**EECS225d Audio Signal Processing in Humans and Machines** **Spring 1995**

Lecture 8: February 3, 1995

*Lecturer: Professor John Ohala* *Scribe: Andrew Fung*

## 8.1 Examples of phonetic variation

We assume, or create, here a phonetic alphabet in which different sounds convey different meanings, i.e., one symbol is equivalent to one and the same sound. The *International Phonetic Alphabet* (IPA) is such a script. Scripts that are familiar to us, such as standard English or French, are not such languages because, for example, the character *h* represents a different sound in the words *house* and *shirt*.

Following are several examples of phonetic variation:

- *That house has a green hue.* Note the different sounds of the letter *h* despite the single orthographic representation. This is caused by the *y*-like sound following the *h* in the word *hue*.
- *The lamp fell.* Here, similarly, the *l* represents a different sound at the beginning of the word *lamp* and at the end of the word *fell*. The latter *l* features added constriction at the back of the mouth.
- *The sky is calm. This guy is calm.* Since it is difficult to differentiate between these two sentences when said aloud, the use of a script such as the IPA can not tell the difference between the sentences. By using a script where one sound corresponds to one symbol, these sentences can be represented using the exact same IPA symbols.
- *Eugene met a tiger on the mountain trail.* In this example, the different pronunciations of the English letter *t* are quite obvious. The *t* in the phrase *met a* becomes a tap, and the two words are said almost blurred together. The *t* in *tiger* is aspirated, as opposed to the same symbol in *trail* which is affricated.

We can note, however, that many of these differences in pronunciation are minor, and that these differences do not create new words; instead, changes such as saying an aspirated *t* where it should be affricated come across as the accents of a non-native speaker of a given language. However, as native speakers, we can usually decipher the code using predictable contextual variance.

## 8.2 Analyzing word spectra

Some attempts at speech recognition have featured analysis of the spectral content of words. This analysis can be described as a game of predicting contextualized variation with the ear. This type of analysis is an extremely difficult task because there is an infinite set of possible variants.

## 8.3 The vocal tract

The vocal tract is comprised of articulators which, in different positions, create different sounds. These articulators include the lips, teeth, alveolar ridge, nasal cavity, hard and soft palates, uvula, tongue, pharynx, and larynx.

¿From front to back, the possible methods of articulation are:

- *labial* : two lips together, as in /b/
- *labio-dental* : teeth to lips, as in /f/
- *(inter)dental* : teeth, as in /sh/
- *alveolar* : tongue to alveolar ridge, as in /t/
- *retroflex* : tongue flexed behind alveolar ridge, as in /r/
- *palatal* : tongue to soft/hard palate, as in /ch/
- *velar* : tongue against velum, as in /g/
- *glottal* : closure in glottis, as in the phrase *uh-uh*

Different phonemes can be classified by the amount of air flow through different parts of the vocal tract. The following list gives a continuum, from the complete stoppage of air in stops to the completely free air flow in vowels:

- Stops
- Affricatives
- Fricatives
- Nasals
- "Liquids"
- Glides
- Diphthongs
- Vowels

These different classes combine to create about 44 distinctive phonemes. This wide pool of possibilities, especially the large number of vowels (14 or 15), poses one of the greatest problems in word recognition. The other major problem is the naming of letters, i.e., different phonemes may actually correspond to the same English alphabet letter.

## 8.4 Do Phonemes Exist?

Con:

- Phonemes are not found in the physical signal.
- The concept is biased by an orthographic tradition of physical writing of alphabets.

Pro:

- Native speakers' speech errors suggest a unit like the phoneme exists. For example, the error made between the phrases *keep a tape* and *teep a cape* is merely a switch in a unit like the phoneme.
- Native speakers can manipulate something like phonemes in word games, poetry. For example, the use of alliteration (the repetition of a consonant at the beginning of adjoining words) in *sadly sang the sorrowful song* is pleasing because of the repeated use of a phoneme. Similarly, assonance (the use of the same vowel in adjoining syllables) is used by many companies for their trade names, as in *One Touch* or *Fed Ex*.
- Phonological rules would be near impossible to state without the use of phonemes. For example, the rule for plurals specifies an *s* phoneme after voiceless consonants, a "schwa" followed by an *s* used in *faces* after sibilants and affricates, and a *z* phoneme elsewhere.

Resolution: Phonemes are psychological units, the units of lexical representation; these are encoded (and decoded) in other units (drawn from an inventory with a different size) than phonemes.