A SPEECH ANNOTATION AND PHONOLOGICAL ANALYSIS PROGRAM

John Ingram, Jeffery Pittam & Robb Hay
University of Queensland

ABSTRACT - A software system for phonetic annotation and phonological analysis of speech samples is described for application to a longitudinal study of sound change in second language learning.

INTRODUCTION

This paper describes the development of a system for phonetic annotation and phonological analysis of speech samples and discusses methodological issues that arise in connection with its application to a longitudinal study of sound change in second language learning involving Vietnamese immigrants acquiring Australian English. In recent years, a variety of inexpensive microcomputer-based systems for speeech analysis have become available, which meet many of the signal processing requirements for descriptive phonetic and applied linguistic research. More or less contemporaneous with these developments, a variety of multi-tiered text manipulation programs for linguistic analysis have appeared (CHILDES, IT, KWIC-MAJIC, etc.), some of which are specifically designed for phonological applications (e.g.: Pye & Ingram, 1988).

However, for much applied phonetic and linguistic research, the easy manipulation and retrieval of speech signal files and the textual data-base manipulation of their corresponding annotations are equally important and complementary operations. The recoverability of relevant acoustic segments of the original signal files is required for establishing transcription reliability or for investigating signal parameters that correlate with the use of particular annotation symbols. Existing microcomputer-based speech analysis programs have limited annotation facilities that do not meet the needs of phonological analysis, and text-oriented linguistic programs do not permit recovery of the primary speech data. The system reported here provides an integrated environment for acoustic phonetic and phonological analysis of speech samples. In this, we believe, it is unique, at least for systems based upon a PC-DOS platform.

The system comprises:

(1) a speech file annotation and segmentation facility, which is part of a sound manipulation hardware/software package (Ultrasound: ULS).

(2) a data base facility for phonological analysis of annotation files (Ultrasound data base: UdB).

THE ANNOTATION FACILITY

Annotation of speech files is performed with a waveform editor, incorporating standard audio-playback and zoom-windowing

facilities. Three levels of annotation are provided for:

1. Segment annotations: Single case-sensitive ascii keyboard characters, used to encode I.P.A. symbols. These symbols may be associated with acoustic segments of the signal file.

2. Feature annotations: One to five character feature codes, which, for present purposes, are interpreted as phonetic or phonological process codes, representing connected speech processes or phonetic feature changes caused by second language phonological interference effects. Feature annotations are link-listed to segment annotations.

3. File annotations: A text line of annotation or commentary that is associated with the whole file.

Figure 1, which shows a screen dump of the annotation display and its corresponding pull-down help menu provides an idea of the functionality of the annotation editor.

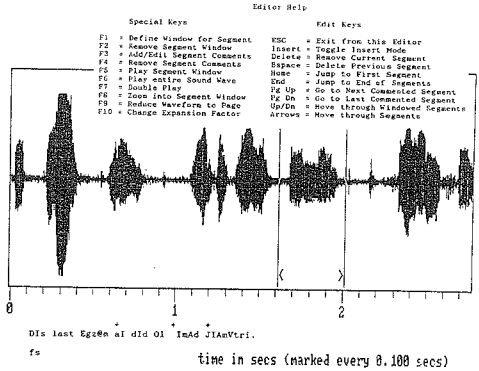

Figure 1: Screen dump of annotation editor and help menu

The annotations are saved as text files, with the same path and filename as their parent signal files, but with a uniform extension (.SEG). The SEG file annotations were originally intended to capture both the conventional **phonological form** of a word, and its **phonetic form** or pronunciation in a given instance or token. Basically, the phonological target of a word is conveyed by its conventional phonemic transcription in IPA symbols, in accordance with entries in the Macquarie dictionary. Phonetic forms are represented as featural annotations, augmenting or modifying the pronunciation indicated by the phonemic transcription.

A conventional phonemic representation corresponds to a broad phonetic transcription, or the approximate pronunciation of a word spoken carefully in isolation, according to generally accepted standards in the speech community. It may therefore be interpreted as an appropriate 'target' pronunciation for language learners and phonetic depatures from it, in terms of

phonetic feature counts, have been shown to provide a good metric of foreign accentedness (Brennan & Brennan, 1981).

Much of the phonetic variation of words in connected speech is attributable either to the word-internal contextual effects of 'allophonic' rules, or to connected speech processes (CSPs) which operate basically across word boundaries and which are globally influenced by speech rate and style. The phonetic effects of allophonic rules and CSPs can be economically represented as featural departures from, or additional annotations upon, phonemic segments. Furthermore, allophonic rules and CSP's are important components of the 'sound pattern' of a language and their presence (or, just as significantly, their omission) in speech will also signal the native/foreign status of a speaker.

Finally, although not by itself a persuasive consideration, a phonemic transcription also provides an economical code for representing the sound pattern of forms in the lexicon, and is therefore a suitable basis for their storage and retrieval. It was for these reasons that a two-tier annotation system was adopted, using broad phonetic transcription to capture the conventional phonological forms or 'targets' of words, augmented by featural annotations to capture the phonetic forms of tokens.

The phonetic process codes or feature annotations, shown in Table 1, reflect two kinds of things, (1) natural processes of phonetic simplification that one might expect to encounter in a variety of linguistic contexts or speech styles, and which have some claim to status in a universal taxonomy of natural phonetic processes, and (2), phonetic or phonological transfer effects or featural changes specific to Vietnamese (e.g.: [pvcd] prevoicing/implosion of voiced stops).

This annotation scheme is sufficiently flexible, but it raises problems for phonological analysis that are specific to language learning and second language learning in particular. Phonetic variation in the speech of second language learners can arise from multiple linguistic sources including, lexical or rule-based phonological interference effects and language-universal processes of phonetic simlification, as well as from extra-linguistic, performative considerations. It is the task of phonological analysis to attempt to isolate and describe the interaction of these various sources of phonetic variation. Success in the enterprise will depend partly upon the data-base tools available for the job.

THE DATABASE FACILITY

UdB provides, via pull-down menus, functions that are familiar to users of data base utilities, for file management, searching, and frequency tabulation. Segmentation files, which may be scattered over drives and directories, are gathered into data base files (identified by the extension .UDB) which are loaded into RAM and may be saved as text files. Previously saved .UDB files may be loaded or 'merged' (i.e.: augmented with additional .SEG files).

Changes common to all segments:

del	Deleted
epen	Epenthesis (added)
sub	Substitution of whole segment
wkn	Weakening of segment

Voicing feature changes:

dvcd	Devoiced	vcd	Voiced	lar	Laryngealized

Manner of Articulation changes:

Stops		Nasals	
asp	Aspirated	stpg	stopping
unasp	Unaspirated	wkn	weakening
affr	Affricated	Approximants	
spir	Spirantized	voc	Vocalized
glot	Glottalized	flp	Flapping of /r/
unrl	Unreleased	fric	Fricativized
ejec	Ejected	Fricatives	
pvcd	Prevoiced	affr	Affricated
flp	Flapping (t or d)	stpg	Stopping
		wkn	Weakening

Consonantal place of articulation changes:

lab	Labialized	dnt	Dentalized	alv	Alveolarized
pal	Palatalized	vel	Velarized	glot	Glottalized
lat	Lateralized				

Vowel feature changes:

rais	Raised	low	Lowered	frt	Fronted
bac	Backed	shrt	Shortened	lng	Lengthened
rnd	Rounded	mnth	Monophthongized	dpth	Diphthongized
rhot	Rhotacised	red	Reduced to schwa	unrd	Unreduced

Syllable changes:

syld	Syllable deletion
clrd	Consonant cluster reduction
resyl	Resyllabification: shift of C from coda to onset

SUPRASEGMENTAL FEATURE CHANGES

strss	Change of word stress
trm	Terminal contour: non-standard

Table 1: Phonetic Process codes/ Featural Annotations

Searches of the data base are controled by user-defined search formulae that specify linguistically significant search objects, such as a specific lexical item, particular phonemes, words begining or ending with specific phonemes or phoneme classes, segments containing certain feature annotations, or feature annotations in a certain segmental enviornment, etc.. Search formulae may be saved, retrieved for re-use, or edited. Search formulae are constructed using DOS-like variables (wildcards) and operators for and/or combinations of search elements.
Users may define macros for establishing classes of segments,

such as: {C} = consonant, {N} = nasal, etc.. This provides a shorthand for classes of segments commonly used in the specification of phonological environments.

Reports of searches can be customized and sent to standard output devices. Frequency displays are type/token tabulations or concordances produced from the output of a search. For example, one might wish to tabulate all the phonetic forms and their token frequencies associated with a specific word or a phonological environment within a word.

METHODOLOGICAL ISSUES:

It is important to distinguish between phonetic features arising from the application of CSPs in Australian English, and those which are attributable to first language (L1) phonological interference effects or developmental phonetic simplifications, that fall outside the range of acceptible usage, and which may be regarded as positive markers of 'foreign accent'. For ease of coding, it is tempting not to annotate normal connected speech processes and to code only those phonetic features which are 'non-standard' and which represent 'pronunciation errors'. This is the practice that was adopted for most of the initial analysis of the data.

However, a second language learner's failure to apply L2 allophonic rules or CSPs where expected is also significant for the perception of foreign accent. So for example, one of the most frequently encountered non-standard features of pronunciation of Vietnamese English is the retention of full vowel quality in unstressed environments, which is captured by the feature annotation [unrd]. Another example is the occassional failure to apply obligatory rules of allophonic variation, such as the aspiration of initial voiceless stops, which was captured by the feature [unasp]. Thus, allophonic rules and CSP's of Australian English were noted only insofar as their absence in expected environments may have potentially contributed to the perception of foreign accent. But their occurrences in expected environments were not recorded.

The decision to exclude from coding accent-neutral or expected sub-phonemic phonetic changes means that the feature annotated transcriptions constitute something less than full phonetic representations. While this simplifies the task of transcription, there is some sacrifice of phonetic detail which is relevant for the assessment of L2 phonological accommodation and may have implications for the intelligibility of the speech of some subjects. Learning the language-specific conditions of application of connected speech processes in the target language is an important component of L2 phonological accommodation. The younger speakers in our sample clearly show this accommodation over the period of the study. It is a prominent feature of their casual vernacular speech. Furthermore, as mentioned previously, broader varieties of AE employ more radical forms of phonetic reduction and elision (Ingram, 1989). When the effects of these acquired L2 CSP's are combined with residual L1 phonetic processes and developmental simplifications, the effect upon the intelligibility of the

speech can be quite deleterious.

To investigate the impact of coding for expected L2 connected speech processes, data from one speaker, H. a 14 year old female who showed rapid acquisiton of Australian English, while still preserving many features of Vietnamese pronunciation, was tagged at the featural level for the presence of expected L2 CSP's, as well as phonetic features that mark accented English. 'Foreign' phonetic features were distinguished from 'assimilated' features/processes by a trailing exclamation mark (e.g.: clrd! = non-standard cluster reduction).

The speech sample of 241 words in length, recorded approximately two years after entry to Australia contained 58 instances of non-native features and 65 instances of AE CSPs, which are summarized in Table 2. A high proportion (approx. 50%) of the AE CSPs involved phonetic reduction of high frequency function words, as would be the case for native English speakers.

Phonetic reduction of function words:
and(4/6) are(1/1) but(5/5) I(1/11) to(3/6) it(1/3)
of(1/1) you(1/3) cos(1/1) what(1/1) just(1/1)
Flapping of /t/:
probabili<u>t</u>y(1/1) star<u>t</u>ed(1/1) wha<u>t</u>ever(3/4)
Cluster reduction: Weakening of final nasals:
la<u>st</u>(3/3) [las] ni<u>n</u>e(1/1) [naĭ]
Deletion of unstressed vowel: Epenthetic [r]:
c<u>er</u>tificate(1/1) [stɪfəkət$^{?}$] yea<u>r</u> eight(1/1) [jiɹæɪt]
Vowel changes (young speakers):
m<u>a</u>ths(2/2) /æ/ -> [ɛ] [mɛs] exam(2/2) /ɛ/ -> [ɪ] [ɪksæ̃m]

Table 2: Aust. English CSPs in Vietnamese Australian
((x/y) = occurrences with feature/total no. occurrences)

In less fluent Vietnamese English speakers, many of these function words would either be omitted from the grammatical structure or appear in phonologically unreduced form. The appearance of the flap [ɾ] as an ambisyllabic form of /t/ seems to signal the attainment of an aspect of English-like syllable structure which many older learners never attain. A similar observation may be made about the appearance of the 'linking' or epenthetic [ɹ] (yea<u>r</u> eight). The deletion of [ə] in unstressed syllables ('c<u>er</u>tificate') is interesting not only as a sociolinguistic marker among Brisbane adolescents (Ingram, 1989), but because it conflicts with native Vietnamese phonotactic constraints. Other signs of accommodation to peer group speech may be seen in certain vowel changes.

REFERENCES:

Ingram J. (1989) Connected speech processes in Australian English. Aust. Journal of Linguistics 9, 21-49.