CONNECTED SPEECH PROCESSES AND CONNECTED SPEECH SYNTHESIS

J. Ingram

Department of English
University of Queensland

ABSTRACT Construction of a data base for the study of connected speech processes (CSP's) in Australian English is described. Application to the problem of speech rate, style, and sociolect sensitive synthesis is discussed.

INTRODUCTION

In recent years, there has been a convergence of interest on the part of Phoneticians and Sociolinguists in the study of connected speech processes (Kerswill, 1987; Dressler & Wodak, 1982). Connected speech processes (CSPs) refer to various phonological proceesses which abound in connected speech: assimilations, deletions, vowel reduction, consonantal lenitions, etc.. CSPs are responsible for a phonemic target such as:

/aɪ æm gouɪŋ tu/

being realized phonetically as:

[ð:ŋənə]

Connected speech processes have been called 'fast speech rules', or 'casual speech rules'. As knowledge of the acoustic structure of segmental contrasts has accumulated, Phonetics has been increasingly preoccupied with the the problem modeling phonological processes in connected speech. However, systematic empirical studies of CSP's based upon appropriate corpora of natural speech are lacking, despite the fact that such studies could provide a useful data base upon which to develop systems of connected speech synthesis by rule, capable of modeling accent, dialectal, and stylistic variation in speech.

The importance of CSP's for an account of sociolinguistic variation in Australian English is implicit in the 'classical' work of Mitchell, (1946) and Mitchell & Delbridge (1965), in which the prevalence of 'quite normal phonetic processes' of 'assimilation' and 'elision' were found to differentiate speech on the major 'broad', 'general', to 'cultivated' socio-stylistic continuum.

Kerswill (1987) provides an insightful discussion of the relationship between CSP's and the dimensions of rate, register, and speech variety (accent or dialect type), with exemplification from Durham (U.K.) English. He points out that although CSP's represent natural phonological processes that 'are likely to be explicable in terms of vocal tract characteristics and the motor control mechanism, as well as being influenced by speaking rate and articulatory care', they are also constrained in their application in language or dialect specific ways. Durham vernacular exhibits a number of natural CSP's which are not found in RP. Also, RP has some CSP's which are not found in Durham vernacular.

THE PRESENT STUDY

The data presented here is drawn from a larger corpus of material, The Brisbane Speech Survey, a study of sociolinguisitc variation in the speech of Brisbane schoolchildren. Adolescents were interviewed in self-selected friendship groups of three, at school, by one of the principal investigators or a project research assistant.

Participating schools in the wider survey, were selected to be sociologically representative of the Brisbane speech community, but for this analysis only two schools were chosen, to provide clearly contrastive groups of speakers and to enhance prospects of detecting sociolectal differences that may exist in the community as a whole. School A is a large, co-educational, private (fee collecting) school, of high prestige in the community, located in one of the most prosperous residential suburbs of Brisbane. School B is a large government (non-fee collecting) high school situated in a predominantly working-class suburb on the outskirts of the city.

Group interviews generated a good deal of recorded material which was then scanned for passages of sustained and uninterupted casual speech, where speakers were engaged by what they were saying and apparently unselfconscious of the recording situation. The basic unit of analysis chosen for the study of CSP's was what we dubbed the 'mini-discourse'. A 'mini-discourse' was defined as a passage of uninterupted monologue in which the speaker holds the floor for a period of time. Such passages were usually short narratives, relating an incident or illustrating a point the speaker wanted to make. They were usually bounded by turn taking. The mean length of a mini-discourse was 65 words and approximately 12 seconds in duration.

Once identified, a mini-discourse was dubbed on to a separate tape, transcribed orthographically, digitized for acoustic phonetic analysis, and phonologically coded for the presence of CSPs. A core of 24 mini discourses form the basis for the present analysis, selected so as to provide a balanced sample for the factors of School (A or B), Sex (M or F), and Grade (8 or 10).

METHOD

A combined methodology of acoustic-phonetic and phonological analysis was used to obtain a textual data base, coded for the presence of CSPs, which could then be interrogated either from the perspective of particular phonological processes, or the lexical items involved.

Each mini-discourse was segmented into convenient working units that usually, but not necessarily, corresponded to phonological phrases. These speech segments were digitized (8-bit quantization, 20 KHz sampling rate) using a PC based signal analysis system.

A textual data base was constructed, as a repository for phonetic/phonological observations on the occurrence of CSPs in the speech segment (phrase) currently under analysis. This was done using a multi-linear text editing and concordance program running on a second PC. The mini-discourses had been previously transcribed orthographically at the time of their selection. The program provides a multiple-line format for performing keyword-in-context searching and concordancing of a textual data base. The orthographic text line was supplemented by a line of phonemic transcription, using the transcription conventions of the Macquarie Dictionary, and by another line of CSP codes, based on a taxonomy of

connected speech processes devised by the author for the current data base and guided by previous research (Appendix).

The aim in coding the CSP's was, with certain qualifications, to provide an explicit annotation of the phonetic changes involved in a mapping between the phonemic representation of a word and its its phonetic realization, conventionally represented in terms of a narrow transcription. This goal was approximated more closely for consonants than vowels, and for segmental features than for suprasegmental features.

RESULTS

As expected, the vast majority of CSPs represent lenition processes of one kind or another. The data analysis has been concerned with identifying the linguistic environments associated with the application of particular CSPs, and also with the distribution of phonetic variants of particular lexical items which arise as a result of application of CSPs. The ability to interrogate the data base from the viewpoint of particular CSP's or from the viewpoint of particular lexical items, was the reason for using a multilinear text concordance program. This dual perspecitve on the data is important for modeling CSPs, because it is the high frequency words which are particularly prone to reduction in running speech. Thus, it is an open question whether the phonetic variation involved in speech rate effects and style shifting is best accounted for by a simple lexical variant look-up, or by incorporating segmental phonological rules, which is computationally more expensive, or, conceivably, by some combination of the two strategies. It seems clear, for example, that the observed phonetic variation associated with Vowel Reduction in the data base can only be accounted for by grammatical and pragmatic constraints applying to particular lexical items that are subject to vowel reduction. Vowel Reduction was confined almost exclusivly to function words, partly as a consequence of the coding procedures whereby unstressed vowels in lexical content words were coded as schwas, following the practice of the Macquarie Dictionary.

Excluding the definite and indefinite articles, which almost invariably occurred in reduced form, the 10 most frequent words manifesting vowel reduction or deletion were (in order of occurrence):

Word	Number of reduced forms	Proportion of total occurrence
and	151	.95
was	53	1.00
to	44	.80
it(s)	29	.76
you	29	.83
of	25	.89
that	19	.33
(be)cause	8	.50
were	7	.58
just	4	.19

Table 1. Ten most frequently reduced forms

It is clear from the above table that there is substantial variation in the incidence of vowel reduction for different words. Grammatical function appeared to be the important consideration in the case of 'that', for which the following grammatical roles were observed:

Role	Exemplification
Complementizer:	'I used to say that they hurt me feet..'
Relativizer:	'There's this lodge that..'
Demonstrative Pronoun:	'She doesn't tell anyone else that.'
Demonstrative Article:	'..that bloke there.'
Intensifier:	'not really that funny.'
Tag:	'They were throwing spears and that.'

As a subortinator, 'that' is highly prone to phonetic reduction, but as a pro-form or in its other grammatical roles, it may not be reduced, as the following table indicates:

Function	Reduced	Unreduced
Complementizer	12	1
Relativizer	7	4
Demonstrative & Other	0	33

Table 2. Grammatical Role and Phonetic Reducability of 'that'

Although the numbers are too small for statistical confidence, it seems that the relativizer offers more scope for stylistic variation in phonetic reducability than 'that' in its other grammatical roles.

A combination of pragmatic and grammatical considerations appear to be influential in the reducability of 'and', which was classified in terms of 4 degrees of phonetic reduction and the following grammatical features:

Role/Environment	Degree of Phonetic Reduction			
	[n]	[ən]	[æn]	[ænd]
Independent clause	25	35	23	8
Dependent clause	10	3	2	0
Phrasal	12	14	4	0
Unclassified	4	5	4	1

Table 3. Grammatical role and Reducability of 'and'

Table 3 shows a tendency for the degree of reduction of 'and' to be inversely related to the rank of the constituents being conjoined. All phrasal conjunctions showed some degree of phonetic reduction. However, there remains a substantial amount of variability unaccounted for by considerations of grammatical rank. Of the 9 observed cases of completely unreduced 'and' ([ænd]), 5 were found to occur in the environment 'and um..' where the slow rate of speech and the pause filler indicated that the speaker was planning what to say next. Hence, pragmatic factors that affect speech rate, appear to influence phonetic reducability in some cases.

Interaction of CSPs

Another aspect of the data relevant for the problem of modeling CSP's concerns the way that they interact to enable or disable one anther's application. For example, weakening of /t/ may occur in several ways, depending on phonological environment. Flapping is to be expected intervocalically, particularly when the /t/ is preceded by a stressed vowel and followed by an unstressed vowel or liquid. Glottalization of /t/ is most likely to occur before an alveolar nasal. Phrase final position favours unreleased /t/.
Flapping is enabled by /h/-dropping and disabled by schwa deletion, which

is, in turn, enabled by alveolarization of /ŋ/ as the following attested examples show:

'that he'	'sitting'	'sitting'
h deletion		alveolarization /ŋ/
flapping	flapping	
[ðəɾi]	[sɪɾɪŋ]	[sɪʔn]
		schwa deletion
		glottalization /t/

These kinds of interaction are usually handled by postulating rule ordering in formal phonological models.

Social class and sex differences

A test was made of differences in the incidence of CSP's from the speakers in the two schools. Significant differences in the expected direction were observed for Alveolarization of /ŋ/ (the ING variable) and /h/ deletion, two well attested markers of sociolectal differentiation. The principal finding was that the WC adolescents appear to permit a greater degree of phonetic reduction in casual speech. There were were no significant differences between the groups on the milder forms of reduction, but the WC group seemed more prepared to permit more radical forms of lenition that resulted in the loss, rather than the mere weakening of segmental features.

When the mini-discourses were partitioned by sex of the speaker, only marginal differences in the incidence of CSPs were found.

PROSPECTS AND LIMITATIONS

It is hoped that kind of detailed information on the distribution of CSP's in running speech will prove useful in the development of a speech synthesis system capable of accomodating speech rate, style, and sociolectal variation. However, there are certain limitations in the data base as it stands and certain design problems to be overcome before the knowldege gained from this type of empirical investigation can be incorporated into a synthesis by rule system.

There is no coding of prosodic features, except insofar as they are indirectly reflected in segmental features such as vowel reduction. This may prove to be a serious limitation bcause prosodic boundaries are known to block the application of a number of CSP's.

The annotation system is discrete and symbolic, whereas the phonetic effects that CSP's induce are, for the most part, graded and most directly expressed in parametric terms. The feature-style notation used to capture CSP's has greater flexibility than traditional narrow phonetic transcription. The task of providing acoustic specifications for particular CSP codes is not far advanced. The main use of the digitized signal files to date has been to facilitate auditory judgements and provide spectrographic or waveform confirmation/disconfirmation of perceptual features in difficult cases. While further work obviously needs to be done on this aspect of the problem, there is a strategic advantage to not committing oneself to a particular parametric representation of the speech signal, but instead, of expressing regularities of speech rate, style, and sociolectal variation in a symbolic form that could be translated into a variety of parametric representations.

REFERENCES

DRESSLER, W. & WODAK, R. (1982) Sociophonological methods in the study of sociolinguistic variation in the study of Viennese German, Language in Society 11, 339-370.

KERSWILL, P.E. (1987) Levels of linguistic variation in Durham, Journal of Linguistics. 23, 25-49.

APPENDIX

Classification of Observed CSP's

1. Lenitions, Reductions and Deletions:
   - Vowels:
     - 1.1 Vowel reduction
     - 1.2 Schwa deletion
     - 1.3 Monophthongization (/aɪ/,/aʊ/)
   - Syllables:
     - 1.4 Unstressed syllable deletion
     - 1.5 Consonant cluster reduction
   - Stops:
     - 1.6 Flapping of /t/
     - 1.7 Glottalization of /t/
     - 1.8 Unreleased /t/
     - 1.9 Deletion of /t/
     - 1.10 Unaspirated /t/
   - Dentals and Non-sibilant fricatives:
     - 1.11 Weakening of /ð/
     - 1.12 /ð/ deletion
     - 1.13 Weakening of /v/
     - 1.14 Deletion of /v/
     - 1.15 Deletion of /h/
   - Sonorants:
     - 1.16 Vocalization of /l/
     - 1.17 Glide deletion

2. Assimilations:
   - 2.1 Final obstruent devoicing
   - 2.2 Approximant devoicing
   - 2.3 Palatalization
   - 2.4 Stop assimilation

3. Linking Processes:
   - 3.1 Epenthesis of /r/
   - 3.2 Epenthesis of glide
   - 3.3 Epenthesis of stop

4. Strengthening Processes:
   - 4.1 Aspiration of stops

5. Phonetically unmotivated/obscure Processes:
   - 5.1 Alveolarization of /ŋ/
   - 5.2 /ʌ/ to /ɪ/ in 'just'
   - 5.3 /aɪ/ to /i/ in 'my'