



Phonetics for Speech Technology

- Jack Godfrey
- Texas Instruments, Multimedia Technologies Laboratory



Outline

- Motivation
- Terminology, principles and methods of representation
- Units, features, cues
- Special problems of spontaneous speech
- Practicum: Switchboard transcripts and lexicon



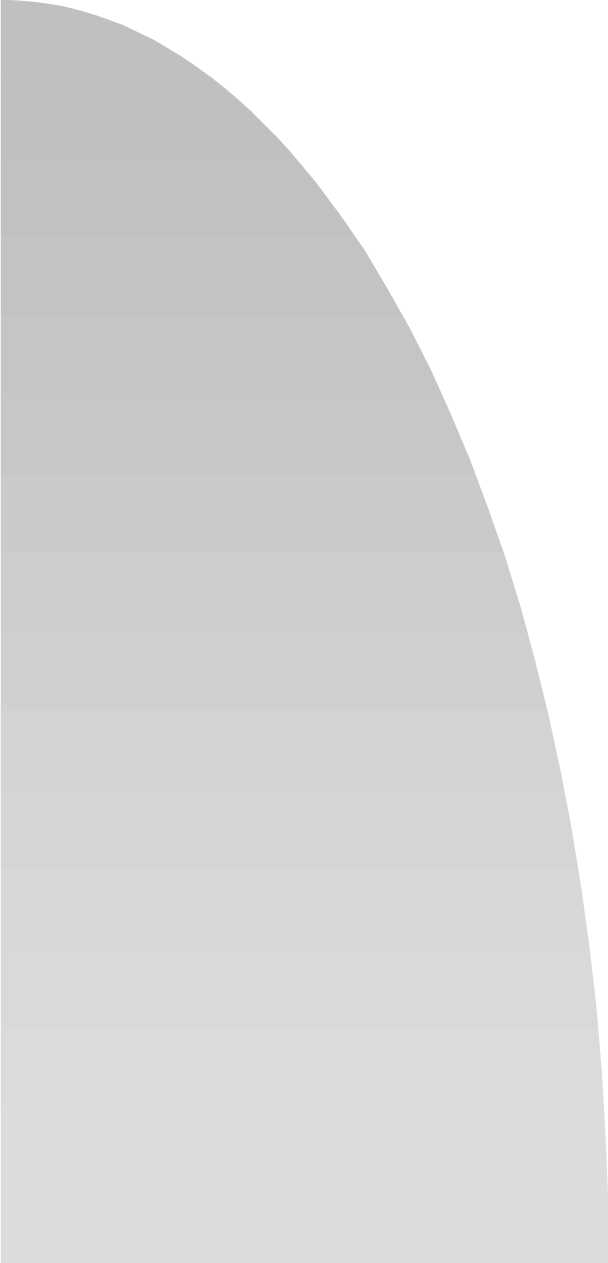
Motivation

- Linguistics and Phonetics: definition and scope
- Subfields: definition and scope
- Related fields
- Phonetics in Speech Recognition and Synthesis



Key Ideas in the Development of Phonetics

- Phonemes, Distinctive Features
(Troubetzkoy, Jakobson, Martinet)
(Linguistics, 1930s)
- Source-Filter, Formants & VT shapes
(Fant, Peterson, Stevens, Klatt)
(Acoustics, 1940s-60s)
- Special Processing, Locus Theory
(Haskins Lab) (Psychology, 1960s-80s)
- Prosody (Pierrehumbert, Liberman, Ladd) (Linguistics, 1980s)
- Articulatory “reorganization” (Haskins: Fowler, Goldstein) (Physiology, 1990s)



The “Problem” of Invariance

- Intuition: a simple cipher, n sounds \leftrightarrow n symbols
- 50 years experience: speech is a complex “code on the phonemes”
- Understand the code to constrain the decoder
- Jakobson: “We speak in order to be heard in order to be understood”
- For humans, the lack of invariance is not a problem but a solution



Inductive Approach: transcribe a little, think a little

- At least two types of transcription: phonemic and phonetic
- Phonetic: Broad <--> Narrow
- Phonemic: one distinctive segment, one symbol
- Phonetic: one relevant detail, one symbol/diacritic
- Roughly, broad phonetic = allophonic



•Examples

- Use the ISIP Phone Inventory or any other system
- Listen and transcribe



Allophones: representing regular variation

- Allophones: predictable variants of a phoneme
 - ◆ In complementary distribution
 - ◆ e.g., aspiration of E word-initial stops
 - ◆ or in free variation
 - ◆ e.g., release of English word-final stops
- Differ from each other in one or more discrete, usually *redundant* features
- Rules bring variation to light



• A few simple examples

- ◆ light/dark // in English (vs. Russian) [*velarization*]
- ◆ variants of /t/ in English [*glottalization; aspiration; closure*]
 - ◆ tie, sty, strive, item, enter, site, settle, ten -- all different
- ◆ vowel duration before final C [length]
 - ◆ wide, white; rise, rice
- ◆ German ach, ich; but not E [h], [ng] (complementary but unrelated)



•More on allophones

- Allophone: a *variant* of a phoneme *predictable by rule* (Ladefoged)
- May or may not be obviously “motivated” (e.g., by ease of articulation)
- Implies separate **targets**, as opposed to
 - ◆ simple contextual influence (mutual adjustments)
 - ◆ coarticulation (see the specific definition below)
- But there is no bright line between them (eg. aspirated sonorants in **clash, crash**)



•Other kinds of variation

- Contextual influence: targets modified to accommodate others
 - ◆ keen, kin, ken; asp. sonorants (play, crash)? .
- Assimilation: a redundant feature change to match a neighbor
 - ◆ dental n in tenth; labiodental m in emphasis”
- Coarticulation: simultaneous operation of articulators for different segments
 - ◆ (rounding before /u.o/ --> screw, spoon
 - ◆ nasalization before /n, m, ng/ in camp, can't, pined



•Linguistic and nonlinguistic factors

- Prosodic status: stressed; accented; phrase-final
- Discourse status: focus, emphasis, contrast
- Dialect (aa/ao; V+r: flap)
- Pragmatics: style; information load; rate
- Physiological: sex; VT size
- Environment and channel (Lombard effect)



Some experimental evidence

- The effect of prosodic position and style on vowel duration: House and Fairbanks (1953) vs. Crystal and House (1988?)
- The effect of vocal tract shape and dialect on vowel cues: (Ladefoged and Broadbent, 1956)



The transcription examples revisited

- Try to enumerate some of the allophones and detect other variants
- Try changing style, rate, prosody
- Add some symbols for phonetic detail



Terminology: first cut

- Phoneme: minimal segment that can distinguish meaning
- Allophone: variant of phoneme, predictable by rule, complementary distribution
- Phonemic transcription: one symbol per choice (of segment)
- Broad and Narrow Phonetic transcription: one symbol per *relevant* detail
- Coarticulation (narrow definition)
 - ◆ Left to right vs. right to left



Phonology and Distinctive Features

- Distinctive feature: a minimal property of speech sounds, by which to distinguish word meanings in a language
- Property: a category of sound classification in a language
- Formal proof: find a phonological rule *requiring* only the proposed feature



Examples

- Find the rule for forming English plurals -
-> discover the sibilants and voicing
- Sibilants as a class (s, z, sh, zh) ->
property “sibilance” = a distinctive feature
- Find the rule for forming English past
tense --> discover alveolar + stop,
voicing
- Find the rule for “vowels before r” -->
discover feature “tense”



Distinctive Features: continued

- Rules act on related objects (“natural classes”)
- Features are thus based in reality: auditory and/or articulatory cues
- Are the cues invariant? only relationally
 - ◆ Relative to choices at this position
 - ◆ bilabial, labiodental --> labial
- What about control parameters?
Ask a neuroscientist...



Other implications

- DFs useful for looking at cross-language generalizations
- New insights: autosegmental and metrical frameworks, syllable-based phenomena
- Theory is far ahead of empirical data
- Feature-cue relationship may be complex, abstract
- ADTs to evaluate feature systems?



New definition of phoneme

- **Simultaneous bundle of distinctive features**
- Whether binary or n-ary is not clear
- Each feature has one or more correlates (cues), *relative to the current configuration*
- Cues need not be simultaneous in time
- Examples: voicing; nasality; tones



Examples of DF systems

- Ladefoged: emphasis on physical more than formal
 - ◆ L's scheme of % values for correlates
- Halle et al.: formal, binary; theory-driven



Transcription: Pros

- Phonemic transcription needed because of English orthography
- Allophonic or broad phonetic transcription: catalogs high level effects such as deletion of consonants, whispered and reduced vowels
- Points up the frequency of dropped segments, which affects triphone training



Transcription: Pros (continued)

- Phonetic or narrow transcription of at least some data gives a sampling of effects to be seen in the training set: extremely short vowels, elisions, lenition of stops, reductions of common phrases, etc.
- ICSI examples



Transcription: Cons

- Can become a reconstruction (What I would have said)
- Requires extensive training to do, or to use effectively
- Work product tends not to be used by those who don't understand it.
- Lexicon must match transcription to be used in training
- No gold standard to check validity
- Conversational speech is *terra incognita*



Special problems of spontaneous speech

- Simplifications and Deletions
- Especially at word boundaries
- Especially on high-probability words (in context)
- Especially on unstressed syllables
- The “islands of reliability” approach
- Lesson: the Marslen-Wilson experiments



SWITCHBOARD practicum

- Select two passages of very different styles by listening
- Check orthographic transcription
- Make a phonemic transcription, as if for a standard dictionary
 - ◆ Note where this is “reconstruction”
- Make a broad phonetic transcription, identify major allophones
- Make a narrower transcription, noting effects of style and rate
- Discuss issues