INTRODUCTION TO ACOUSTIC PHONETICS 4

Hilary Term, week 8
8 March 2006

1. Acoustic cues for perception of sounds

Acoustic phonetics studies, amongst other things, what aspects of speech signal are important for perception of speech by listeners, i.e. it seeks to discover acoustic correlates of perceived speech, **acoustic cues**.

Acoustic cues represent relations between physical quantities, not absolute values. E.g., vowel quality is cued by relations between formant frequencies, not some absolute frequency value.

Acoustic cues are always used in combination. Weight assigned to a particular cue for a particular feature varies depending on context and situation. Also members of the same language community do not necessarily use the same cues for a given distinction.

Investigation of acoustic cues involves finding out which aspects of acoustic signal the brain uses to recognize and categorize particular sound. This can be done in several ways:

1) Study of spectrograms. It can demonstrate that certain feature is present in particular sound and that it is found whenever the sound occurs again in the same context. However it cannot prove whether listeners actually use this feature as a cue.

2) Isolating and controlling a possible cue. Features are controlled and modified using speech synthesis.

3) Mapping perceptual space on the basis of confusion matrices. It elucidates correlations between dimensions of the map and acoustic properties that are suspect cues. Advantage of this method is that it is based on naturally produced speech.

2.1. Voiced/voiceless distinction

Acoustic cues for voicing feature vary considerably depending on the context.

1) Main cue for *prevocalic* consonants is **voice onset time (VOT)** – time interval between release burst and start of the vocal fold vibration.

2) Consonant sounds in *intervocalic* position is the only context in English when the distinction is actually cued by the **presence/absence of the vocal fold vibration** – in voiced consonants the vibration continues throughout the articulation while in the voiceless consonants it halts for a period of time.

3) Voiced/voiceless distinction in *postvocalic* consonants is cued by the **duration** of the preceding vowel. Vowels are significantly longer before voiced consonants.

Another cue for the voiced/voiceless distinction that holds for all contexts is **relative intensity**. All voiceless consonants have greater intensity as compared to their voiced counterparts.

There are various secondary cues for perception of voicing. Lisker (1986) summarized all potential cues for voicing distinction for English stop consonants in trochees (see table below).

During consonant	Duration of closure
	Duration of glottal signal
	Intensity of glottal signal
Before consonant	Duration of vowel
	Duration of first-formant (F_1) transition
	F_1 offset frequency
	F_1 transition offset time
	Timing of voice offset
	Fundamental frequency (f_0 contour)
	Decay time of signal
After consonant	Release burst intensity
	Timing of voice onset (VOT)
	Onset of F_1 transition
	F_1 onset frequency
	F_1 transition duration
	f_0 contour

2.2. Vowel differentiation

The principle cue for vowel quality is **formant pattern**, i.e. relations between its formant frequencies. Relation between F_1 and F_2 frequencies is especially important. Large separation between F_1 and F_2 and narrow divide between F_2 and F_3 cue front vowels. Conversely, back vowels have small difference between F_1 and F_2 and large difference between F_2 and F_3. Central vowels show a uniform formant pattern, i.e. formant frequencies are equally spaced. Close vowels have high F_1 frequency while open vowels have low F_1 frequency.

Another cue employed in many languages is relative **duration** that helps to distinguish between short and long vowels (including diphthongs).

Intensity differences can also serve as a secondary cue for differentiating between open and closed vowels, especially in situations when there is a high level of masking noise. Total range of variation is only about 7dB but it still contributes to vowel recognition.

2.3. Place of articulation

Two major cues for place of articulation in consonants are **formant transitions/loci** and **noise spectrum**. Nasals also have additional cue – location of **antiformants**.

2.3.1. Formant transitions/loci as a cue for place of articulation

All formant transitions carry important information that can be used as cue for place of articulation. However it was found that F_2 transition/locus is particularly important for differentiation of consonants. **Locus** is a hypothetical intersection point derived by superimposing trajectories of formant transition from consonants in different vowel contexts, e.g., [di], [de], [da], [do], [du].

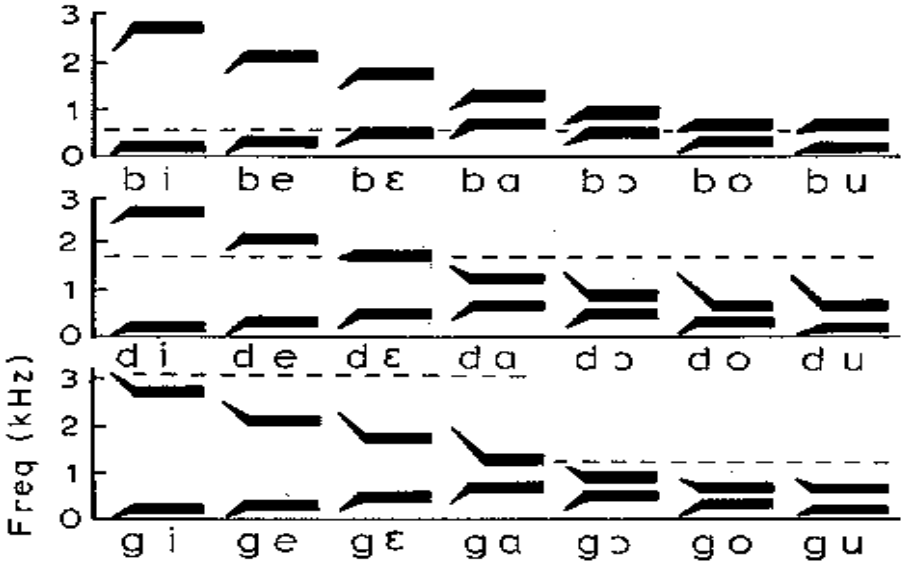

Fig. 1. F_1 and F_2 transition patterns used to synthesize [b], [d], and [g] followed by various vowels

F_1 increases from near-zero value to F_1 frequency of following vowel for all types of consonants.

F_2 rises for labials and falls for velars. Transitions for alveolars can be rising (front vowels), flat (mid vowels) or falling (back vowels) depending on following vowel. Similarly F_2 locus frequency is low for labials, high for velars and mid range for alveolars.

F_3 is rising for labials and velars and falling for alveolars, providing an acoustic reason for grouping them into a natural class.

2.3.2. Noise spectrum as a cue for place of articulation

Spectral characteristics of noise depend on location of noise generator and therefore serve as cue for this location. Labials have diffuse flat or falling spectrum with peak in lower frequencies. Alveolars are characterized by diffuse rising spectrum with high-frequency peak. Velars have compact spectrum with peak in mid-frequency region.

2.4. Manner of articulation differences

Period of **silence** (or near-silence) ranging from 40 to 120 ms cues stops and affricates. These two classes are distinguished by **duration of frication noise** following burst release noise – it is longer in affricates.

Fricatives are marked by **presence of turbulent noise** of appreciable duration ranging from 70 to 140 ms on average.

Nasals, liquids and approximants are distinguished from other consonants by their **periodic source** parameter and absence of noise. Nasals and liquids are cued by very **low F_1**, **weak formants** and presence of **antiformants**. **Low F_3** is cue for trills and taps. Approximants have vowel-like formant frequencies with **rapid transitions**. They do not differ drastically in their intensity from neighbouring vowels, as is the case with nasals and liquids.

2.5. Tone and intonation

Main acoustic cue for tone/intonation perception is $\boldsymbol{f_0}$ but **amplitude contours** may also be important.

2.6. Stress

Perception of stress is linked to three acoustic parameters: intensity, duration and f_0. But no one-to-one correspondence between any single acoustic feature and stress exists. To complicate matters even further, various languages use opposite acoustic relations for stress marking. For example, in English stress is cued by higher tone while in Hindi it is cued by lower tone.

3. Non-auditory aspects of speech perception

Perception is not limited to audition, perceptual representations are product of auditory and visual input. McGurk effect (called after Harry McGurk who first demonstrated it) is perceptual illusion when combination of audio stimulus [ba] and video stimulus [ga] is perceived as [da]. It happens because visual experience of watching talking face also activates primary auditory cortex, showing that visual information may be used in speech understanding and is integrated with auditory speech at a very early stage in processing.

Reading:

Cooper, F.S., Liberman, A.M., and Borst, J.M. (1951) The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Science*, **37**, pp. 318-325.

Fry, D.B. (1979) *The physics of speech*. Cambridge: Cambridge University Press (chapter 11).

Johnson, K. (2003) *Acoustic and auditory phonetics*. 2nd edition. Oxford: Blackwell (chapter 3-4).

Kent, R. D., Dembowski, J., and Lass, N. J. (1996) The acoustic characteristics of American English. In N. J. Lass (ed.), *Principles of experimental phonetics*, pp. 185-225.

Lisker, L. (1986.) “Voicing” in English: a catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech* **29** (1), pp 3-11.

McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, pp. 746-748.