



Discovering the acoustic correlates of phonological contrasts

John Coleman

Phonetics Laboratory, University of Oxford, U.K.
john.coleman@phonetics.oxford.ac.uk

Recently, a number of speech researchers have argued that words are stored in the mental lexicon in the form of numerous, detailed, variant exemplars. This is offered as an alternative to the more conventional view that variant tokens of the pronunciation of a word must undergo some sort of normalization to a single, clean, prototypical, usually symbolic/phonemic representation prior to lexical access. Phonemic classification is, nevertheless, an ability of many speakers (especially alphabetic literates). Consequently, those of us who argue for an exemplar-based view of the lexicon must also demonstrate how phonological contrasts could be derived or inferred on the basis of multiple tokens of words, stored as detailed phonetic representations. In this paper I shall describe a signal-processing method that more-or-less does just that.

Additionally, it has become evident in recent years that many subtle aspects of phonological contrasts remain to be properly examined, and more may remain to be discovered. These 'subtle aspects' are small but systematic differences in the phonetic realisation of contrasts located some distance away from the vowel or consonant thought (in segmental phonology) to be the principal signifier of the contrast. The discovery of the non-local phonetic differences was just serendipitous. This is worrying: what other surprises may lie in store, when we re-examine other contrasts with equal thoroughness? To address this concern, a database of citation pronunciations of English words, each spoken in five repetitions by a single speaker, was 'mined' for further data on non-local correlates of every phonemic contrast in English, in an attempt to discover any further examples of such long-distance effects.

The essence of the algorithm is this: given five tokens of one word (e.g. *pit*) and five tokens of another (e.g. *bit*), determine for which points in time and for which acoustic parameters the two groups of five are significantly different. With five tokens of each word in hand, it is easy to calculate the mean and standard error of each parameter at each time frame. We then determine the 95% confidence intervals of the distributions of the five data points taken from each word for each parameter: where the confidence intervals from the two words are disjoint, we are almost certain that there is a significant acoustic difference between the words. However, for this method to work at all, it is first necessary to precisely align in time all ten tokens of a given word pair. Therefore, the parameters of the ten tokens in a given word pair were each warped against all the others, using dynamic time warping software.

Most of the correlates of contrast investigated this way were local contrasts, consistent with prior studies of speech acoustics. For example, *grab* and *grub* were found to differ in their vowels, *hair* and *share* in their initial consonants, and *tap* and *tab* in their final consonants. In addition, numerous examples of coarticulatory variation in immediately preceding and following segments were found. The most interesting results concerned longer-distance coarticulatory correlates of two features, [voice] and [anterior]. In many word pairs contrasting in so-called "final consonant voicing" (e.g. *sheath/sheathe*, *pup/pub*, *newt/nude*, and *smelt/smelled*), extensive anticipatory coarticulation was found as early as the word-initial consonants, and often in the preceding vowel, i.e. at the end of the preceding word. Regarding [anterior], many instances were found in which acoustic differences were found in the /t/ of the word *utter*, immediately preceding a contrast such as *said/shed*, or

dear/jeer. This phenomenon is very similar to descriptions of [anterior] consonant harmony (sometimes called “sibilant harmony”) in the indigenous American languages Chumash, Tahltan, Navajo and Chiricahua and the African language Zayse. Naturally, it is a little surprising to find a similar phenomenon in English, and shows up one of the inherent limitations of the segmental view of speech.