Proceedings of The Institute of Acoustics

DURATIONAL CORRELATES OF SYLLABLE QUANTITY PATTERNS

M.R. Smith

Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, U.K.

INTRODUCTION

Abercrombie [1] reported a phenomenal difference of polysyllables among regional varieties of Standard English, specifically that syllable quantity patterns for a particular lexical item may serve as dialect markers. He noted that the word "Peter" has a long-short pattern of syllable lengths in Yorkshire, a short-long pattern in Lowland Scots and possibly a pattern of two equal lengths in RP. Such a claim has implications for a study of phonetic regularities that may be exploited in machine recognition of speech. The report is also relevant to the efficient use of phonological regularities associated with dialects, if syllable quantity patterns are reliable specifiers of (some) dialects. Further, an understanding of the acoustic support for the phenomenal differences may help us design more efficient acoustic parsers.

EXPLOITING SYLLABLE QUANTITY PATTERNS

The successful exploitation of this phenomenon requires solving problems of two types, one linguistic and perceptual and the other acoustic. It is a problem in systematic phonetics to establish the type of regularity that quantity patterns may be. It is a problem in acoustic analysis to find a useful description of the physical signal which relates to the differences in patterns. The latter problem is not independent of the former.

Although the claims for quantity patterns concern not only disyllables, these are the primary focus of the present work. As described [1] there are three types of quantity pattern for disyllables within the unit of a foot. (A foot is defined as beginning with a stressed syllable and containing the syllables following exclusive of the next stressed syllable. In the example above, "Peter" is contained in a single foot. All of the stimuli discussed here are of this form.) One quantity pattern, Type A, occurs in RP for the words "shilling", and "never", which have a short-long pattern. Type B occurs for the words "greater", and "firmly", which are given a pattern of equal syllable lengths. Type C is assigned to a foot structure with a word division between the two syllables and for which the pattern is long short, such as in the phrase "tea for" (two). In this formulation, quantity patterns do not apply to isolated disyllabic words stressed on the second syllable. Such words would have monosyllabic feet for which the syntagmatic relative patterns would be meaningless. A different formulation is required to handle such words in isolation. In context, the second syllable of the words will be the first syllable of the foot. When such a word is followed by unstressed syllables, this new foot will receive one of the available analyses.

The quantity patterns Type A and B for RP are further specified in terms of the phonetic segmental structure of the words. Type A patterns occur with the structure (C)VCV(C), where the first vowel may be in the set of functionally short, undiphthongized vowels [e, ɛ, a, ɔ, u , o]. (Abercrombie notes that [a]

may not be a functionally short vowel for many talkers). Type B patterns are associated with the structures (C)VCC(C)V(C) and (C)V(C)V(C) where the latter structure has as a first vowel one of the functionally long vowels or diphthongs, as in "Peter".

Such quantity patterns may be of use in machine recognition of speech either directly by limiting the class of possible lexical candidates for a signal, or indirectly in identifying or confirming a talker's dialect from which further rules can be applied to constrain proposed lexical analyses of the signal.

The direct application of quantity patterns to lexical identification may succeed in a variety of ways. Because the quantity patterns are independent of the lexical stress pattern, they can provide additional partitions of the set of disyllables with first syllable stress. Since this set is a large part of the disyllabic vocabulary of English, it is worth pursuing means for its further division. Quantity patterns would have to be specified as lexical information (or be derivable from other lexical information) for matching purposes.

Quantity patterns may also be involved directly in analyses if the joint occurrence of a quantity pattern and lexical item holds further implications either for the analyses of other portions of the signal or for non-lexical local analyses. Consider the hypothetical case of a word normally associated with one quantity pattern, which when associated with the other, marks a syntactic boundary that is otherwise locally ambiguous. It is an empirical issue whether quantity patterns will participate in lexical analyses in such ways. Further, the existence of constraints on sequences of such patterns has never been explored.

Less directly related to lexical analyses, quantity patterns may serve as dialect markers. As such, they may allow us to invoke whatever other information of segmental or lexical regularities are associated with dialect differences which, in turn, may constrain the local or non-local analyses of a signal. Further, quantity patterns qua dialect specifiers may figure in limited ways within the scope of speech recognition by machine, as for example in a training procedure but not in direct processing.

To be useful in a direct way, quantity patterns must be invariant (or predictably variable) with respect to the factors which influence the realization of lexical items: rate of speech, speech style, position of a word in a longer utterance, phonetic context, and, of course, dialect of the talker. For quantity patterns to be useful in the indirect sense as dialect specifiers, dialects must be differentiable in specific ways.

Three ways in which dialects may differ at the phonetic and phonological systemic level are described in O'Connor [4]. Dialects may differ non-systemically, as when each has the same phonetic inventory but the distribution of segments in particular lexical items differs in a manner captured by a correspondence rule. For example, the Cockney occurrence of [ax] corresponds exceptionlessly to the occurrence in RP of [ex]. Dialects may differ systemically when they do not have the same inventory, as for example, the difference in number and quality of the vowels in Scottish and RP. Dialects

may also be distinguished by sub-systemic differences which are localized to specific phonetic environments. Further, selectional differences may exist without systemic differences, as in the differences between Northern and Southern British English pronunciations of the words "glass" and "laugh" involving the use of [ae] and [ɑ], which are available in both inventories.

Abercrombie [1] noted that Type A patterns may be unavailable in some varieties of Standard English. This would be a case of systemic difference. If the observation were confirmed, then the occurrence of a Type A pattern in a signal would serve as information about which dialects could be excluded from consideration. The absence of Type A in any signal in itself would not be reliable evidence because of the many other factors aside from the dialect of the talker that may govern the opportunities for the pattern to occur, such as constraints on vowel quality of a particular lexical item.

It is more probable that dialects will show non-systemic differences of the type in the example above for the word "Peter". Dialects with the same inventory of possible quantity patterns may show different associations of the patterns with what may be considered the same lexical item. This entails either (or both) that the correspondence of segmental structure with a quantity pattern differs across the dialects, or that the segmental, specifically the vocalic qualities, of particular items in their underlying representations differ across dialects. For example, if "Peter" has a Type A pattern in one dialect and a Type B pattern in another [1], either the association of CVCV(C) with Type A is not the same in both dialects, or the vowel quality in the first syllable is functionally short for the first dialect and functionally long for the second dialect. Other analyses, such as a difference in the number of medial consonants, (which could also motivate a difference in quantity patterns without a dialect difference in the correspondence of segmental sequences with quantity patterns) are not available in this example, although they may be for other lexical items.

If dialects are found to have different correspondences of segmental sequences (conditioned by vowel quality) with quantity patterns, then a quantity pattern together with vowel quality information for a signal will serve to identify or at least partition the dialects. To illustrate this, consider a dialect for which CVCVC with short vowels in the first syllable is associated with the short-long pattern of Type A and further that Type A occurs in this dialect only with these conditions. It will be sufficient for identifying this dialect (though perhaps not uniquely) to find a Type A pattern in a signal for which the first vowel is phonetically short and there is a single medial consonant. It will be sufficient to exclude this dialect to find a Type A pattern in a signal for which the vowel is long or for which there is more than one medial consonant. The work in identifying (and excluding) dialects depends on the full specification of quantity patterns for the set of dialects of interest.

The correspondence of quantity patterns and segmental constraints within dialects of British English remains to be determined. This is not the only obstacle to exploiting quantity patterns in machine recognition of speech. The link between the acoustic signal and the quantity patterns also requires investigation. This is the problem in focus in the current research.

DURATIONAL CORRELATES OF SYLLABLE QUANTITY PATTERNS

ACOUSTIC CORRELATES OF QUANTITY PATTERNS

For Abercrombie [1] quantity patterns are perceptual attributes of a linguistically structured signal. Plausible candidates for the acoustic correlates of quantity patterns appear to be relative durations of portions of the signal and the interactions of temporal structure with resonance characteristics that specify the vowels [3]. If perceptual tests were to reveal that listeners use perceived syllable length (perhaps normalized for intrinsic attributes of the component segments [3]), then values of relative temporal measures could be further adjusted for the coocurrence of physical attributes which influence listeners' judgements of syllable length. Most important among plausible non-temporal conditioning factors would be the fundamental frequency contour. It has been established that listeners' reports of vowel duration are influenced by dynamic characteristics of F0, as are those of stress patterns and syllable structure.

Although Abercrombie makes use of the notion syllable in his descriptions, he does not provide a way of determining syllable structure. The segmental sequence constraints for quantity patterns do not make use of syllable boundaries. Thus CVCVC sequences may be either CV-CVC, CVC-VC, or possibly CVC-CVC. The absence of syllable demarcations has no effect on the correspondence formulations as given. Unspecified syllable structure has serious implications for determining which portions of the acoustic signal are relevant to specifying quantity patterns, if indeed it is syllable structure that is relevant to listeners.

It is possible that quantity patterns are particularly difficult attributes for listeners to detect or use. The existence of such patterns is not mentioned in a number of works on the temporal regularities of polysyllabic utterances in English [2,3,4]. Reports of the quantity patterns of signals would be convenient for the interpretation of the acoustic analyses. The reports are not logically necessary for either the task of lexical analysis or dialect specification. What is required are consistent values from talkers' productions, values which reflect a unity for talkers from the same dialect while they are diverse across dialects, and values which reflect little variance across tokens of a word spoken by talkers sharing a dialect while the values differ for words of different syllable structure, segmental sequences, or stress patterns. Put differently, to be useful, the acoustic values need only to make the required separations of lexical items and dialects and to preserve the required identities. The values need not map onto the impressionistic descriptions of the quantity patterns in any way; the values need not be large for the long-short pattern nor small for the short-long pattern for example. It is an empirical issue to determine how transparently related are the values observed and the possible quantity patterns.

If for convenience it is accepted that vowels and, as claimed, syllables are relevant linguistic units to guide acoustic segmentation of the signal, and further that quantity patterns are best related to temporal attributes of the signal, then at least two questions must be answered in order to provide acoustic correlates of quantity patterns in the temporal domain. It must be determined how best to express the temporal relational patterns, and, not independently, how best to relate observed values to the separation of

dialects. In this study, fourteen expressions, seven each involving syllable durations and vowel durations are tested. For relating values of these expressions to the separation of dialects, a critical point model and a set point (target) model are tested.

The fact that quantity patterns may differ for words of the same stress pattern does not imply that the acoustic correlates of the stress and quantity patterns are independent. Among the expressions tested are those known to be related to the stress patterns, such as ratios of the duration of the first vowel to that of the second [3].

Experimental methods

The stimuli were eight disyllables from a dialect study of 192 isolated words (eight occurrences of each of 24 vowel qualities in different lexical items). Six of the disyllables have a pattern of first syllable stress and two have a pattern of second syllable stress.

Four adult male native speakers of British English participated in the production study. One each is a speaker of Southern Standard English and of Midlands English (Birmingham); two are speakers of RP. Additionally speakers are being recruited from these and other dialect regions. Talkers read aloud the isolated words in two different orders. As a consequence, the position of six of the disyllables varies between list final and non-final sequence.

Tokens of the word "father" were chosen for analyses because this word was predicted to have different quantity patterns in these dialects. It has a Type B pattern in RP. The two repetitions by each talker were filtered at 4.5 kHz and digitized at a 10 kHz sampling rate. The waveforms were displayed and measured on the editing system developed for the Speech Lab of the University Engineering Department.

Waveforms were partitioned into four acoustic sections which identified the acoustic onset and offset of the initial fricative, the first vowel, the medial fricative, and the final vowel. Together with the total duration of each token, these measurements were used in thirteen expressions of the temporal structure of the tokens. Mean values for each talker for each expression are shown in Table 1. S1 and S2 represent the duration of the first and second syllables, respectively. For this word, S2 includes the duration of the medial fricative. V1 and V2 represent the duration of first and second vowels, respectively. T and S represent the total (S1+S2) and sum (V1+V2), respectively. DS and DV represent the difference in syllable durations (S1-S2) and in vowel durations (V1-V2), respectively. The Southern Standard talker, the Midlands talker, and the two RP talkers are represented by T1, T2, T3 and T4, respectively.

Some expressions do not show relative invariance for the two RP speakers and hence can be eliminated from the list of plausible acoustic correlates of syllable quantity patterns. Total word duration and the ratio of the vowel durations divided by the sum of the vowels do not provide the desired non-overlapping ranges for between dialect specifications. Similarly, the simple ratio of the two vowels fails to distinguish the Southern Standard and Midlands talkers.

DURATIONAL CORRELATES OF SYLLABLE QUANTITY PATTERNS

Table 1. Mean durational values for "father"

	T1	T2	T3	T4	
Total msec	540	573	559	590	
S1/T	.57	.59	.54	.53	
S1-S2 msec	75	100	49	32	
S1/S2	1.3	1.4	1.2	1.1	
(S1/S2)/T	2.4	2.5	2.1	1.8	(x .001)
DS/T	1.4	1.7	.9	.5	(x .1)
(S1/S2)/DS	1.7	1.4	2.4	3.4	(x .01)
Sum msec	377	365	398	406	
V1/S	.52	.53	.54	.57	
V1-V2	16	23	33	58	
V1/V2	1.1	1.1	1.2	1.3	
(V1/V2)/S	2.8	3.1	2.9	3.4	(x .001)
DV/S	.4	.6	.8	1.4	(x .1)
(V1/V2)/DV	6.8	4.9	3.5	2.3	(x .01)

Our first impressions that the pattern of equal syllable quantities in RP might have temporal acoustic correlates receive some confirmation from the fact that the mean simple ratio of the syllables approaches unity. The simple difference in syllable durations approaches zero. Non-independently, the ratio of the first syllable to the total word duration approaches one half. This pattern contrasts with that obtained from considering the vowel durations. For the RP talkers, the ratio of the two vowels does not approach unity, as it does for the Southern Standard talker. For the partitioning assumed for this word, the syllable-based expressions more closely mirror our intuitions about quantity patterns than do the vowel-based expressions, first by uniting the data from two talkers and secondly, by the relatively transparent relationship between the form of the expressions and the pattern of equality. As noted above, however, this last aspect is not a necessary one.

It is not possible with the limited data available to rank the expressions in terms of the success with which they separate tokens from different dialects. Analyses of additional tokens are underway. None of the expressions reveals a set point or target value for the two RP speakers. It is, however, difficult to appreciate how much of a difference makes a difference because of the limited data. For each of the five ratios involving syllable durations, there exists a critical point for which the values of RP talkers are on one side of the critical point and the values of the other two dialects are on the other side. For example, the simple ratio of the two syllables has a putative critical point at 1.2, above which the tokens are not from RP and equal to or below which the tokens do come from RP. In this case, it is also possible that a second critical point exists (1.35) which distinguishes the tokens from Southern Standard and Midlands English. There is no logical requirement that the expression whose values distinguish one dialect from another need be the same for all pairs of dialects.

DURATIONAL CORRELATES OF SYLLABLE QUANTITY PATTERNS

In summary, expressions of syllable durations were found to distinguish tokens of the same word spoken by talkers from different dialects. Further, these expressions have values which are interpretable with respect to intuitions about syllable quantity patterns more naturally than are those from expressions of vowel durations. With limited data, critical points can be found for some expressions which unite tokens from a single dialect. Many of the factors known to affect observed durations of vowels and syllables remain to be studied with respect to these candidate expressions. Among them are the rate of speech and the intrinsic characteristics of the component vowels. Although only acoustic analyses have been performed and perceptual studies remain, acoustic correlates of quantity patterns have been shown to have usefulness in the identification of dialects.

REFERENCES

[1] D. Abercrombie, 'Syllable quantity and enclitics in English', In Honor of Daniel Jones, (Eds.) Abercrombie, Fry, MacCarthy, Scott and Trim, Longmans: London, 216-222, (1964).
[2] D. Crystal, Prosodic Systems and Intonation in English, Cambridge Univ. Press, 108-125, 152-179, (1969).
[3] I. Lehiste, Suprasegmentals, MIT Press: Massachusetts, 6-53, (1970).
[4] J.D. O'Connor, Phonetics, Penguin Books: England, 152-175, 180-188, 194-202, (1973).