STOCHASTIC SUPRASEGMENTALS: RELATIONSHIP BETWEEN THE SPECTRAL CHARACTERISTICS OF VOWELS, REDUNDANCY AND PROSODIC STRUCTURE

Matthew P. Aylett

Department of Linguistics, University of Edinburgh
and Rhetorical Systems Ltd.
email: matthewa@cogsci.ed.ac.uk

ABSTRACT

Previous work has shown a relationship between syllabic duration, redundancy within speech, and prosodic structure [1]. In addition, a spectral care of articulation measure of vowels in spontaneous speech has supported these duration results and suggest that care of articulation varies inversely with redundancy and conversely with prosodic prominence. However, these spectral measures remain inconclusive due to measurement difficulties. In this paper a simpler spectral measurement is presented as a metric of care of articulation and applied to three vowels from each corner of the vowel triangle. Prosodic and redundancy factors can predict up to 5% of the variance of this new measurement, supporting the inverse redundancy result. However, in contrast to syllabic duration, whether prosodic boundaries are controlled for or not, the predictive power of prosodic factors and redundancy factors remain relatively independent. This suggests that 1. phrase final syllables, despite lengthening, do not show increased care of articulation if measured spectrally, and 2. unlike duration, redundancy factors affect the spectral characteristics of vowels independently of prosodic structure.

1. INTRODUCTION

Two factors which are known to affect care of articulation are prosodic structure and language redundancy.

Prominence has a direct effect on acoustic factors linked with clear speech and careful articulation. Prominent syllables are longer [1, 2, 3] and the vowels are less spectrally reduced [3, 2, 4].

Articulatory studies have also associated prominence with changes in articulation. The duration, velocity and spatial extensiveness of jaw opening is increased [2] and the openness of the vocal tract increases. This results in increased acoustic power and more extreme spectral features in vowels [2]. de Jong argues that this shift in spectral features is made in order to increase perceptual clarity and is better regarded as hyper-articulation than a simple increase in amplitude.

A number of studies have persuasively shown that more predictable sections of speech exhibit the same acoustic reduction and shortening that is common in hypospeech and avoided in hyperspeech e.g. [5, 6, 7, 8]. Measures of redundancy have varied from a subjects ability to guess a word, the effects of given and new in dialogue, to regularities in the lexicon.

However, few corpus studies have been carried out which examine these factors ([7] a notable exception). This is partly due to the difficulty in both controlling and automatically measuring the acoustic variation associated with predictability and prosodic prominence. [1] demonstrated that strong results could be obtained for simple duration measurements but found it more difficult to examine spectral change in vowels. In this paper a simple spectral measurement is explored across a large corpus of spontaneous speech [9] and the relationship between this spectral measure of care of articulation, redundancy and prosodic structure, is explored.

2. METHOD

A combination of automatic prosodic coding together with care of articulation and language redundancy metrics were applied to each syllable in the HCRC Map Task Corpus [9].

2.1. Prosodic Structure

The HCRC Map Task Corpus is word segmented. Syllable boundaries (for polysyllabic words) were determined using autosegmentation. A dictionary containing a canonical phonemic representation for each word was used to guess the probable segmental contents of each word. The prosodic variables used were as follows:

[wboun:] Word boundary. This corresponds to a ToBI break index of 1.

[Aipboun:] Automatically coded Full Intonational Phrase Boundary. If the syllable was followed by a pause it was

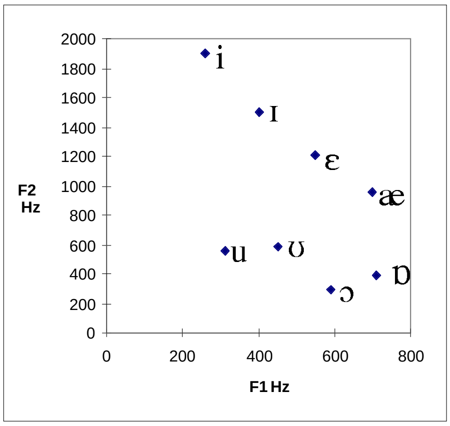

Fig. 1. The Vowel Space

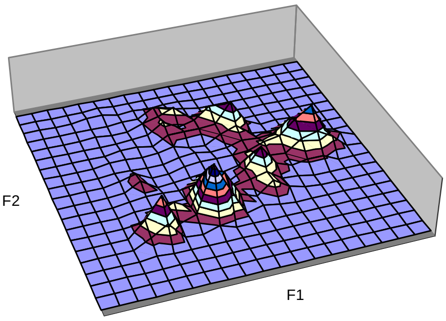

Fig. 2. 3d scatter diagram of the vowel space

regarded as having a high likelihood of being followed by a full intonational phrase boundary.
[lexstr:] Lexical stress. Whether the syllable is lexically stressed.
[Aacc:] Automatically coded Phrasal Accent. If the syllable was lexically stressed **and** open class, it was marked as having a high likelihood of having a phrasal accent.

2.2. Redundancy

Two measurements of redundancy were used:

The Word Level (wf): Log of Word Frequency. More frequent words should be more easy to predict and thus be more redundant. Each syllable was associated with the COBUILD word frequency of the word it was part of.

The Syllable Level (trigram): Syllabic Trigram Measurement. Using the spoken part of BNC (British National Corpus), the transition probability of guessing a third syllable on the basis of the first two. This measurement gave some idea of predictability produced by frequent sequences of words and the redundancy of later syllables in polysyllabic words.

2.3. Care of Articulation Measurement: Simple Vowel Quality

Different vowels have different characteristic spectral qualities. Areas within the spectrum of a vowel with relatively high energy frequency components (i.e areas around these peaks) are termed formants. In vowels the frequency of formants, generally the first and second formant (F1, F2), can be used to categorise vowels.

This two dimensional space can be referred to as the vowel space (Figure 1). The triangular shape made by the three vowels /i, u, ɒ/ (heed, who'd, hod) is often referred to as the vowel triangle. A scatter plot of F1/F2 values from vowels in citation speech show how actual values produced relate to the vowel space. If the density of the scatter is plotted as a third dimension, a 3D plot of the vowel space is produced (Figure 2). In this plot the hills show locations of high density. In general the values of F1 and F2 making up each hill will correspond to an example of a particular vowel.

In studies that examined F1/F2 values in carefully and less carefully articulated vowels e.g. [4], it was found that the vowel formants tend to be less extreme in less carefully articulated speech and closer the centre of the vowel triangle. This *centralisation* could be caused by the formant not reaching the extreme vowel target that it would in carefully articulated speech. Thus if we can compare the F1/F2 value of speakers vowel in spontaneous speech with the values in carefully articulated citation speech this difference can be used as a measure of care of articulation.

For each of the 32 male[1] speakers in the HCRC Map Task Corpus a list of words is read slowly and carefully. A formant tracker is applied to this speech and the mean and variance of F1 and F2, for three vowels (one from each corner of the vowelspace - /i, u, ɒ/ (heed, who'd, hod) are calculated. The values are taken from the central point of the vowel after segmentation using HTK.

These values are then regarded as typical values for hypa-articulated vowels and the Gaussian distribution of F1 and F2 regarded as a model of this vowel in the vowel triangle. The probability of a vowel of the correct type falling into this distribution is then regarded as a measure of the hypa-articulation of the vowel.

[1] A comparison between the automatic formant tracker and formants coded by two phoneticians showed that the formant tracks were quite unreliable (58% agreement with hand coding for F1, 84% for F2 [1]). This was particularly true for female speakers and so only male speech was analysed.

3. RESULTS

3.1. Does vowel spectral clarity relate to prosodic structure?

Using the probability of the vowel as being part of the citation distribution of its F1/F2 as a dependent variable a multiple linear regression was carried out against prosodic factors. Overall prosodic factors predicted a small proportion of the variance $(r = 0.1083, r^2 = 0.0117, F(4, 9380) = 27, p < 0.001)$. Individual factors varied in their significance (2 tailed t test) **lexstr** - $p < 0.001$, **Aacc** - NS, **wboun** - $p < 0.001$, **Aipboun** - $p < 0.05$. Figure 3 shows the mean spectral clarity in each prosodic context. Lexical stress is associated with a strong rise in spectral clarity, while a following word boundary reduces it. Intonational phrase boundaries do not appear to have a marked affect. In is interesting to compare this result with that of log syllabic duration (see Figure 3). Here we can see the marked affect on duration of prominence but also, in particular, prosodic boundaries. In addition, prosodic factors predicted a much larger proportion of the variance of log syllabic duration (42%) $(r = 0.6481, r^2 = 0.42, F(4, 9380) = 1698, p < 0.001)$.

3.2. Does vowel spectral clarity relate to redundancy?

A similar regression test was carried out comparing the redundancy factors with spectral clarity. Again a small proportion of the variance was predicted $(r = 0.1075, r^2 = 0.0116, F(4, 9380) = 54, p < 0.001)$. Only the syllabic trigram proved a significant factor, **trigram** - $p < 0.001$. However if syllables with lax vowels are included in the analysis word frequency does have a significant affect on log syllabic duration suggesting a large proportion of this is due to unstressed function words. Figure 4 shows the mean spectral clarity measure for each quartile of the trigram measurement. As we can see the more probable a syllable given the two previous syllables the lower the spectral clarity according to our simple measurement. If we compare this to the duration results (Figure 4) we see that although the effect is mirrored more strongly in the first three quartiles in the last quartile, the most probable trigrams, the duration has tended to increase.

3.3. Looking more closely at individual vowels

The power of the regression model for spectral clarity is disappointing. If we break the regression down by vowel we find the model varies greatly in how well it predicts the spectral clarity of different vowels.

Looking at Table 1 we see that it is the /i/ vowel which has contributed mostly to any success the regression model has enjoyed. Although still of low predictive power it is possible to merge the prosodic and redundancy models to give the combined predictive power and to examine the independence of the redundancy and the prosodic factors in predicting the spectral characteristics of /i/ in spontaneous speech.

A combined model, including all prosodic and redundancy factors, predicts 5% of the variance $(r = 0.2223, r^2 = 0.0494, F(6, 3034) = 26, p < 0.001)$. Table 2 compares the independent contribution of redundancy and prosodic factors in predicting spectral results and syllabic log duration results both with and without controlling for intonational phrase breaks.

Unlike a similar analysis in [1], carried out over all syllables in the corpus, the shared, or non independent contribution of the two models does not change much in either context.

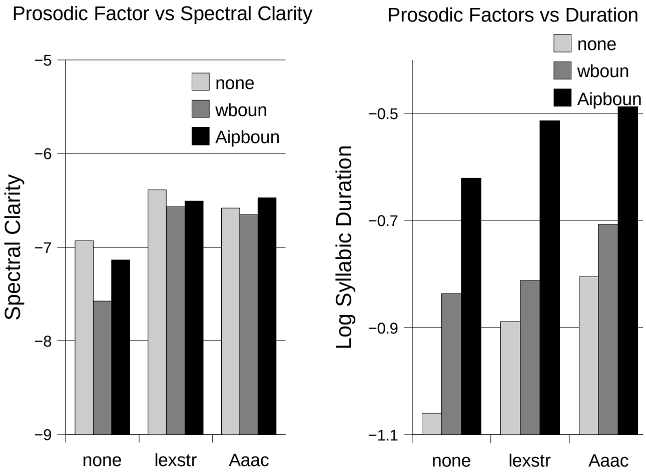

Fig. 3. Prosodic effects on spectral clarity and duration

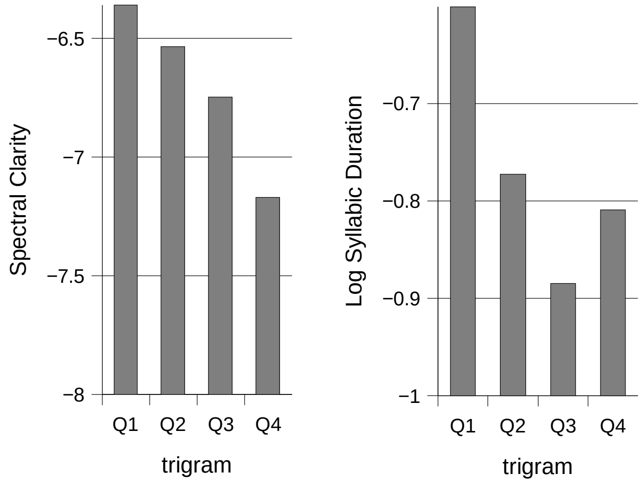

Fig. 4. Trigram redundancy effects on spectral clarity and duration

Vowel	Regression Coefs.					
	Prosodic Structure			Redundancy		
	r	r^2	$pvalue$	r	r^2	$pvalue$
i	0.184	0.034	0.001	0.161	0.026	0.001
u	0.086	0.007	0.05	0.090	0.008	0.005
ɒ	0.064	0.004	0.001	0.025	0.000	NS

Table 1. Differences between vowels and regression model's predictive power.

Percentage of variance predicted by Model			
	Pros. Model	Red. Model	Shared
Syllabic Log Duration	27%	6.5%	11.6%
Syllabic Log Duration - no IPs	15%	11.3%	5.9%
Spectral Clarity	2.3%	1.5%	1.1%
Spectral Clarity - no IPs	2.8%	2.8%	0.9%

Table 2. Differences between the regression model's predictive power.

3.4. Summary of Results

1. The spectral clarity, measured by a simple F1/F2 model, is significantly related to Prosodic factors. Lexical stress makes the clarity greater, word boundaries reduce it, intonational phrases increase it slightly.
2. This result is in marked contrast to duration which increases both with prominence and boundary strength. Intonational phrase boundaries having a major affect on duration.
3. The spectral clarity, measured by a simple F1/F2 model, is significantly related to redundancy measured by syllabic trigram probability. The more likely a syllable the more reduced the vowels spectral clarity.
4. This result follows duration results more closely, although a difference is noted in the most reduced/ most probable quartile of the data.
5. These results are very different for each of the three vowels. This suggests that it is difficult to generalise the spectral measure over different vowels. Results for the /i/ are very much stronger than for the other two vowels.
6. Unlike log syllabic duration when measured over all syllables with IPs controlled, the redundancy and prosodic models both contribute independent predictive power to the duration and spectral results with shared power varying from 13-25% of the overall predictive power (In contrast to 27-62% [1] when looking at log syllabic duration in all syllables in the corpus).

4. DISCUSSION

The results presented here go someway to supporting the view that care of articulation is closely related to redundancy in speech and to prosodic structure. However, in contrast to duration, producing robust results for spectral measures remains elusive. This can be ascribed to problems in automatically measuring these spectral values (formant trackers are far from reliable when applied to spontaneous speech from casual dialogue). However the problem of normalising spectral values for different vowels to represent care of articulation is also a major concern. In the results presented here it is difficult to establish whether results were different for /i/ because the F1/F2 values are more widely separated in general, making them both easier to measure and to compare, or because the measure is invalid for the other two vowels.

A strong relationship between prosodic structure, redundancy and care of articulation supports the idea that, in English, prosodic structure is the means with which constraints caused by a robust signal requirement are expressed in spontaneous speech. However, although such a relationship does appear to exist [1] and although the results presented here support this, a less noisy, theoretically motivated spectral measure of care of articulation is required.

5. REFERENCES

[1] Matthew P. Aylett, *Stochastic Suprasegmentals (http://www.cogsci.ed.ac.uk/m̃atthewa/thesis_sum.html)*, Ph.D. thesis, University of Edinburgh, 2000.

[2] Kenneth J. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation.," *JASA*, vol. 97, pp. 491–504, 1995.

[3] D. R. van Bergem, "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," *Speech Communication*, vol. 12, pp. 1–23, 1988.

[4] Seung-Jae Moon and Björn Lindblom, "Interaction between duration, context and speaking style in English stressed vowels," *JASA*, vol. 96, pp. 40–55, 1994.

[5] P. Lieberman, "Some effects of semantic and grammatical context on the production and perception of speech," *Language and Speech*, vol. 6, pp. 172–187, 1963.

[6] D.B. Pisoni, H.C. Nusbaum, P.A. Luce, and L.M. Slowiaczek, "Speech perception, word recognition, and the structure of the lexicon.," *Speech Communication*, vol. 4, pp. 75–95, 1985.

[7] A. Bell, D. Jurafsky, E. Fosler-Lussier, C. Girand, and D. Gildea, "Forms of english function words - effects of disfluencies, turn position, age and sex, and predictability," in *ICPhs99*, 1999.

[8] Ellen Bard and Matthew Aylett, "The dissociation of deaccenting, givenness and syntactic role in spontaneous speech.," in *ICPhs99*, 1999.

[9] Anne H. Anderson et al, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1991.