PERCEPTION OF CASUAL SPEECH IN A FOREIGN LANGUAGE

Linda Shockey
The University of Reading, U.K.

ABSTRACT
Previous experiments with gated speech have shown that native speakers of English are very skillful at interpreting informal English phrases containing conversational speech reductions. Results suggest that they have (at least) two modes of parsing an incoming signal: a local mode in which immediate phonetic/phonological environment is taken into account and a global mode in which the entire utterance is scanned for coherence. The former is used most frequently when perceiving one's own language, but the latter is available for interpretation of unclear input. When non-native speakers of English are presented with the same gated English sentences, their interpretations are surprisingly inaccurate and appear to reflect much greater use of global scanning than is found in the native speakers. The present paper asks 'Do non-natives do better on natural speech containing only simple vocabulary?'

1. INTRODUCTION

My research has been aimed at answering the question “How does casual, conversational speech differ from citation form speech?” in the hope that the answers will prove interesting both theoretically and practically. Applications to automatic speech recognition and speech synthesis are obvious, as are potential contributions to our understanding of normal speech perception.

It has been regularly observed that casual speech is phonetically less extreme than citation form speech: measurement of vowels and of consonant transitions suggest considerable centralisation [1]. Electro-palatography has shown less - sometimes startlingly less - overall contact between the tongue and the palate in casual speech in English and other languages [2]. Perceptually, it has been shown that words which are receiving their nth mention are less acoustically distinct than words which are newly introduced [3,4].

In this paper, I consider cases of phonological difference. My research and that of others indicates that minimal feature changes such as devoicing as well as more major structural changes such as reduction of final clusters are frequent in casual speech.

How is potentially ambiguous reduced speech understood? When a speaker of English hears a phrase like [wʊdˈn̩tʃuː], how do they know whether it is “wouldn’t you” or “wooden shoe”? Clearly, context plays a part, but how much and what sort of information do we need to gain access to the lexicon?

2. EXPERIMENTAL TECHNIQUE

A technique I have used to explore this question [5,6,7] is gating, which consists of recording a phrase or sentence, cutting off the majority of its end, and then first presenting the ‘amputated’ sentence on its own, then adding 50 msec. of speech at a time to the end of the utterance until the whole thing is restored. A few seconds of silence are introduced after each stimulus, and the subjects are asked to write what they hear in normal English spelling each time a stimulus is presented.

An early sentence which I tried was “The screenplay didn’t resemble the book at all,” pronounced [dɪˈskrɪmplɪdɪn ˈɛzəmblɪdəˈbʊkətɔl], with two major differences from citation form: ‘screen’ was pronounced ‘scream’ and ‘didn’t’ was reduced to something close to ‘din.’ (There was a very short period of non-nasal closure before the velum was lowered). Native speakers had very little trouble understanding the gated sentence (81% got it entirely correct) and demonstrated two strategies for doing so:

1. Early reporters: most listeners heard ‘scream’ until the conditioning factor for the m (i.e. the [p]) appeared, whereupon they began to report hearing ‘screen.’
2. Late reporters: a much smaller group of listeners did not report hearing ‘screen’ until several more words were revealed. These behaved as if they were not depending on phonetic information alone, but were resorting to a more global scan. [8,9]

Interestingly, many more listeners used Strategy 2 to recognise “didn’t,” which was more highly reduced and for which the conditioning factors were not so obvious. The recognition point for this word was spread out from just after the end of the word itself to much later in the sentence.

The same stimuli were then tried on a group of non-native speakers, who showed a remarkable inability to understand the sentence. 62% of the subjects changed ‘scream’ to ‘screen,’ but only 25% recognised ‘didn’t.’ In addition, most of the...
subjects were late reporters: reduced forms were recognised, on the whole, long after the end of the relevant words.

These experiments were criticised on two grounds. First, the vocabulary was difficult for non-natives: ‘screenplay’ is not a frequent term in EFL materials, and even ‘resemble’ could be relatively unfamiliar. Second, the stimuli were produced by the author, a native speaker of American English. This could have introduced difficulties on because 1) they were not taken from natural discourse and 2) the accent may not have been familiar to foreign students in England.

3. PRESENT EXPERIMENT
A short sentence (‘So it was quite good fun actually on the wedding, though’) was selected from a monologue by a native speaker of Standard Southern British. This sentence is made up of uncontroversial vocabulary items and was spoken with enough phonological reduction to be interesting. It also exhibits slightly unusual syntax, which reduces the chance of right answers through guessing. Phonetically, the sentence was [sɪˈskwərəʊdˈfɪmetəʊˌwɪmənəˌwɛdɪŋdəʊ], where the schwa in ‘was’ was rounded and the ‘d’ in ‘good’ did not achieve full closure. The second [n] is dental, as is the [d] in ‘though.’

The sentence was gated beginning just after the initial ‘s’ and presented to nine native speakers of Greek who were students or staff at Reading University. All had lived in England for some time and used English regularly as a teaching/learning medium and in their social lives.

4. RESULTS
Results were very similar to those found for non-native teaching/learning medium and in their social lives. ‘So it was’ at the beginning. Eight perceived ‘quite,’ four whole sentence perfectly, largely because of the highly reduced sentences reported above: of the nine, none understood the sentence was gated beginning just after the initial ‘s’ though. The second [n] is dental, as is the [d] in ‘though.’

whether 2 above is inevitable is a point to consider: will intentional focus on reduced speech while studying a foreign language give adult learners access to native-speaker strategies or will attempts to understand another language always involve “perceptual noise” which results in the use of relatively inefficient global strategies?

REFERENCES