech mode identification in utterances of sentence length, partial sentences, and noun phrases. Moreover, the set of talkers in the current study is much larger.
than in those studies, which only examined between 1 and 7 talkers, most of whom were male. In all studies, the authors noted that there were large individual differences among the talkers in the particular attributes that differentiated their speech modes. In order to obtain samples of spontaneous speech, a set of 20 talkers (10 male) produced descriptions of the Giver maps from the HCRC Map Task (Anderson et al., 1991). Each map consisted of a set of labeled iconic landmarks with a path drawn from a starting point, around various landmarks to a finish. The talkers were asked to describe the path on their map in sufficient detail that another person could draw the path on a matching map. These descriptions were transcribed orthographically, and the talkers returned to the laboratory to read their transcriptions aloud at least one week later.

In order to assess a listener’s reliance on phonetic attributes when identifying speaking mode, it was necessary to identify pairs of sentence-like utterances containing landmark label phrases that matched in lexical composition and relative fluency. For each talker, 4 sentence pairs that met these criteria were identified. Then, 4 versions of each sentence pair were generated for use in the listening tests: the full sentences, the first half of each sentence, the second half of each sentence, and the isolated landmark label phrases. These items were used to create an AX perceptual identification task. On each trial, a listener heard both versions of an item and identified which one (first or second) was produced spontaneously. The test blocked items by version type (full sentence, first half, second half, or landmark label), and each listener provided judgments for all 4 types of items from all 20 talkers. However, within a listening test, no sentence fragment was repeated, thus different sets of listeners provided judgments for all combinations of sentence fragments and talkers.

Overall, listeners were able to identify speaking mode for only the full sentences. Identification of speaking mode for briefer utterances, in particular, the isolated landmark label phrases, was no better than chance. Moreover, there were large individual differences across talkers in how well their speech mode could be identified. These findings indicate that although speaking mode has been found to impact phonetic factors, such as vowel dispersion, and lead to increased intelligibility relative to casual or spontaneous speech, listeners are unable to identify speaking mode as such in utterances that are briefer than sentences. Therefore, the segmental and suprasegmental differences between spontaneous and read speech modes appear to support enhanced intelligibility, but identification of speaking mode appears to rely more heavily on suprasegmental attributes present in full sentences.

ACKNOWLEDGMENTS

Completion of this work was funded in part by NSF PAC award #BCS-1229033. The authors thank Hannah Gash, Kelly Jordan, Eva Lewandowski, Rolliene Mallari, Catherine Ode, and Caitlin Scanlon for assistance with data collection.

REFERENCES


