# RECONSTRUCTION OF BASE FORMS IN PERCEPTION OF CASUAL SPEECH

Linda Shockey, Linguistic Science
Anthony Watkins, Psychology
University of Reading, UK

## ABSTRACT

Perception of normal, relaxed speech involves relating phonologically reduced forms to their mental representations (assuming that lexical storage does not simply involve making a list of all pronunciations of all base forms). Many researchers into how the lexicon is accessed assume that word recognition normally occurs within the phonological boundary of the word being processed and that therfore it is only in exceptional cases that a decision about the identity of a word is postponed. The research reported here suggests that such delayed recognition may be a very commonly-used strategy for understanding of conversational speech forms.

## THE EFFECT OF CASUAL SPEECH PHONOLOGY ON PRONUNCIATION

It has been demonstrated [4,6] that there is phonetic reduction in words which have once been focal but have since passed to a lower information status: the first time a word is used, its articulation is more precise and the resulting acoustic signal more distinct than in subse- quent tokens of the same word. By 'phonetic' we mean that the effect can be described in terms of of vocal tract inertia and ease of articulation: since the topic is known, it is not necessary to make the effort to achieve a maximal pronunciation after the first token. We expect the same to happen in all languages, though there may be differences of degree.

Phonetic effects are not the only ones which one finds in relaxed, connected speech: there are also language-specific reductions which occur in predictable environments and which appear to be controlled by cognitive mechanisms rather than by physical ones. These we term phonological reductions because they seem to be part of the linguistic plan of a particular language. While they may not make a change in meaning, they contribute to acceptable relaxed pronunciation. They help to make a native speaker sound native. Among these in English are effects such as changing /t/ to [?] before another consonant in syllable-final position, as in "hatbox" pronounced [hæ?bɒks].

Casual speech processes cause changes in everyday conversational speech which make some of the forms found quite different from their dictionary represen- tation or "citation form." They can, for example, cause ambiguities: the distinction between /n/ and /m/ is often not observed before bilabial consonants. This means that "screen play" and "scream play" are often not pronounced differently.

More extreme differences are possible: the word "handbag" is often pronounced "hambag" The /d/ is deleted or suppressed, and the /n/ which remains changes to match the following /b/, as in the example above. The word "can't" is often pronounced [kɒ?], without the [n], and with the final /t/ changed to a glottal stop

Phonological effects are common in casual speech, but some models of speech perception (e.g. [2]) assume fully-specified input which is processed in a linear order: there are no segments absent from the signal (though overlap of gestures can occur), nor are there any segments which are not present in the phonemic inventory of the language, but which appear as the result of phonological processes, such as the nasalised [ɒ] in [kɒ̃?].

Some researchers [6,7] have begun to explore the changes which will have to be made in lexical access models in order to accommodate phonological variation, and this paper is a contribution to that explorationl.

### Sequential lexical access

It is believed that "Listeners generally recognise words before hearing them completely," [9] . A special case is made for homonyms, which have to be disambiguated by following information [1].

But personal experience tells us that it is quite possible to revise our notion of what was heard based on subsequent information, especially when we are listening under unfelicitous conditions, e.g. to a foreign language with which we are only adequately familiar, to our own language in a noisy environment or even to gated sentences. Experimental work by Grosjean [3] and Bard et al [1] supports this intuition.

We hope to do a series of studies aimed at finding out how ambiguities caused by phonological reduction are resolved by listeners, and how, in general, reduced forms are related to the fully-articulated forms which (presumably) constitute entries in the mental lexicon. We assume that the scope of material used to unravel these reductions varies with the degree of reduction: as the phonetic information becomes less dependable, more semantic information is called for. We also suspect that subjects vary a great deal in the extent to which they depend on one or the other of these sources.

## THE PILOT EXPERIMENT

The following sentence was produced by the experimenter and recorded digitally:

*The screen play didn't resemble the book at all.*

The word "screen" was pronounced with a final **m**, as is normally done in this environment in unself-conscious speech.

The sentence was then presented to 10 subjects using a gating technique. There were 33 equal gates, beginning in the middle of the vowel of "screen." All of the subjects originally judged the first word in the sentence to be "scream." When the segment (p) which is the motivation for the /n/ to /m/ change was presented, three of the subjects reversed their judgement from /m/ to /n/. Even though this is subsequent to the end of the word, it is easily explained by extremely local factors: the notion of 'underspecification' [7] could explain this result. Eight subjects had changed [m] to /n/ by the end of the word "play," but two withheld judgement on the /m/ or /n/ decision until clearer, non-phonological, information was available from other sources.

The word "didn't" was even more reduced, (to [dĩn]) and here there was even greater evidence of late recognition: six opted for "didn't" at various stages in the word "resemble," and two subjects couldn't commit themselves until after "book" was recognised.

It is thus clear that here the identity of the lexical item was not resolved before its end. Further, it seems that it was not resolved purely through phonological knowledge, though an implicit knowledge of the sorts of reductions permissible in English was called for.

A similar experiment using the sentence:

*The scream play was part of primal therapy*

showed that many subjects perform an [m] to /n/ transformation on this case also, even though it is inappropriate, and that this transformation must be reversed by subsequent semantic input for the sentence to be understood. This suggests both that the influence of the phonetic conditioning factor is very strong and that reversal of a decision based on subsequent information must be a robust part of our linguistic competence. As Warren [10] has it, "...succesful speech perception cannot proceed as a Markovian process, with perception occurring first on lower and then higher levels of organisation. Processing of this nature does not benefit fully from the redundancy of the message and does not permit the correction of mistakes."

This pilot experiment will be followed by others which explore whether other conversational speech processes (such as tapping, palatalisation, and devoicing) are perceived similarly and the consequences for models of word recognition.

## Bibliography

[1] Bard, E.G., Shillcock, R., and Altmann, G.T.M., (1988) "The Recognition of Words After Their Acoustic Offsets in Spontaneous Speech: Effects of Subsequent Context," *Perception and Psychophysics* 44, pp. 395 - 408

[2] Cole, R. and Jakimik, V., (1979) "A Model of Speech Perception" in R. Cole (Ed.), *Perception and Production of Fluent Speech,* Erlbaum

[3] Grosjean, François, (1985) "The Recognition of Words After Their Acoustic Offset: Evidence and Implications," *Perception and Psychophysics* 38(4), pp. 299 - 310

[4] Fowler, C.A. and Housum, J, (1987), "Talkers' Signalling of "New and "Old" Words in Speech.....," *Journal of Memory and Language* 26, pp. 489 - 504

[5] Frauenfelder, U.H. and Tyler, L.K., (1987) "The Process of Spoken Word Recognition: an Introduction," *Cognition* 25, pp. 1 - 20

[6] Frauenfelder, U.H. and Lahiri, A., (1989), "Understanding Words and Word Recognition: Does Phonology Help?" Chapter 10 in W. Marslen-Wilson (ed), *Lexical Representation and Process,* MIT Press

[7] Lahiri and Marslen-Wilson, (1991), "The Mental Representation of Lexical Form: A Phonological Approach to the Recognition Lexicon" *Cognition* 38: pp. 245 - 294

[8] Lieberman, P., (1963) "Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech," *Language and Speech* 6, pp. 172 - 187

[9] Marslen-Wilson, W.D., (1987) "Functional Parallelism in Spoken Word Recognition," *Cognition 25,* 1987, 71 - 102

[10] Warren, Richard, (1982), *Auditory Perception, A New Synthesis,* Pergamon