STOCHASTIC SUPRASEGMENTALS: RELATIONSHIPS BETWEEN REDUNDANCY, PROSODIC STRUCTURE AND CARE OF ARTICULATION IN SPONTANEOUS SPEECH

Matthew Aylett

Department of Linguistics,
University of Edinburgh
email: matthewa@cogsci.ed.ac.uk

ABSTRACT

Within spontaneous speech there are wide variations in the articulation of the same word by the same speaker. This paper explores two related factors which influence variation in articulation, prosodic structure and redundancy.

We argue that the constraint of producing robust communication while efficiently expending articulatory effort leads to an inverse relationship between language redundancy and care of articulation. The inverse relationship improves robustness by spreading the information more evenly across the speech signal leading to a smoother signal redundancy profile. We argue that prosodic prominence is a linguistic means of achieving smooth signal redundancy. Prosodic prominence increases care of articulation and coincides with unpredictable sections of speech. By doing so, prosodic prominence leads to a smoother signal redundancy.

Results confirm the strong relationship between prosodic prominence and care of articulation as well as an inverse relationship between language redundancy and care of articulation. In addition, when variation in prosodic boundaries is controlled for, language redundancy can predict up to 65% of the variance in raw syllabic duration. This is comparable with 64% predicted by prosodic prominence (accent, lexical stress and vowel type). Moreover most (62%) of this predictive power is shared.

This suggests that, in English, prosodic structure is the means with which constraints caused by a robust signal requirement are expressed in spontaneous speech.

1. INTRODUCTION

In order to explain phonetic variation Lindblom [5] in his H&H (hyper- and hypospeech) theory presents the idea that differing degrees of articulatory effort are used in different circumstances. Lindblom argues that a speaker assesses the needs of a listener and balances the effort used in producing speech against the need for producing speech which is sufficiently discriminable. In doing so the speaker alters articulation in response to communicative and situational demands along a continuum of hyper- and hypospeech. Lindblom argues that communication is more efficient if the speaker expends less articulatory effort for easily discriminable sections of speech and more articulatory effort when speech is more difficult to discriminate. However if we regard speech as 'language encoded into an acoustic signal' we can regard articulation as the encoding process, language as the information we wish to encode, and speech as the resulting signal. In these terms *articulatory effort* is related to the redundancy of this encoding process and *sufficient discriminability* to the overall signal redundancy. It is then the efficiency of this encoding process *given a noisy environment* which dictates the form of the resulting signal rather than the listeners needs per se.

1.1. Redundancy

Redundancy only has a meaning with regards to a statistical model. In language we can build different models for different levels of structure. From a phonetic perspective it is useful to distinguish between a model based on the compositional structure of language to one based on the acoustic observations of a section of speech.

The first, *the language model*, is the likelihood of a word, syllable or phoneme appearing in the speech stream. The second is *the acoustic model* which is the likelihood of specific acoustic observations being connected with a word, syllable or phoneme.

These two models are combined to produce a third level of redundancy, *signal redundancy*. Signal redundancy, the overall redundancy in the signal, is calculated by combining the language model and the acoustic model. Because signal redundancy is the combination of these two previous models and because it is good for signal redundancy to be smooth to combat noise this leads to my central hypothesis. *For signal redundancy to tend to smoothness requires that sections of speech which are very language redundant will tend to be sections of speech which are less acoustically redundant and thereby less salient and distinctive*.

From this perspective, variation in articulation is caused by the general requirement of producing a robust signal in a noisy environment. It is the demands of the individual language, in its structure and compositionality, which leads to such changes in articulation. Prosodic structure could be the linguistic means of controlling such variation because, 1) calculating language redundancy online is hard, 2) such redundancy can be decomposed linguistically[1], and 3) there is extensive psycholinguistic evidence that we are directly aware of prosodic structure.

To help clarify this information theoretic viewpoint, it is useful to

[1] For example redundancy at the lexical level (how predictable is that word given the lexicon) words can be calculated separately and then used to calculate redundancy at the phrase level (how predictable is that word given the previous words in the phrase).

compare what could be regarded as a traditional view of prosody with the model suggested by this hypothesis. Figure 1 (taken from [6]) shows a traditional view of prosody. Here a whole set of different factors are controlling how prosodic structure influences the phonetic shape of utterances.

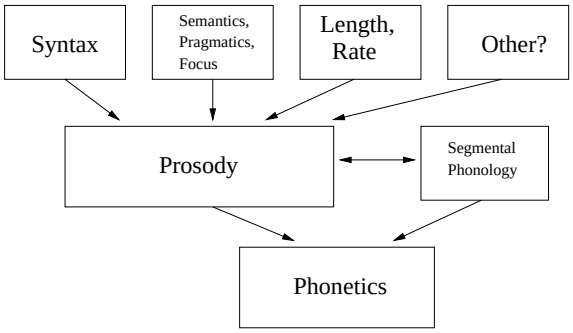

Figure 1: One view of the role of the prosodic component of the grammar [taken from 6, page 237].

In contrast, Figure 2 shows how a smooth redundancy hypothesis could be modelled. The component affecting prosodic structure is language redundancy rather than a set of different factors such as syntax, semantics, speech rate etc. This redundancy component is then encoded into prosodic structure in order to make the signal redundancy smooth and the communication robust.

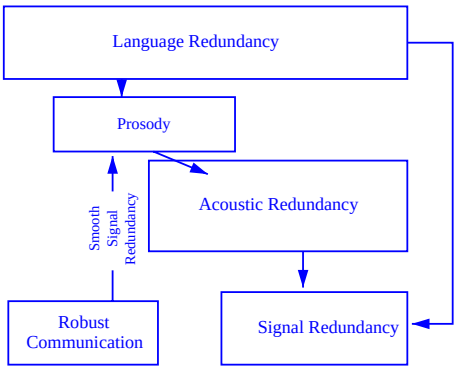

Figure 2: Smooth redundancy hypothesis.

In order to test this redundancy model we need to examine the relationship between prosodic structure, care of articulation and language redundancy. The more prosodic structure predicts the same changes in care of articulation as language redundancy, the more convincing the argument that prosodic structure is there to affect these changes.

2. METHOD

In order to investigate this hypothesis we have examined a very large corpus of spontaneous speech, the HCRC Map Corpus [3]. A combination of automatic prosodic coding together with care of articulation and language redundancy metrics were applied to each syllable on the corpus.

2.1. Prosodic Structure

The HCRC map task 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 [2]. 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 regarded as having a high likelihood of being followed by a full intonational phrase boundary.

vtype: Vowel type. Whether the vowel is full or reduced (where reduced equals /I,ə/ in lexically unstressed syllables).

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

In this work three measurements were taken based on word frequency, syllabic trigram probability and givenness. The aim of these measurements was not to present a theoretical model of redundancy in language but rather to approximate such redundancy. The metrics cover redundancy at:

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 in later syllables of polysyllabic words.

The Discourse Level (**men**): Givenness. This relates to the introduction of a referent in a dialogue. The more this referent is mentioned the more 'given', and thus predictable, it becomes (in this case a referent was a landmark on a map).

2.3. Care of Articulation

In general more carefully articulated speech or 'clear speech' is longer. Word duration is greater in 'clear speech' than when the same word is spoken in spontaneous or citation speech [7]. However, although lengthening tends to occur as a side effect of more

[2] A small section of the corpus, 679 full intonational phrases, were hand coded. The results for these hand coded materials, and an evaluation of the automatic coding, is presented in Aylett [2]).

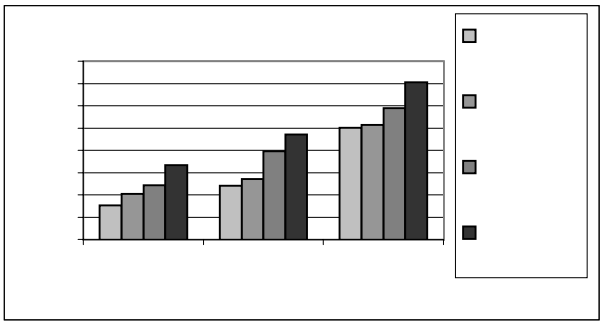

Figure 3: Automatic prosodic factors vs raw syllabic duration.

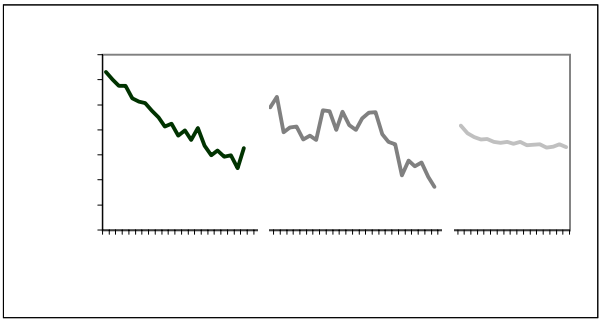

Figure 4: Redundancy factors vs raw syllabic duration.

carefully articulated speech, it can also occur when care is not being taken [4][3].

Despite this, for practical reasons, we will focus on duration. In the results reported here, the raw duration in milliseconds of the syllable is used to represent care of articulation[4].

3. RESULTS

3.1. Prosody

In terms of duration change prosodic factors predicts up to 42% of raw syllabic duration in automatically coded materials (See Table 1), although it was found that a large amount of interdependency exists between different factors. The greater the prominence and the greater the boundary following the syllable the longer the syllable. Figure 3 shows the mean values of raw syllabic duration for different prosodic categories.

[3]Earlier results obtained for a spectral measurement based on vowel quality were inconclusive [2]. In the light of this, and, with no other dependable spectral technique available that could be applied to spontaneous speech, a duration measurement was regarded as the most dependable and practical solution to measuring care of articulation within a large corpus. Firstly, because of the overwhelming evidence of a strong correlation between care of articulation and duration, and secondly, the fact that areas where this relationship is more suspect, such as phrase final syllables, are analysed separately in this study.

[4]In both [2] and [1] a normalised duration score based on chained log distributions was also examined. However as results from this normalised duration score were similar to the raw duration in most cases only results for raw syllabic duration will be presented here.

3.2. Redundancy

For all materials, trigram and word frequency factors predicted 14% of the raw duration variation (see Table 2). However when prosodic boundaries were controlled, this rose to 40%. For *landmark referents* with controlled prosodic boundaries, redundancy factors predict 65% of raw syllabic duration change. The more predictable a syllable the less carefully articulated a syllable is (see Table 1 and Figure 4).

PROSODIC FACTORS All Materials			
	Regression Results	$r = 0.6473$	$r^2 = 0.4190$
Auto Prosodic Factor	Independent Contrib. to r^2	F(1,169461)	p value
vtype	01.08%	3139.49	0.001
lexstr	00.83%	2421.31	0.001
Aacc	01.49%	4335.15	0.001
wboun	03.62%	10561.72	0.001
Aipboun	19.72%	57523.91	0.001
REDUNDANCY FACTORS (Prosodic Boundaries Controlled)			
All Materials			
	Regression Results	$r = 0.6081$	$r^2 = 0.3698$
Redundancy Factor	Independent Contrib. to r^2	F(1,89531)	p value
wf	10.11%	14361.29	0.001
trigram.	01.93%	2736.84	0.001
Mention Coded Materials			
	Regression Results	$r = 0.8085$	$r^2 = 0.6536$
Redundancy Factor	Independent Contrib. to r^2	F(1,12294)	p value
wf	06.06%	2150.52	0.001
trigram	00.74%	263.66	0.001
men	00.33%	116.28	0.001

Table 1: Regression analysis of prosodic and redundancy factors and raw syllabic duration.

3.3. Independence

Comparing the independent contribution of redundancy factors and prosodic factors to predicting duration (see Table 2 and Figure 5) it was found that:

1. Most of the contribution made by redundancy factors is implicitly represented by prosodic factors. However a significant but small percentage (7-10%) predicted even by these very simple redundancy metrics was not represented by prosodic factors.
2. Prosodic factors made a large independent contribution to predicting duration change above that representing redundancy (about 35% compared to a shared prosodic/redundancy contribution of 7% over all sets of materials, see Table 2).
3. This independent contribution of the prosodic factors was much smaller for syllables where prosodic boundaries were controlled for (2-4% over all sets of materials, see Table 2). This suggests a major role of prosodic prominence is to smooth signal redundancy by controlling care of articulation in a way which implicitly mirrors language redundancy factors.

Independent and Shared Contribution				
	Prosodic Boundaries NOT Controlled			
Materials	r^2	Independent		Shared
		Pros.	Red.	
All	49.01%	34.49%	7.11%	7.41%
M	61.06%	29.76%	9.64%	21.66%
	Prosodic Boundaries Controlled			
Materials	r^2	Independent		Shared
		Pros.	Red.	
All	41.44%	4.46%	9.70%	27.28%
M	67.83%	2.47%	3.62%	61.74%

Table 2: Independent contributions of redundancy - *Red.* and prosodic models - *Pros.* in predicting the variance of raw syllabic duration. The non-independent, shared contribution is shown under *Shared*. M: Materials with mention coding. All: Materials with automatic prosodic coding and a trigram/word frequency redundancy model. All results are significant.

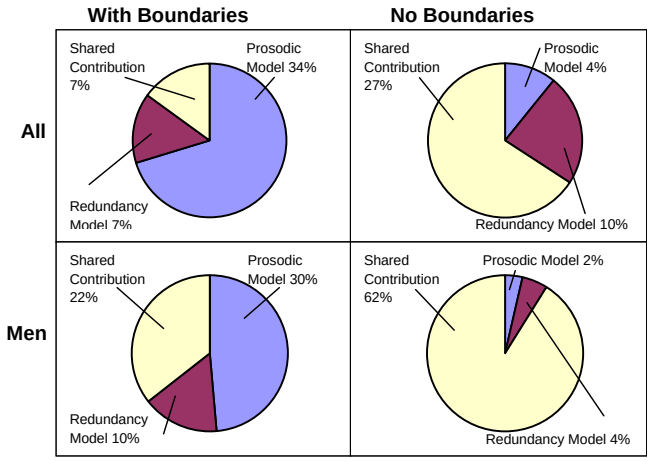

Figure 5: Shared and independent contribution of prosodic and redundancy models in predicting raw syllabic duration for all materials (All) and mention materials (Men) with and without boundaries controlled.

4. DISCUSSION

Redundancy factors were most successful at predicting raw syllabic duration in syllables occurring in references to landmarks (the mention set) when prosodic boundaries were controlled for. In this case the redundancy model predicted 65% of the variation ($r = 0.8085$). This predictive power seems high. Especially when you consider it is based on such simple redundancy measurements.

For the same materials, prosodic structure also accounted for about the same amount of variance (64%, $r = 0.8013$). When we consider a joint model of redundancy factors and prosodic factors we find that an enormous 62% of the predictive power is shared.

Although not as extreme, results over the other materials supported the extent to which prosodic factors embodied these redundancy factors. When boundaries were considered, the shared predictive power fell to about a third or less of the variance predicted, when boundaries were controlled, this rose to two thirds or more of the variance predicted.

If we take a critical look at the more traditional view of prosody embodied in figure 1 and compare it with the redundancy model shown in figure 2 we can make some interesting observations.

Firstly the traditional view does not offer a theoretical framework for why some things affect prosody and others do not. Each area, syntax, semantics, discourse structure are treated independently in this traditional view. The reasons some syntactic factors affect prosody and some do not are not related to the reasons some semantic factors affect prosody and some do not. By looking at language in terms of redundancy we can relate these different factors to each other. Concepts as diverse as focus, syntactic structure, word class, length of utterance and word frequency can be looked at in terms of a predictive model and thus in terms of language redundancy. In addition, the reason language redundancy should affect care of articulation and thus be expressed in terms of prosodic structure follows persuasively from the requirements of getting information from A to B within a noisy environment. This does not mean that other factors outside redundancy do not affect prosody, for example psycholinguistic or phonological constraints, however it does shed some light on why we have prominence.

The results from this work suggest that prosody acts as a interface between the compositional structure of language and the constraints of producing a robust and effective signal. This role of prosody in smoothing signal redundancy is crucial to why prosodic structure is as it is and why it works as it does.

REFERENCES

1. Matthew Aylett. Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and syllabic duration. In *ICPhs99*, 1999.
2. Matthew P. Aylett. *Stochastic Suprasegmentals*. PhD thesis, University of Edinburgh, 2000. http://www.cogsci.ed.ac.uk/∼matthewa/thesis_sum.html
3. Anne H. Anderson et al. The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366, 1991.
4. J.E. Flege. Effects of speaking rate on tongue position and velocity of movement in vowel production. *The Journal of the Acoustical Society of America*, 84:901–916, 1988.
5. Björn Lindblom. Explaining phonetic variation: a sketch of the H & H theory. In William J. Hardcastle and Alain Marchal, editors, *Speech Production and Speech Modelling*, pages 403–439. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
6. Stefanie Shattuck-Hufnagel and Alice E. Turk. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25:193–247, 1996.
7. R. M. Uchanski, S. Choi, L.D. Braida, C. M. Reed, and N. I. Durlach. Speaking clearly for the hard of hearing IV: Further studies in the role of speaking rate. *Journal of Speech and Hearing Research*, 39:494–509, 1996.