Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 50 (2008) 323–336
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

A new method for eliciting three speaking styles in the laboratory

James D. Harnsberger *,[1], Richard Wright [2], David B. Pisoni

Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, United States

Received 4 May 2007; received in revised form 1 November 2007; accepted 2 November 2007

Abstract

In this study, a method was developed to elicit three different speaking styles, reduced, citation, and hyperarticulated, using controlled sentence materials in a laboratory setting. In the first set of experiments, the reduced style was elicited by having 12 talkers read a sentence while carrying out a distractor task that involved recalling from short-term memory an individually-calibrated number of digits. The citation style corresponded to read speech in the laboratory. The hyperarticulated style was elicited by prompting talkers (twice) to reread the sentences more carefully. The results of perceptual tests with naïve listeners and an acoustic analysis showed that 6 of the 12 talkers produced a reduced style of speech for the test sentences in the distractor task relative to the same sentences in the citation style condition. In addition, all talkers consistently produced sentences in the citation and hyperarticulated styles. In the second set of experiments, the reduced style was elicited by increasing the number of digits in the distractor task by one (a heavier cognitive load). The procedures for eliciting citation and hyperarticulated sentences remained unchanged. Ten talkers were recorded in the second experiment. The results showed that 6 out of 10 talkers differentiated all three styles as predicted. In addition, all talkers consistently produced sentences in the citation and hyperarticulated styles. Overall, the results demonstrate that it is possible to elicit controlled sentence stimulus materials varying in speaking style in a laboratory setting, although the method requires further refinement to elicit these styles more consistently from individual participants.
Published by Elsevier B.V.

Keywords: Speaking styles; Speech perception; Speaking rate; Vowel dispersion

1. Introduction

Traditionally in studies of speech production and perception that use natural speech, utterances are recorded under highly controlled conditions in a laboratory setting. Control over the audio recording conditions and the nature of the materials recorded (particular syllables, words, sentences) serves to limit sources of error in the data collection process, or to avoid particular confounds that might render the results uninterpretable. Control over the quality and structure of the materials also insures that any results can be replicated in other laboratories, a key aspect of any experiment. However, it has long been recognized that the style of speaking elicited from talkers reading linguistic material aloud in a laboratory setting differs systematically from other speaking styles that occur naturally, such as more reduced styles of speech that can be observed in unmonitored conversations (Summers et al., 1988; Picheny et al., 1989; Byrd, 1994) and hyperarticulated or clear speech, whether directed to normal and hearing-impaired listeners (Picheny et al., 1985, 1986, 1989; Payton et al., 1994; Uchanski et al., 1996), infants (Fernald et al., 1989; Kuhl et al., 1997), non-native speakers (Bradlow and Bent, 2002; Uther et al., 2007), or automatic speech recognizers (Oviatt et al., 1998a,b). Differences between reduced and laboratory citation speaking styles can include the duration of the utterance and its constituent words, pausing, and the degree of centralization in the quality of vowels, to name a few.

While citation speech is fairly uniform in its characteristics from word to word and repetition, conversational

* Corresponding author. Tel.: +1 352 392 2113x271; fax: +1 352 846 0243.
E-mail address: jharns@ufl.edu (J.D. Harnsberger).
[1] Now at the University of Florida, Gainesville, FL, United States.
[2] Now at the University of Washington, Seattle, WA, United States.

0167-6393/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.specom.2007.11.001

speech is typified by a wide range of hyperarticulation and reduction at the word and syllable levels (Krull, 1989; Duez, 1992; Wassink et al., 2007). The probability of any one word being reduced (*hypoarticulated*) or hyperarticulated is widely thought to be related to its informational load (i.e., Lindblom, 1996; Jurafsky et al., 2001; Aylett and Turk, 2004); the less important the word is to a particular utterance's meaning and the more predicable it is from context, the more hypoarticulated it typically is. Likewise, if a word is particularly important to the meaning of an utterance and if it is not predictable from the preceding set of words, it is likely to be hyperarticulated. Thus, the use of conversational speech in controlled experiments poses another problem to the experimenter: It is difficult to control the degree of reduction or hyperarticulation on any particular word. Like reduced speech, hyperarticulated speech displays features that distinguish it from citation speech, including vowel space expansion, a decrease in speaking rate, increase in the number and duration of pauses, an increase in energy in the mid-frequency range of the spectrum, and a decrease in disfluencies, among others. Such hyperarticulated speech effects have been observed primarily in American English, but a few studies observe similar patterns in other languages (e.g., Croatian – Smiljanić and Bradlow, 2005; German – Köster, 2001; Jamaican – Wassink et al., 2007). It is a style adaptively employed by speakers to enhance recognition and comprehension by the listener given the listening environment at hand (Lindblom, 1990; Moon and Lindblom, 1994; Lindblom, 1996; Jurafsky et al., 2001; Aylett and Turk, 2004). It also occurs frequently in human–computer interaction when automatic speech recognizers produce errors; speakers will shift to a hyperarticulated speaking style which, while helpful in human speech communication, can actually limit the success of human–computer interaction given the automatic speech recognizers are not commonly trained on this style (Hirschberg et al., 2004).

The differences between natural speaking styles and speech elicited in laboratory reading tasks also pose a problem for theories of speech perception and spoken word recognition by human listeners, most of which have been formulated from studies using controlled speech materials: To what extent do these findings generalize to the speaking styles that people produce and perceive in natural settings? The perception of variability that exists among speech styles has not been studied in detail, no doubt due to the problem of eliciting naturalistic speech in the decidedly unnatural manner and the setting of reading aloud in a laboratory, although these issues have begun to be addressed in computer/machine word recognition (Ostendorf et al., 1996; Schriberg, 2001; Liu et al., 2006; Bates et al., 2007). Other types of "nonlinguistic" variability have been shown to have an effect on speech perception and spoken word recognition, including talker, rate, and stimulus variability (Mullennix and Pisoni, 1990; Nygaard et al., 1995; Bradlow et al., 1999). These studies suggest that listeners encode in long-term memory significant episodic details and properties of speech signals that they encounter, and that these details influence the subsequent perception and recognition of speech. If listeners are sensitive to highly detailed, episodic properties of speech, then variation in those properties due to speaking style differences may also play an important role in speech processing, one that has thus far been neglected in speech perception and spoken word recognition research.

One factor limiting the study of the perception of speaking styles has been a methodological one: How does a researcher elicit different speaking styles, including more reduced and naturalistic speech, while at the same time controlling for the particular syllables, words, or sentences to be studied? Recording natural conversation, or guided conversation on a particular topic, has been used in the study of sociolinguistic variability in speech production, namely, in the elicitation of stigmatized, or less prestigious, sounds, words, or syntactic structures of a dialect (Labov, 1972; Milroy, 1987). In other methods, subjects have been asked to participate in and narrate a task (Hirschberg and Nakatani, 1996). Such procedures have been useful in eliciting particular intonational forms and in studying specific aspects of discourse structure (Swerts and Collier, 1992; Speer et al., 1999). However, none of these methods can guarantee the elicitation of specific sentences. What is needed is a technique to elicit a variety of speaking styles found outside the laboratory, particularly more hypoarticulated styles, while controlling for linguistic content. Speech samples elicited by such a method would represent an important point on the continuum between experimental control and ecological validity.

To address this research need, a method has been developed for eliciting sentences in different speech styles in the laboratory while controlling for the particular sentence materials used. The range of speech styles includes a reduced, or hypoarticulated style, that should more closely resemble the speech style employed in natural settings in conversation than laboratory read speech. The first version of this method was described by Brink et al. (1998). Brink et al. (1998) attempted to elicit three speaking styles, namely, reduced, or hypoarticulated speech; citation, or read speech, a style that is normally used in reading controlled materials in a laboratory setting; and hyperarticulated speech. Each style was elicited in a separate condition of the experiment.

Brink et al. (1998) elicited reduced speech by having subjects read a sentence while engaging in a concurrent digit span task that involved remembering a digit sequence of five to seven digits in length that was presented immediately prior to the sentence. After reading the sentence, subjects were asked to recall the digit sequence in the same order in which it was presented. The digit span task was a distractor task, chosen to place the subject under a cognitive load while reading a sentence. It was chosen as the distractor task because, in piloting, it was successful in producing the desired speech style while minimizing talker disfluencies. The logic behind the use of any distractor task to elicit a reduced speaking style was based on two sets of