SPOKEN LANGUAGE VARIABILITY

IMPLICATIONS FOR MODELING SPEECH PERCEPTION

ABSTRACT

Keith Johnson
Ohio State University
Department of Linguistics & Center for Cognitive Science

IMPORTANCE FOR MODELING SPEECH PERCEPTION
**FUNCTION WORDS**

Beyond the system of contrasts that the lexicon defines, the larger syntactic and semantic context delimits a range of permissible speech reduction patterns. This is especially apparent in the realization of function words because they follow different patterns of reduction than do content words (Lee, [7], p. 100). For example, the word **to** send it in (which at a lexical level of reception is unquestionably present to native speakers) is a release burst and a bit longer [s] in the following word, which occurs in the phrase "votes or something".

Beyond the system of contrasts that the lexicon defines, the larger syntactic and semantic context delimits a range of permissible speech reduction patterns. This is especially apparent in the realization of function words because they follow different patterns of reduction than do content words (Lee, [7], p. 100). For example, the word **to** send it in (which at a lexical level of reception is unquestionably present to native speakers) is a release burst and a bit longer [s] in the following word, which occurs in the phrase "votes or something".

**Figure 4.** Contextual retroflexion in the phrase "votes or something".
of a word pronunciation, in which the utterer exaggerates some aspect
in the VC database where amutation instances of highly

6. EMPIRICAL OBSERVATIONS

words.

The examples discussed in this section illustrate that function
words can be phonetically reformulated, may be gradiently reduced,
and in some cases may be signaled primarily by rhythm. These
words can be distinguished from one another by those elements of the
preceding phonetic context that are stored, and in some cases may be
metaphorically reinterpreted. Any empirical results obtained

The word "have" is realized in /[ts/.

Figure 4. The word "to" is realized as [ts].

Figure 5. Cues, including F3 amplitude and rhythm, for the /h/ of
"have" are weak but present.

Figure 6. A "rhythmic word" in "sort of".

EMPHATIC PRONUNCIATIONS

In the VC database there are numerous instances of highly
emphatic productions in which the talker exaggerates some aspect
of a word pronunciation, in which the utterer exaggerates some aspect
8. IDIOSYNCRATIC PHONETIC SPACE

Another kind of variability that we have found in the ViC is roughly consistent with that found in another talker's productions with a closure of the voiceless fricative /s/. What is idiosyncratic for this speaker is that the spectrum for /s/ has more low frequency noise than usual. What is idiosyncratic for this speaker is that the spectrum for /s/ has more low frequency noise than usual.

It is sometimes said that prosody provides a parallel channel with a variety of prosodic dimensions (intonation, rhythm, stress, etc.) that is separate from the lexically specified information. For example, the pitch of a question mark can be used to convey a strong interrogative meaning or a weak request. In addition, the length of a word's stress can convey different levels of emphasis. For example, a longer stress on a specific syllable can indicate that the word is more important or that it is being repeated. These prosodic features interact with the lexical meaning of the word to create a more complete understanding of the speaker's intent.

If we consider the production in Figure 7, we see that the word is pronounced with a clear emphasis on the first syllable. This can be interpreted as indicating that the speaker is particularly interested in this word, or that it is important to the context. The fact that the word is pronounced with a strong emphasis on the first syllable suggests that it may be a key point in the speaker's message. This is consistent with the idea that prosody can provide additional information about the speaker's intent, beyond what is contained in the lexical meaning of the word.
Figure 9 is reminiscent of Sapir's (1921) discussion of "the intuitive 'placing' of the sounds with reference to one another." In this discussion he touches on individual variation. "It is true that no two individuals have precisely the same pronunciation of a language, but it is equally true that they aim to make the same sound discriminations, so that if the qualitative differences of the sounds that make up A's pattern from those that make up B's are perceptible to a minute analysis, the relations that obtain between the elements in the two patterns are the same."

The relational invariance that Sapir appeals to is directly analogous to ratio coding schemes of vowel normalization that have been proposed in more recent decades. But the case of "township trustee" does not quite correspond to a ratio coding scheme, because it is not as if the playback speed of the recording has been altered (the typical situation inviting of a ratio scheme). Instead the talker simply has a somewhat idiosyncratic pronunciation of only one phoneme [s]. To "normalize" to this talker's production of [s] the listener needs to warp only a small portion of the perceptual space.

The two phrases in this utterance. The first phrase is pronounced by the speaker of the phrase "township trustee." The second phrase is pronounced by the speaker of the phrase "a little skirt on her." The first phrase is spoken more softly and with a breathy phonation type (as indicated by the ratio of the first two harmonics). The second phrase is spoken more loudly, and with a creaky phonation type (as indicated by the ratio of the first two harmonics).

Figure 9. A production of the phrase "township trustee" showing that [s] is spectrally similar to [S] for this talker.

9. MEANINGFUL VARIATION

The theme of this paper has been that variability in spoken language is "meaningful." I have shown examples from unmonitored conversational speech that illustrate this point at a number of levels of analysis - that variability is shaped by a language's system of contrasts, the syntactic function of words, or the emphasis that a talker wants to convey. The examples in the last two sections also highlighted the fact that individuals differ. Johnson, Strand & D'Imperio (1994) put forward the idea that people "perform" gender in their speech. In keeping with this perspective I would like to conclude with an example of spoken language variability that arises from this performance aspect of ordinary speech.

The phrases shown in figure 10 were spoken in the context of a young mother talking about how people dress their children. She said, "...put a sweater on her on Tuesday, and a little skirt on her on Wednesday." The words by themselves don't convey the talker's meaning in this utterance. The first phrase is spoken rather softly, with a breathy phonation type (as indicated by the ratio of the first two harmonics) and with a higher F0. The second phrase is spoken more loudly, with a creaky phonation type (as indicated by the ratio of the first two harmonics). The effect of these changes is to indicate that the talker is making fun of the behavior she is describing - treating a child like a dress-up doll. This information is conveyed by the phonetic differences between the two phrases in this utterance.

People use subtle phonetic cues to convey nuanced information about the communicative context when they talk to each other. Thus, in contrast to the traditional view that speech perception theory merely needs to account for phoneme recovery, this look at conversational speech suggests that variability in spoken language is not merely random noise as a simple linear associator model would assume, and it is not merely "lawful" as a mechanistic phoneme transmission view would assume. Rather, people use subtle phonetic cues to convey nuanced information about the communicative context when they talk to each other.
Spoken language varies as talkers blend together complex linguistic and supralinguistic messages. This rich "meaningfulness" of variability is a worthy focus of study which presents some interesting deep challenges to conventional theories of speech perception and auditory word recognition.

Acknowledgements

The Variation in Conversation corpus is the result of a joint collaborative project of the Departments of Linguistics and Psychology at Ohio State University. Funding was provided by Grant No. R01 DC04330-01 from the National Institute of Deafness and Other Communication Disorders, to Mark Pitt, Elizabeth Hume, and Keith Johnson. Personnel on the project include Scott Keisling, William Raymond, Jennifer Muller, Matt Mashay, and Robin Dautricourt.

Keith Johnson, 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210. Email: kjohnson@ling.osu.edu.

References


