The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech*

Matthew Aylett

University of Edinburgh and Rhetorical Systems Ltd., Edinburgh

Alice Turk

University of Edinburgh

Key words

articulation

duration

prosody

redundancy

speech

Abstract

This paper explores two related factors which influence variation in duration, prosodic structure and redundancy in spontaneous speech. We argue that the constraint of producing robust communication while efficiently expending articulatory effort leads to an inverse relationship between language redundancy and duration. The inverse relationship improves communication robustness by spreading information more evenly across the speech signal, yielding a smoother signal redundancy profile.

We argue that prosodic prominence is a linguistic means of achieving smooth signal redundancy. Prosodic prominence increases syllable duration and coincides to a large extent with unpredictable sections of speech, and thus leads to a smoother signal redundancy.

The results of linear regressions carried out between measures of redundancy, syllable duration and prosodic structure in a large corpus of spontaneous speech confirm: (1) an inverse relationship between language redundancy and duration, and (2) a strong relationship between prosodic prominence and duration.

The fact that a large proportion of the variance predicted by language redundancy and prosodic prominence is nonunique suggests that, in English, prosodic prominence structure is the means with which constraints caused by a robust signal requirement are expressed in spontaneous speech.

* *Acknowledgments:* We would like to thank Ellen Bard for helpful discussion of ideas presented in this paper, and Dan Jurafsky, Rena Krakow, Bjorn Lindblom, for useful comments on previous versions.

Address for correspondence. Dr. Matthew Aylett, Rhetorical Systems Ltd., Crichton's Close, Canongate, Edinburgh EH8 8DT; e-mail: <matthewa@cogsci.ed.ac.uk>.

1 Introduction

It has long been recognized that the production of speech sounds of a given language is characterized by a large degree of between and within-speaker variability. Much of speech research in the last half-century has been devoted to the description of this variability, and to the development of theories which predict where, when, and why particular variants of segments occur. The present paper focuses on two types of within-speaker variability: Variability resulting from differences in predictability between elements in a given speech string (i.e., differences in redundancy), and variability resulting from differences in prosodic context.

We argue that the acoustic consequences of differences in redundancy can be explained functionally within an information theoretic framework, by the drive for speakers to achieve robust information transfer in a potentially noisy environment while conserving effort. We argue that these pressures encourage speakers to produce utterances whose elements have similar probabilities of recognition (i.e., utterances with a smooth signal redundancy profile). We present supporting evidence showing that phrase-medial syllables with high language redundancy (i.e., highly predictable from lexical, syntactic, semantic, and pragmatic factors) are shorter than less predictable elements. We then show that variability in duration due to language redundancy is strongly related to variability associated with prosodic prominence.[1] We argue that a strong, but imperfect relationship between language redundancy and prosodic prominence is to be expected if prominence is the means speakers use to implement a smooth signal redundancy profile.

1.1 *Background*

It is well documented that syllable duration is related both to predictability due to linguistic and discourse context and to prosodic prominence structure. Many researchers have observed that less predictable elements in utterances tend to be articulated more carefully, and more slowly as a result, than more predictable elements (Balota, Boland, & Shields, 1989; Bard & Aylett, 1999; Fowler & Housum, 1987; Fowler, Levy, & Brown, 1997; Goldinger & Summers, 1989; Hunnicut, 1985; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985; Samual & Troicki, 1998; Sotillo, 1997; Wright, 1997). For example, Lieberman (1963) found that “nine” in the utterance “I would like nine please” is longer and more carefully articulated than “nine” in the well-known proverb “a stitch of time saves nine”. Along the same lines, prosodically prominent elements (either phrasally or lexically stressed), tend to be longer, louder and articulated with more care than unstressed elements (Beckman, 1986; Fry, 1955; Sluijter, 1995; Turk & Sawusch, 1997; Turk & White, 1999) (see Shattuck-Hufnagel & Turk, 1996, for a review).

[1] We use the term prosodic prominence to refer to phonological stress at all levels. The term therefore includes both lexical and phrasal stress.

We propose that an inverse relationship between syllable duration and predictability arising from lexical, syntactic, semantic and pragmatic factors (what we term language redundancy) is to be expected, since it provides an efficient way of ensuring that elements with low levels of language redundancy are produced for a longer period of time and perhaps with more salient acoustic characteristics, and will thus be likely to be recognized. Of course, another way of ensuring that elements with low levels of language redundancy are recognized correctly would be for speakers to produce all elements in an utterance with maximal duration and clarity. However producing speech in this way would be highly inefficient compared to producing more predictable elements with reduced duration and effort.

Proponents of Information Theory (Pierce, 1961; Shannon, 1948) argue that the most efficient way of ensuring robust information transmission in noisy environments is to have smooth signal redundancy (which can be thought of in terms of a smooth distribution of the probability of recognition) throughout an utterance. To understand why this should be so, let us consider a simple example, which makes a few simplifying assumptions for illustration's sake. Let us compare strings of elements whose overall probability of recognition, p(recognition), add up to 1: string AB, whose elements each have p(recognition) of .5, and thus has a smooth redundancy profile, and string CD whose elements have p(recognition) of .25 and .75 respectively, and thus does not have a smooth redundancy profile. Because the probability of correct recognition of both elements in the string in the correct order is a product of the p(recognition) of each element in the string (assuming that the p(recognition) of each element is independent of the other),the probability of correct recognition of string AB is (.5 * .5) = .25, whereas the probability of correct recognition of string CD is lower: (.25 * .75) = .1875. For an utterance to have a smooth *signal* redundancy profile, each element in the utterance must therefore have the same inverse relationship between language and acoustic redundancy.

To summarize, studies in the literature which show an inverse relationship between language redundancy and measures of care of articulation support our hypothesis that acoustic redundancy (e.g., longer durations, more salient gestures) compensates for low levels of language redundancy in such a way that signal redundancy (probability of recognition) is evenly distributed throughout each utterance.

This idea suggests that the well-documented relationship between prosodic prominence and articulatory and/or acoustic attributes such as care of articulation or durational differences might originate from an association between prominence and language redundancy. Indeed, such a relationship appears to exist. Pan and Hirschberg (2000) showed that phrasal stress placement in English is predictable in part from word bigram predictability. And furthermore, in many languages, focused, and hence, less predictable, elements tend to be phrasally stressed (see Ladd, 1996, for a review). Lexical stress patterns can also be related to predictability in many cases. For example, in English, stress falls on the first syllable of approximately 70% of content words in spontaneous speech, where segments are relatively difficult to predict, compared to segments that occur later in the word (Cutler & Carter, 1987). In addition, the reduction of monosyllabic function words in languages like English can also be attributed to their high syntactic redundancy and frequency of usage, and the lack of reduction of open class, content words may reflect their low syntactic redundancy and relatively infrequent usage.

The relationship between redundancy and prosodic prominence is clearly imperfect, however. For example, many English words have noninitial lexical stress despite the fact that their initial syllables may be difficult to predict, and phrases regularly exhibit default "broad" focus stress patterns in which nuclear phrasal stress falls near the end of the phrase (see Cruttenden, 1986; Ladd, 1996), regardless of the predictability of the words which bear nuclear stress. These examples suggest an independent role for prosody in the speech production process.

What we propose is that speakers use prosodic prominence as a means of implementing an inverse relationship between language and acoustic redundancy so that the resulting signal redundancy is smooth throughout an utterance. On this view, the imperfect correspondence between prominence structure and redundancy is a reflection of both the indirect way in which redundancy influences the acoustic signal and learned, language-specific conventions about stress placement.

The Smooth Signal Redundancy Hypothesis

A There is an inverse relationship between language redundancy and acoustic redundancy (as manifested by syllable duration).

B Prosodic prominence smoothes signal redundancy by controlling syllabic duration.

To help clarify part B) of the hypothesis, it is useful to compare what could be regarded as a traditional view of prosodic implementation with the model suggested by this hypothesis. Figure 1 is modified from a similar figure in Shattuck-Hufnagel & Turk (1996) and shows a traditional view of prosody's role in speech production. In this model, a set of different factors controls the influences of prosodic structure on the phonetic characteristics of utterances.

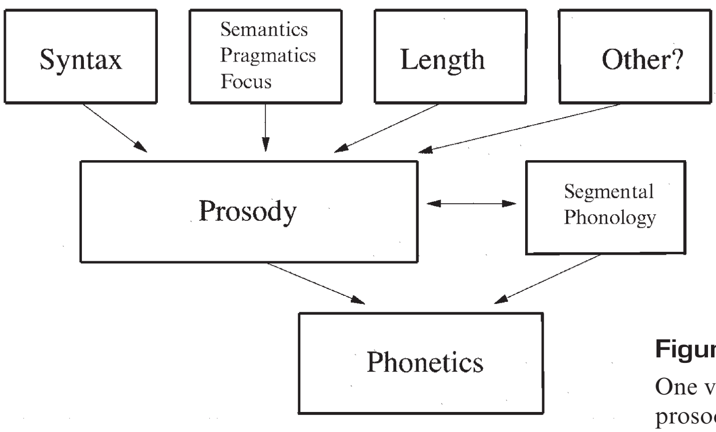

Figure 1

One view of the role of the prosodic component of the grammar (based on Shattuck-Hufnagel & Turk, 1996)

In contrast, Figure 2 shows how smooth signal redundancy could be achieved. The components affecting prosodic structure are language redundancy and robust communication rather than a set of different factors such as syntax, semantics, speech rate, and so forth. These two components are then encoded into prosodic structure in order to make the signal redundancy smooth and the communication robust. Note that robust communication includes both checking, which relates to prosodic boundary implementation, and smooth signal redundancy. A discussion of checking is beyond the scope of this paper, readers are referred to Aylett (2000) for a discussion of this topic.

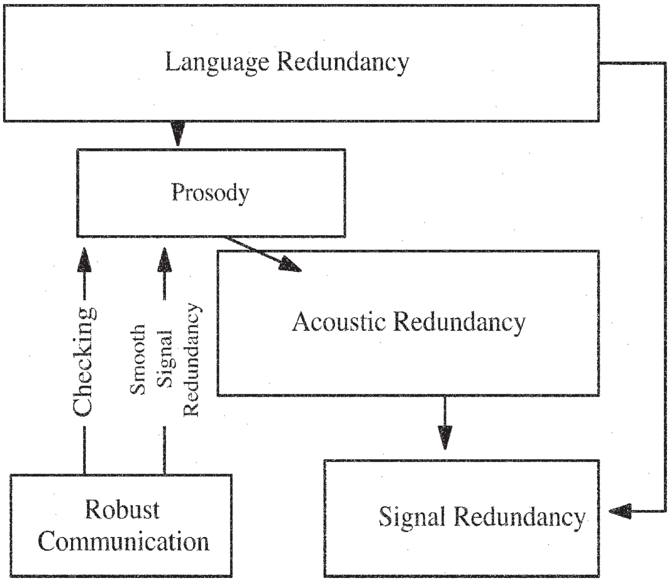

Figure 2
The smooth signal redundancy hypothesis

Despite the apparent fundamental differences between these models, they can be related to each other. In Figure 3, the traditional prosodic model is amalgamated with the smooth redundancy model. This amalgamated diagram highlights a difference between the meaning of the arrows in the original diagrams. In Figure 1, the arrows represent the processes in a production model. For example, if a major syntactic boundary occurs in the utterance to be produced, the language system adapts its prosodic structure accordingly. In Figure 2, the arrows represent more general conditioning processes, which could either reflect an historical influence of language redundancy on prosodic structure, or a direct, on-line influence of the type shown in Figure 1. For example, lexical stress, a prosodic factor, will tend to be word-initial as a result of redundancy factors which played a role in the evolution of the lexicon and the English prosodic system. In contrast, phrasal prosodic factors, such as phrasal stress/accent placement, are more likely to be conditioned by redundancy factors in a similar way to the factors that are shown to condition prosody in Figure 1. However, it is often unclear to what extent these factors directly alter prosodic structure during production and to what extent prosodic phonology has already evolved to take such factors into account. For example, reference reuse may be sufficient to cause de-accenting

(Ladd, 1996, p175), without the need to calculate, on-line, the actual redundancy of the repeated word in a particular context given the dialog structure.

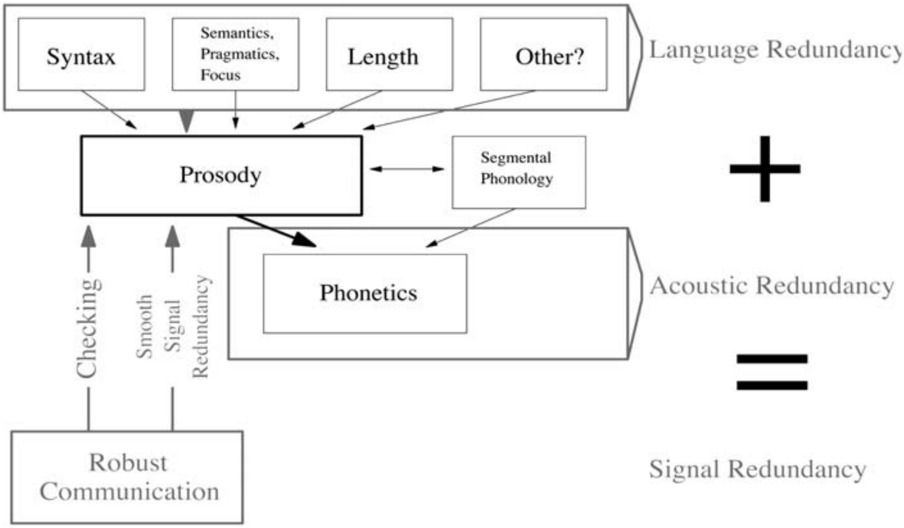

Figure 3

How the smooth signal redundancy model could be amalgamated with more traditional views of prosody (based on Shattuck-Hufnagel & Turk, 1996)

Despite these complexities, the diagrams do help illustrate the potential relationship between a redundancy-based model and a more traditional model of prosody's role in speech production.

In this paper, we test the smooth signal redundancy hypothesis by examining the relationships between prosodic prominence structure, language redundancy, and syllabic duration, and address the following questions:

1. To what extent does language redundancy relate to syllable duration?
2. To what extent does prosodic prominence relate to and thus arguably control syllable duration?
3. Does language redundancy relate to syllable duration independently of prosodic prominence?

The more prosodic prominence predicts the same changes in syllabic duration as language redundancy, the more convincing the argument that prosodic prominence affects these changes. If the strong version of part B) of the smooth signal redundancy hypothesis holds, we expect prosodic structure to account for all of the variance in syllable duration accounted for by redundancy.

Questions 1 and 2 have been addressed to some extent in the literature, but never using the same materials, as is crucial for answering Question 3. Furthermore, most studies of redundancy and prominence have focused on single factors, for example, contextual predictability factors or a single type of prominence, for example, either phrasal or lexical stress. Our hypotheses require materials which are coded for redundancy and prominence at many levels; we attempt to provide a comprehensive model of prominence structure and redundancy for our materials, within practical limitations.

2 Method

Although many studies have already shown the influence of selected redundancy factors on duration and other variables and other studies have shown the influence of prominence on duration, it is still unclear whether these two types of variables actually account for the same type of variability. Because of this, our study tests the effect of redundancy and prominence factors on the same materials, so the influences can be compared. In addition, because most current studies of these factors' relationship with duration are based on laboratory speech, and offer only limited opportunity for variation in multiple levels of redundancy and prominence factors, we chose to use a large corpus of task-oriented dialogs between Glaswegian English speakers, the HCRC Map Task Corpus (Anderson et al., 1991). The corpus consists of about 15 hours of spontaneous speech spoken by 64 speakers, and contains approximately 200,000 syllables. As described below, a combination of hand and automatic prosodic coding together with language redundancy and duration metrics were applied to this material. Every data point in our analysis thus represents an individual syllable with prosodic, redundancy, and duration information associated with it. Several multiple linear regressions with the duration metrics as dependent variables were used to test the predictive power of both redundancy and prosodic factors. Prosodic and combined redundancy / prosodic analyses were conducted on the whole corpus and on a phrase-medial subset of the corpus, in order to assess and control for phrase-boundary-related influences on syllable duration. Redundancy analyses were conducted on a phrase-medial subset alone. In the combined prosodic / redundancy analyses, we also assessed differences in the unique contribution of prosodic prominence and redundancy to our statistical models.

In the following sections, we describe the syllabic duration measurements we used as dependent variables, as well as the measurements we used for redundancy, and our prosodic structure coding scheme.

2.1 *Duration*

Over 70% of all syllable boundaries in our corpus were hand-segmented as a result of hand segmentation of word boundaries, where 60 – 70% of words in our corpus were monosyllabic. Word-medial syllable boundaries were determined using an HMM model for phone auto-segmentation, where phones were subsequently combined into syllables using the maximal onset principle. Word-medial syllable boundaries had an error of less than 30 ms compared to hand segmented boundaries.

We included two durational measures in our analyses: one raw syllable duration measurement in milliseconds (DUR1), and one measurement normalized for syllabic content (DUR2). This second measure (DUR2) was developed to account for factors such as the dependency of segmental duration on segmental identity (e.g., high vowels tend to be shorter than low vowels, and voiced stops tend to be shorter than fricatives), and on the number of segments contained within each syllable. Normalizing for segmental identity turned out to be problematic because segmental identity covaries to a large extent both with word frequency, and with the number of segments within each syllable. For example, the characteristic mean duration of /ð/ observed in our corpus can be attributed to the segment itself, to the high frequency of words such as ‘the’, or to the fact that 94% of syllables containing /ð/ are two or three segments in length (as opposed to 53% of syllables containing /s/ and 73% of syllables containing /k/). Because it is extremely difficult to model the independent contributions of these factors given problems of data sparsity (see Campbell, 1992; van Santen, 1998; Wightman, Shattuch-Hufnagel, Ostendorf, & Price, 1992, amongst others), and because the covariance of segmental identity with number of segments in each syllable turned out to be widespread, we chose to normalize only for the number of segments in each syllable, and to ignore segmental identity. This choice was motivated in part by the fact that some of our segmentation choices introduced further durational complications as far as segmental identity was concerned: Because vowels following stop consonants were judged to begin at voicing onset, vowels following voiceless, aspirated stops were shorter than vowels following voiced stops. Leaving segmental identity out of our model meant that we could ignore this type of segmental contextual interdependency. Support for this approach can be found in Aylett and Bull (1998), who found that a durational model based on the number of segments within each syllable predicted durational variation across speakers better than a model based on individual segment duration distributions.

The normalization process we used was as follows:

- Phonemic content was ignored. Each phoneme was regarded as being identical, that is, as having the same log distribution, $\mu = -2.7478$ (64 ms), $\sigma = 0.5702$ (-1 *SD* = 36 ms, $+1$ *SD* = 113 ms), representing its characteristic duration. This distribution was calculated on the basis of hand segmentation of two map task dialogs and represents the variance and mean of all phonemes.

- The number of segments in a syllable were associated with a characteristic expansion of the syllable’s length (see Table 1). These expansion factors were determined on the basis of experimental results presented in Campbell (1992).

- The overall expected duration of a syllable was calculated by multiplying the mean of the segmental log distribution by the number of segments in the syllable and the expansion factor associated with this number of segments.

- The normalized measurement was then arrived at by calculating how much this expected duration needed to be reduced or increased in order to be equal to the observed duration. Rather than describing this in percentage terms (say 130% longer than expected) it was described in terms of the number of *SD*s each segment would need to be altered away from the log mean in the log domain for the syllable to be this length. The effect of this is to treat each syllable as a set of connected springs, each representing a segment, and the normalized score can be regarded as the amount of force required to expand or compress the spring to fit an observed length. This chained log distribution approach is based on Campbell and Isard (1991) (see Equation 1).

Equation 1

$$d = \sum_{i=1}^{n} \exp^{(\mu(i)+k\sigma(i))} M$$

where: n = the number of segments in a syllable,
k = our normalized duration score,
μ = the mean log duration of a segment(–2.7478 (64 ms)),
(σ = the SD of the log distribution of a segment's duration (0.5702 (– 1 SD = 36 ms, + 1 SD = 113 ms))
M = a segmental multiplier (see Table 1)

TABLE 1

Multipliers for different number of segments in a syllable. For example if a segment is in a three segment syllable the multiplier is 1.00, if it is in a four segment syllable the multiplier is 0.93 (see Equation 1). These multipliers are derived from duration results presented by Campbell (1992)

	Multipliers				
Number of Segments	1	2	3	4	5 +
M	1.60	1.14	1.00	0.93	0.87

2.2 *Redundancy*

The predictability of a syllable in running speech is dependent on many factors. Without understanding all the dependencies between semantics, syntax, pragmatics and the structure of language any measure of redundancy is an approximation. 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 syllable level (syllabic trigram probability), at the word level (log of word frequency) and also at the discourse level (order of mention of referents). These measurements will give a representative, robust and simple measure of redundancy allowing a large scale quantitative corpus analysis.

Studies in the literature suggest that similar measures to the ones we use here affect durational characteristics. Fidelholtz (1975) and Booij (1995) both argue for a word frequency effect on vowel reduction in content words, in a direction pointing towards shorter durations for high frequency words as compared to low frequency words. Bell et al. (2002) offer experimental support for this view from segment durations. Pan and Hirschberg (2000) showed that word bigram predictability affected phrasal stress placement (and presumably duration), and many researchers have found that givenness reduces duration (e.g., Bard & Aylett, 1999; Fowler & Housum, 1987).

Independent redundancy variables used in the analysis. The redundancy variables used in the *Results* section and the names we have assigned to them are listed below.

wf: Log of Word Frequency. More frequent words should be more easy to predict and thus more redundant. Each syllable was associated with the COBUILD word frequency of the word it was part of (see Baayen, Piepenbrock, & Gulikers, 1995 for details).

trigram: Syllabic Trigram Measurement. Using the spoken part of BNC (British National Corpus)[2] the transitional probability of guessing a third syllable on the basis of the first two was calculated. The CMU-Cambridge toolkit was used to calculate trigram probability using Good Turing and backoff.[3] This measurement gives some idea of predictability produced by frequent sequences of words and the redundancy in later syllables of polysyllabic words. Together with word frequency this measurement gives a more interword sense of redundancy.

men: Givenness. Both word frequency and trigram measurements can be regarded as low level measures of redundancy, in that they take no consideration of the meaning in language or of the flow of meaning in a stream of speech. In contrast, givenness is related to the introduction of a referent in a dialog. The more this referent is mentioned the more "given" it becomes. This final measurement of redundancy measures how many times a referent (in this case a landmark on a map e.g., "white mountain," "east lake," "it"—when referring to a landmark) has been mentioned.

2.3
Prosodic structure coding

Our materials were coded for prosodic prominence (stress) and boundary structure at the word and phrase level using a combination of hand and automatic techniques (described below). We assume a theory of prominence structure which incorporates distinctions between reduced and full vowels, lexically stressed versus unstressed syllables, and nuclear versus non-nuclear phrasal stresses/accents (cf. Ladefoged, 1982; Cruttenden, 1986). We assumed a culminative prominence hierarchy, that is, that only full vowels bear lexical stress, that only lexically stressed syllables bear phrasal stress/accent, and that only phrasally-stressed/accented syllables can bear primary (nuclear) phrasal stresses/accents. Most of these distinctions have been shown to have a measurable effect on duration: lexically stressed vowels are longer than unstressed vowels, even when unaccented (Sluijter, 1995), and syllables which bear phrasal stress are longer than those without phrasal stress (Beckman, 1986, among others).

2 The BNC is designed to represent as wide a range of modern British English as possible. The corpus comprises 4,124 texts, of which 863 are transcribed from spoken conversations or monologs. <http://info.ox.ac.uk/bnc/what/index.html> for details.

3 Good Turing and backoff are means of dealing with data sparsity and missing contexts when calculating trigram statistics (see Clarkson & Rosenfeld, 1997 for details).

Furthermore, Wightman et al. (1992) and Beckman and Edwards (1990) suggest that nuclear accented syllables are longer than non-nuclear accented syllables. The durational behavior of lexically unstressed reduced versus full vowels is not well-documented, but we have reason to believe that these would also show durational differences, if only because of differences in vowel target location (Lindblom, 1967).

We assume a theory of prosodic constituent structure containing four levels of nested constituents, as embodied in the ToBI break-index convention. These four levels correspond roughly to the four levels of Prosodic Word, Minor Phrase, Major Phrase, and Full Intonational Phrase in the Selkirk (1978) theory of prosodic constituent structure, and the highest two levels correspond to the (Beckman & Pierrehumbert, 1986) theory of Intermediate and Full Intonational phrases. Because preboundary lengthening is well-documented at most levels of the prosodic hierarchy, see Ladd and Campbell (1991), and Wightman et al. (1992), and many phrase-initial segments are longer than phrase-medial segments (Fougeron & Keating, 1997), we included these boundary factors in regression analyses of duration on prosodic variables, and controlled for these factors when studying the effects of redundancy on duration.

When comparing the influence of prosodic and redundancy factors on duration, we ran analyses twice, once on all materials (including phrase-edge syllables), and once on phrase-medial syllables. Because our Smooth Signal Redundancy Hypothesis predicts a relationship between prosody and redundancy for prosodic prominence, but not for prosodic boundaries, we expect a large unique contribution of prosodic factors when all materials are included but a much smaller contribution when only phrase-medial syllables are considered.

Automatic coding. The existing word segmentation and transcription of the HCRC map task corpus was used as the basis for phoneme and syllable segmentation as well as for the coding of lexical stress. The CELEX on-line dictionary (Baayen, Piepenbrock, & Gulikers, 1995), which contains canonical phonemic representations for each word was used to guess the probable segmental contents of each word. Lexical stress and syllabification based on the maximal onset principle were subsequently assigned as indicated in the CELEX on-line dictionary, and a hidden Markov model (HMM) speech recognizer with a model for each segment pretrained from previous speech was used to posit the likely boundaries of each phoneme. Word boundaries coinciding with pauses were labeled automatically as full intonational phrase boundaries.

Hand coding. 679 full intonational phrases containing 3190 words were also coded for constituent boundaries (break indices 0–4) and accents/phrasal stresses using GlaToBI (Mayo, Aylett, & Ladd, 1997). GlaToBI is a variant of the ToBI tone and break index coding system which was adapted for the Glaswegian accent of Scottish English (Beckman & Ayers, 1993). Differences between ToBI and GlaToBI are primarily differences between accent and boundary tone type, and are irrelevant to the work reported here.

Independent prosodic variables used in the analysis. The following prosodic variables were used to predict syllable duration variation in our analyses.

- Prosodic boundaries: Binary variables[4]

  wboun: Prosodic word boundary corresponding to a ToBI break index of 1.

  p2boun: Minor Phrase corresponding to a ToBI break index of 2.

  p3boun: Major Phrase/Intermediate Intonational Phrase corresponding to a ToBI break index of 3.

  ipboun: Full Intonational Phrase Boundary corresponding to a ToBI break index of 4.

  Aipboun: Automatically coded Full Intonational Phrase Boundary. For nonhand coded materials, syllables followed by a pause were coded as being followed by a full intonational phrase boundary.

These boundary variables are hierarchical. Thus if **ipboun** is 1 meaning a full intonational phrase boundary is present after a syllable **wboun, p2boun** and **p3boun** must also be set to 1.

- Prominence: Binary variables

  vtype: Vowel type. This variable indicates whether the vowel is full or reduced (where reduced equals /ɪ, ə/ in lexically unstressed syllables) and corresponds to the first level of prominence described by Ladefoged (1982).

  lexstr: Lexical stress. This variable indicates whether the syllable is lexically stressed and corresponds to the second level of prominence described by Ladefoged and the first level of prominence as described by Cruttenden (1986). (**lexstr** is not strictly a binary variable, as primary lexical stress is coded as a 1 and secondary stress is coded as 0.5.). lexstr differs from **vtype** because some full vowels can be unstressed, for example, /i/ in *spongy* /spʌndʒi/.

  acc: Phrasal Accent. This variable indicates whether a phrasal accent has been marked using ToBI and corresponds to the second level of prominence as described by Cruttenden (1986).

  Aacc: Automatically coded Phrasal Accent. For nonhand coded materials, if the syllable was lexically stressed and open class, it was marked as having a high likelihood of having a phrasal accent.

  ppa: Primary Phrasal Accent. The last accent before an intermediate or full intonational phrase boundary (as coded using ToBI) is marked as having primary phrasal stress. This corresponds to the third level of prominence described by Ladefoged (1982) and the third level of prominence as described by Cruttenden (1986). Automatic coding of primary phrasal stress was considered too unreliable.

---

[4] Previous work has shown a relationship between total number of syllables in a word and syllabic duration (e.g., Campbell, 1992). However it was found that if this was included as a factor results were confounded by word and syllable boundary factors (**wboun** = 0/1) (88% of words in the corpus are 1 or 2 syllables in length), as well as the strong negative correlation between total number of syllables in a word and its word frequency. For these reasons the number of syllables factor was removed from this analysis in favor of lexical stress and word boundary information (see Aylett, 2000 for details)

As with the boundary variables, prominence is also in a hierarchical relationship. For example if **lexstr** is one, meaning the syllable is lexically stressed, **vtype** must also be one because reduced vowels cannot occur in lexically stressed syllables. Similarly if **ppa** is one indicating a primary phrasal accent, all other prominence variables are also set to 1.

- Spillover

  spill: This factor is based on work by Turk and Sawusch (1997) and Turk and White (1999), and is used in hand coded prosodic data only. It represents the amount by which durational effects of prominence spill over from an accented syllable onto adjacent syllables. This is mostly in a rightward direction (20%), leftwards by much less (5%), within a word. Across a word boundary a spill of 4% is reported in a rightward direction.

2.4
Materials

Each data point in this analysis is a syllable from the HCRC Map Corpus (Anderson et al., 1991). The syllables are coded with prosodic, redundancy and duration factors. As shown in Figure 4, not all of the syllables in the corpus were coded with the same scheme:

- Any syllables which were not coded for redundancy factors (such as syllables forming words unknown to the BNC corpus) were ignored. Just under 10% of the data (18225 syllables) were removed from the analysis.
- A proportion were prosodically hand coded, so, in addition to the lexical and automatic factors, these syllables were also coded for accent and break index.
- A proportion of syllables, those that are within references to landmarks on the maps used in the dialogs, were also coded for mention.

The overlap between these groups and the entire data set is shown in Figure 4.

To control for the effects of prosodic boundaries on duration, a separate analysis was carried out on a set of syllables for which break index was controlled. In particular, we used only nonphrase-final monosyllabic words that had a following break index of 1. Table 2 shows the number of materials for all possible conditions.

TABLE 2

Number of syllables in each condition

	Total Coded (All)	*Mention Coded (M)*	*Prosodic Coded (P)*	*Mention + Prosodic (P∪M)*
DUR Coded	169464	30508	3638	1521
DUR Coded + Prosodic Boundaries Controlled	89532	12295	1186	205

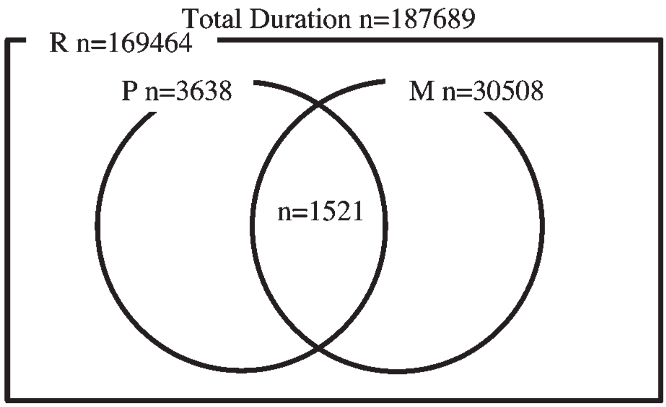

Figure 4

Materials examined in the analysis for the Duration experiment. R: The number of syllables with valid word frequency and trigram information. P: The number of syllables with hand coded prosodic factors. M: The number of syllables coded for mention. (18225 syllables were removed from the analysis because there was no dictionary entry in CELEX, e.g., disfluencies)

2.5 *Statistical Analyses*

The influence of prosodic and redundancy factors on measures of syllabic duration was assessed using multiple regression analyses. Unique contributions of individual factors to the regression models were determined by subtracting the predictive power of the linear regression without the factor(s), from the predictive power of the regression with the factor(s).

Significance of the unique contributions was assessed using maximum likelihood (also termed the likelihood ratio test (Neter, Wasserman, & Kutner, 1990)). The shared contribution of redundancy and prosodic factors was determined by subtracting their unique contributions from the total predictive power of the model (the total r^2).

3 Results

3.1 *Does redundancy condition syllabic duration?*

Table 3 shows that redundancy factors predict between 20% and 65% of variation in syllable duration in nonphrase-final monosyllabic words (followed by a break index of 1), depending on the set of materials under investigation and whether the dependent variable is DUR1 or DUR2. Not surprisingly, redundancy factors predicted more of the variance in duration when mention was included as an independent variable (for mention-coded materials, of course). Figure 5 shows an inverse relationship between all three redundancy factors and duration, as predicted by the Smooth Signal Redundancy Hypothesis: As redundancy increases, duration decreases.

TABLE 3

Regression analysis of redundancy factors. These results are for syllables in phrase medial monosyllabic words followed by a break index of 1

DUR1: Raw syllabic duration

	Regression Results	*$r = .6081$*	*$r^2 = 0.3698$*
Redundancy Factor	Unique Contrib. to r^2	F(1,89531)	p value
wf	10.11%	14361.29	.001
trigram	01.93%	2736.84	.001

DUR1: Raw syllabic duration: Mention coded materials only

	Regression Results	*$r = .8085$*	*$r^2 = 0.6536$*
Redundancy Factor	Unique Contrib. to r^2	F(1,12294)	p value
wf	06.06%	2150.52	.001
trigram	00.74%	263.66	.001
men	00.33%	116.28	.001

DUR2: Normalized syllabic duration

	Regression Results	*$r = .4250$*	*$r^2 = 0.1806$*
Redundancy Factor	Unique Contrib. to r^2	F(1,89531)	p value
wf	04.44%	4850.10	.001
trigram	01.21%	1322.12	.001

DUR2: Normalized syllabic duration: Mention coded materials only

	Regression Results	*$r = .6603$*	*$r^2 = 0.4360$*
Redundancy Factor	Unique Contrib. to r^2	F(1,12294)	p value
wf	05.45%	1187.79	.001
trigram	00.11%	24.75	.001
men	00.90%	196.06	.001

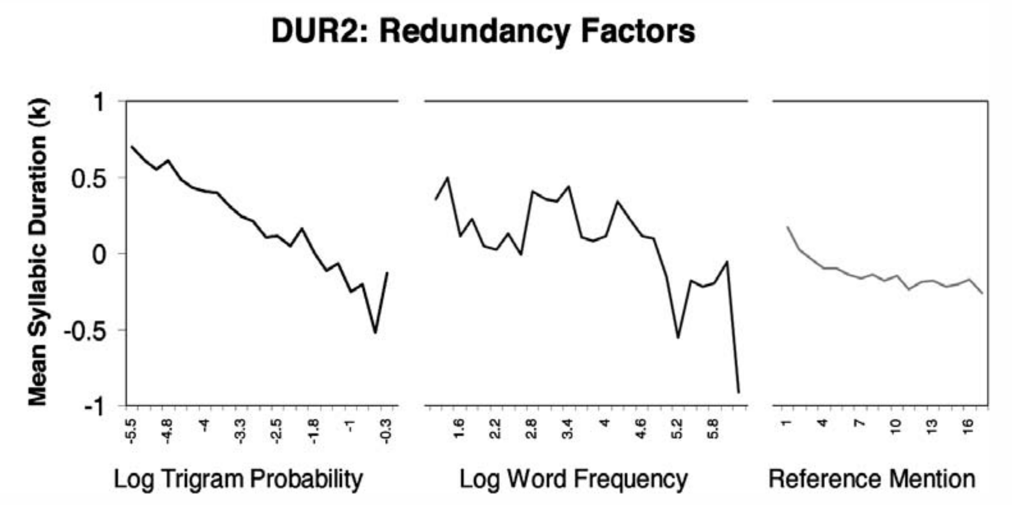

Figure 5

The relationship between redundancy factors and DUR2. Redundancy increases from left to right. Trigram and word frequency factors are calculated over the entire corpus whereas mention is calculated over only mention coded materials

3.2
Does prosody condition syllabic duration?

As expected, a strong relationship was found between durational measures and prosodic structure. As we can see from the overall regression results in Table 4, hand-coded prosodic factors predict between 50–60% of the variation in syllabic duration, depending on whether raw or normalized duration is used as the dependent variable. Most individual factors showed significant unique contributions to the regression model, in line with predictions from previous studies (e.g., Beckman & Edwards, 1990; Sluijter, 1995; Turk & White, 1999; Wightman et al., 1992). However, we observed a large degree of interdependence between prosodic factors, as evidenced by the low unique contribution of individual factors. For example, as shown in Figure 6, it is clear that prosodic boundary strength and degree of prominence have similar (lengthening) effects as far as syllable duration is concerned. Durational results from the whole corpus using automatically coded prosodic variables were similar (see Table 5). These results offer support for the Smooth Signal Redundancy hypothesis, in the sense that the significant relationship between prominence factors and syllabic duration suggests that prominence can control information content. In addition, results for boundary factors highlight the need to control for these factors in analyzing effects of redundancy on duration.

TABLE 4

Regression analysis of hand coded prosodic factors against syllabic duration

DUR1: Raw syllabic duration

	Regression Results	*r = .7710*	*r^2 = 0.5944*
Prosodic Factor	Unique Contrib. to r^2	F(1,3638)	*p* value
vtype	00.09%	8.59	.01
lexstr	01.17%	105.52	.001
acc	03.30%	296.41	.001
spill	01.21%	109.12	.001
ppa	00.01%	0.90	NS
wboun	06.97%	625.73	.001
p2boun	00.72%	64.74	.001
p3boun	00.04%	3.98	.05
ipboun	00.15%	14.02	.001

DUR2: Normalized syllabic duration

	Regression Results	*r = .7238*	*r^2 = 0.5238*
Prosodic Factor	Unique Contrib. to r^2	F(1,3637)	*p* value
vtype	00.45%	34.85	.001
lexstr	04.54%	346.95	.001
acc	02.70%	206.79	.001
spill	02.23%	170.39	.001
ppa	00.00%	0.37	NS
wboun	07.46%	569.88	.001
p2boun	00.74%	56.92	.001
p3boun	00.00%	0.11	NS
ipboun	00.07%	5.85	.05

Figure 6

Effect of automatic prosodic factors (boundary and prominence) on DUR2 over the whole corpus

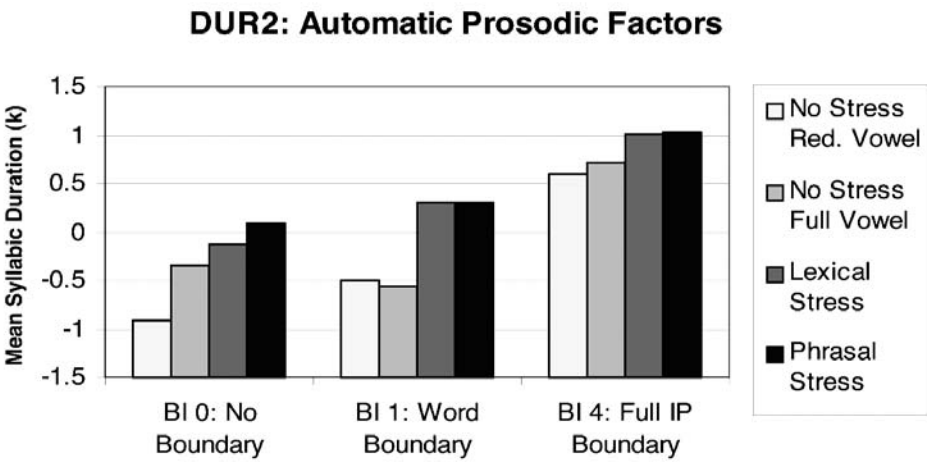

TABLE 5

Regression analysis of automatic coded prosodic factors with syllabic duration

DUR1: Raw syllabic duration

	Regression Results	*r = .6473*	*r^2 = 0.4190*
Auto Prosodic Factor	Unique Contrib. to r^2	F(1,169461)	*p* value
vtype	01.08%	3139.49	.001
lexstr	00.83%	2421.31	.001
Aacc	01.49%	4335.15	.001
wboun	03.62%	10561.72	.001
Aipboun.	19.72%	57523.91	.001

DUR2: Normalized syllabic duration

	Regression Results	*r = .6077*	*r^2 = 0.3693*
Auto Prosodic Factor	Unique Contrib. to r^2	F(1,169461)	*p* value
vtype	00.37%	997.12	.001
lexstr	03.31%	8901.64	.001
Aacc	00.03%	79.00	.001
wboun	01.46%	3926.77	.001
Aipboun.	13.10%	35208.99	.001

4 Further tests of the Smooth Signal Redundancy Hypothesis

4.1 *Relating redundancy and prosodic effects on duration*

Results presented in previous sections suggest that syllabic duration is affected both by prosodic and by redundancy factors. Support for the Smooth Signal Redundancy Hypothesis was found in the inverse relationship between redundancy and measures of syllabic duration and in the significant relationship between prosodic prominence factors and syllabic duration.

In this section, we present results from further tests of the Smooth Signal Redundancy Hypothesis; that is, tests designed to see to what extent prosodic prominence and redundancy are related. Part B of the Smooth Signal Redundancy Hypothesis states that redundancy smoothes care of articulation by controlling prosodic prominence, and thus predicts that most of the variance accounted for by prosodic prominence and redundancy should be shared. In particular, the hypothesis predicts that redundancy should not offer a unique contribution to a joint prosody / redundancy model. Because of language-specific, conventionalized aspects of prominence placement referred to in Section 1.1, we do expect to see a unique contribution of prosodic prominence, but we expect this contribution to be small since we argue that the function of prosodic prominence is primarily to implement redundancy factors.

The multiple regressions testing this prediction for durational measurements were run twice, once on all materials within each data subset shown in Figure 4, and once on materials in which prosodic boundaries were controlled for, that is, on phrase-medial monosyllabic words (followed by a break index of 1). We expected to see a greater unique contribution of prosodic factors in the case where both phrase-edge and phrase-medial materials were included in each data subset, since the close relationship between redundancy and prosody only applies to prosodic prominence, but not to boundaries. On the other hand, we predicted only a small contribution of prominence when boundaries are controlled for, since the Smooth Signal Redundancy Hypothesis claims that the primary function of prominence is to implement redundancy factors. Results of these analyses are shown in Table 6, and subsets of the results are shown in Figure 7. In both types of analyses, prosodic and redundancy factors together predicted between 30% and 68% of the variance in the duration measurements, depending on the type of analysis, the dependent variable (DUR1 or DUR2), and the subset of data being investigated. The fact that the proportion of variance in durational measurements accounted for by a combined redundancy / prosody model is only slightly higher than the proportion of variance accounted for by either prosodic or redundancy factors alone is consistent with the view that redundancy effects are implicitly expressed by prosodic prominence. And indeed, the contribution of redundancy factors to both types of model is small (around 5% on average).

TABLE 6

Independent contributions of redundancy—Red. and prosodic models—Pros. in predicting variance of DUR1 and DUR2 over all materials. The nonindependent, shared contribution is shown under Shared. P: Materials hand coded for prosody. M: Materials with mention coding. All: Materials with automatic prosodic coding and a trigram/word frequency redundancy model. P∪M: Materials both hand coded for prosody and for mention. All results are significant except for the redundancy model's contribution to hand coded and mentioned coded material with respect to DUR

DUR1/2: Independent and shared contribution

	All Materials				Boundaries Controlled			
	r^2	Unique		Shared	r^2	Unique		Shared
Materials		*Pros.*	*Red.*			*Pros.*	*Red.*	
DUR1: P	63.11%	39.52%	3.67%	19.92%	53.17%	12.10%	7.97%	33.10%
DUR1: P∪M	68.35%	38.52%	2.29%	27.54%	53.29%	4.73%	11.27%	37.29%
DUR2: P	53.80%	38.27%	1.42%	14.11%	33.25%	21.13%	1.80%	10.32%
DUR2: P∪M	61.91%	37.18%	1.67%	23.06%	31.44%	16.69%	2.31%^{ns}	12.44%
DUR1: All	49.01%	34.49%	7.11%	7.41%	41.44%	4.46%	9.70%	27.28%
DUR1: M	61.06%	29.76%	9.64%	21.66%	67.83%	2.47%	3.62%	61.74%
DUR2: All	40.22%	31.36%	3.29%	5.57%	29.68%	11.62%	2.84%	15.22%
DUR2: M	51.90%	24.83%	9.12%	17.95%	55.01%	11.41%	1.16%	42.44%
Average	56.17%	34.24%	4.78%	17.15%	45.64%	10.58%	5.08%	29.98%
% of Explained Variance		60.96%	8.5%	30.53%		23.18%	11.14%	66.69%

A comparison of the left and right columns in Figure 7 (and the corresponding figures in Table 6) shows that the unique contribution of prosody to the regression models was large when it included prosodic boundary information (in the All Materials analysis). This result was expected, since redundancy is hypothesized to share predictive power with prosodic prominence, but not with prosodic boundaries. The right side of Figure 7 supports this hypothesis: The unique contribution of prosody to the regression model was small (just over 10% on average, cf. Table 6), when prosodic boundaries were controlled for.

What can also be seen from Figure 7 is that regressions computed over different subsets of materials yielded results that were qualitatively similar in terms of the shared and independent predictive powers of prosody and redundancy. However, the relative magnitudes of the independent and shared contributions differed between dependent variables (DUR1 vs. DUR2), and between subsets of materials. These quantitative differences are difficult to explain, but are likely to depend to some degree

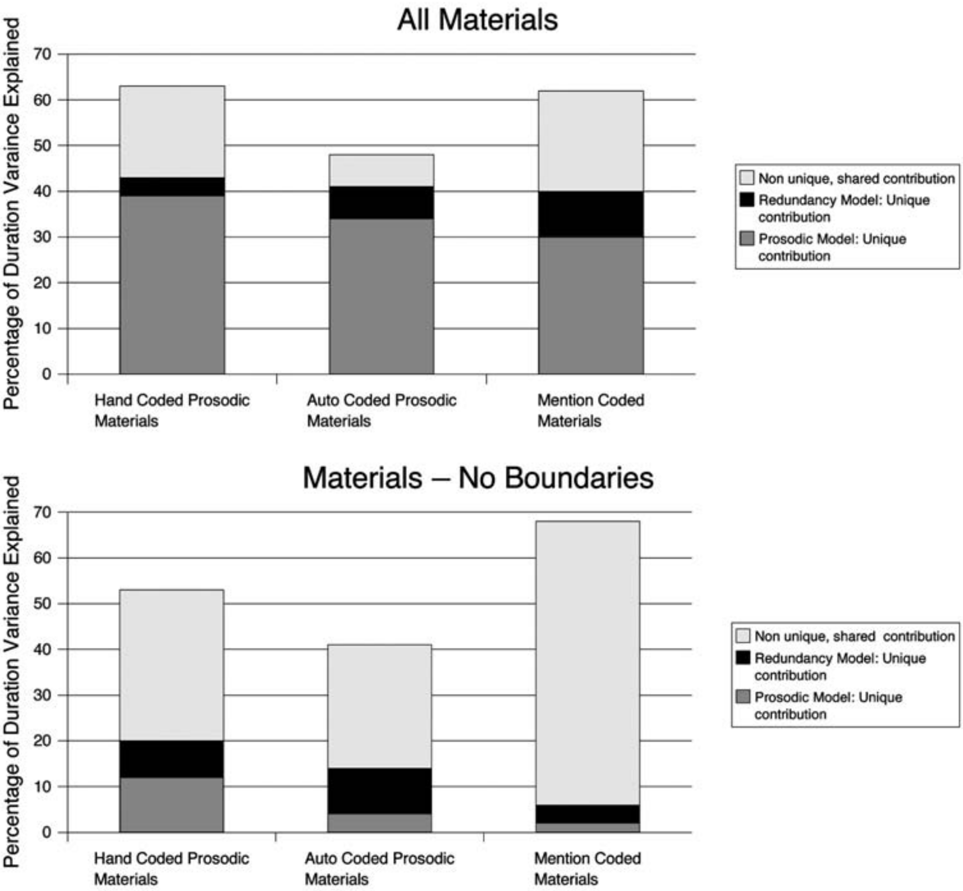

Figure 7

Effect of automatic prosodic factors (boundary and prominence) on DUR2 over the whole corpus

on the relative validity of our models. For example, the prosodic model was likely to have been the most valid for hand coded prosodic materials, and indeed, prosody made the biggest independent contribution to the regression model for these materials. On the other hand, the redundancy model was likely to be the most valid in the Mention Coded Materials subsets; but, in spite of this, redundancy had the smallest independent contribution for most of these subsets, apart from the P∪M intersection for the DUR1 measurement.

Since our results show (1) that redundancy factors account for a large percentage of variance in duration when prosodic boundaries are controlled for, and (2) that prominence factors account for much of the same variance, we have been largely successful in supporting the Smooth Signal Redundancy Hypothesis. However, the small, significant, unique contributions of redundancy factors in combined redundancy/prominence regression models of duration are difficult to account for, since we predicted that prominence should account for *all* of the variance accounted for by redundancy factors. Several possible explanations for these unique effects exist. One possibility is that our models and/or measurements were imperfect. Indeed, our redundancy models were quite simple: They included two, and at the most three out of a host of possible factors. The prosodic model used for our hand-coded data had strong theoretical motivation, and was likely to have been quite good, but was used only for a small subset of materials. Furthermore, the prosodic model we used was quantal in the sense that it predicts four distinct phonetic categories of prominence, and it is possible that speakers implement degrees of prominence (and redundancy) that are more fine-grained than this.[5] As for the measurements we used, the DUR2 measurement was normalized only for a single factor known to affect segment durations, and our DUR1 measurement was not normalized at all.

Another possibility is that the small independent contributions of redundancy factors would have remained, even with perfect models and measurements. On this view, an independent contribution of redundancy would require a direct link between redundancy and care of articulation in addition to the indirect link via prosodic structure.

Although we predicted unique contributions of prominence factors to our models, since we hypothesized that their unique contributions would represent the extent to which a language had phonologized or conventionalized the relationship between redundancy and prominence. However, it is possible that we would have found even smaller unique contributions with better measurements and/or prosodic models. Unfortunately, our present data cannot decide between these alternatives.

5 Summary of results

The results obtained from this work can be summarized as follows:

- Both prosodic factors and redundancy factors have a significant effect on syllabic duration in a large corpus of spontaneous running speech.

- Redundancy factors, looking at a subset of landmark referents with controlled prosodic boundaries, predicted 65% (Table 3) of raw syllabic duration change. Results for other materials varied. For all materials, trigram and word frequency factors predicted 14% (Table 3) of the variation. The more predictable a syllable in terms of low level factors such as word frequency and syllabic trigram and a higher level factor, givenness/mention, the shorter a syllable is.

- Prosodic factors predicted up to 59% (Table 4) of raw syllabic duration in hand coded materials. Results for automatic coding based on lexical information and pauses predicted 42% (Table 5) of the variation.

- Comparing the independent contribution of redundancy factors and prosodic factors to predicting duration it was found that (see Fig. 7, data from Table 6):

[5] We are grateful to Dan Jurafsky for this observation.

1. Most of the contribution made by redundancy factors is implicitly represented by prosodic prominence factors. However a significant but small percentage (5% on average) predicted even by these very simple redundancy metrics was not represented by prominence factors.
2. Prosodic factors, including both prominence and boundary factors, especially when hand coded, made a large unique contribution to predicting duration change above that representing redundancy (about 35% compared to a shared prosodic/redundancy contribution of 17% over all sets of materials).
3. This unique contribution of prosodic prominence factors was much smaller for syllables where prosodic boundaries were controlled for (11% over all sets of materials). This suggests a major role of prosodic prominence structure, without boundaries, is to smooth signal redundancy by controlling duration in a way which implicitly mirrors language redundancy factors. This result supports the Smooth Signal Redundancy Hypothesis.

6 Discussion

In this paper, we presented the Smooth Signal Redundancy Hypothesis as a functional explanation for the use of prosodic prominence (lexical and phrasal stress) in running speech. Results from analyses designed to test this hypothesis show (1) a significant effect of prosodic and redundancy factors on duration in a large corpus of spontaneous running speech, (2) an inverse relationship between redundancy and duration, (3) that a large proportion of the variance in durational measures is accounted for by prosodic and redundancy factors, in spite of the relative simplicity of our models, and (4) that the effects of prosodic prominence and redundancy are to a large extent shared, although small unique effects of both prominence and redundancy were observed.

Although we supported the Smooth Signal Redundancy Hypothesis using duration evidence alone, we believe that the same principles should apply to other measures of saliency/care of articulation, such as formant targets or even F0 excursions.

Our results are consistent with the view that duration, and perhaps care of articulation more generally, are used to compensate for low levels of language redundancy. They also suggest that prosodic prominence structure is used to implement redundancy differences by controlling care of articulation. We argue that the primary reason for these observed relationships is the need for efficient, robust information transmission in a potentially noisy environment. On this view, prosody acts as an interface between the compositional structure of language and the constraints of producing a robust and effective signal. Because we found small unique contributions of redundancy factors to our models that our hypothesis did not predict, we concede that some aspects of redundancy may not be controlled by prominence, but may be implemented more directly. However, as discussed earlier, it is also possible that these unique contributions would disappear with more appropriate models and/or measurements. In addition, we leave open the possibility that factors outside redundancy also might affect prominence structure, for example, psycholinguistic or phonological constraints, but argue that these constraints should be marginal given the importance of reliable information transmission.

A comparison of the more traditional view of prosody shown in Figure 1 and the combined model in Figure 2 highlights some additional advantages of our combined view. Firstly, it offers a unified principle for relating aspects as diverse as focus, word class, length of utterance and word frequency in terms of a predictive model, and thus in terms of redundancy. Furthermore, it offers an explanation for why some aspects of grammatical information affect prosodic prominence structure, and others do not: Only aspects of grammar which affect language redundancy should affect prosodic prominence. Aspects such as structural relations between syntactic constituents which do not necessarily affect language redundancy should not be reflected in prosodic prominence, although they might be reflected in prosodic constituent structure. Given the fact that the need for efficient, robust, information transmission is arguably universal, our framework could potentially shed some light on observed cross-linguistic differences in stress systems and phonetic realization. Differences in stress systems or in patterns of phonetic reduction/centralization are predicted to co-occur with related differences in language redundancy. For example, our hypothesis predicts that so-called syllable-timed languages should have smoother language redundancy profiles than so called stress-timed languages, since their syllables do not vary as much in terms of duration.

In addition, languages with looser constraints on word ordering could use position as a means of smoothing redundancy rather than prominence or duration. For example, rather than having a situation where repeated mentions are attenuated, languages could place repeated mentions further to the front of a phrase where they would be less predictable from context.

We do not begin to test these predictions here, but rather raise them as examples of issues that can be addressed using the Smooth Signal Redundancy framework.

Received: February 27, 2001; first revision received: October 07, 2002; second revision received: August 25, 2003; accepted: October 10, 2003

References

ANDERSON, A. H., BADER, M., BARD, E. G., BOYLE, E., DOHERTY-SNEDDON, G. M., GARROD, S., ISARD, S., KOWTKO, J. C., McALLISTER, J. M., MILLER, J. E., SOTILLO, C. F., THOMPSON, H. S., & WEINERT, R. (1991) The HCRC Map Task Corpus. *Language and Speech*, **34**(4) 351–366.

AYLETT, M., & BULL, M. (1998). The automatic marking of prominence in spontaneous speech using duration and part of speech information. In *Proceedings of ICSLP–98*, 2123–2126.

AYLETT, M. P. (2000). *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech* (http://www.cogsci.ed.ac.uk/matthewa/thesis sum.html). Ph.D. thesis, University of Edinburgh.

BAAYEN, R. H., PIEPENBROCK, R., & GULIKERS, L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. Version 2.5.

BALOTA, D., BOLAND, J., & SHIELDS, L. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, **28**, 14–36.

BARD, E., & AYLETT, M. (1999). The dissociation of deaccenting, givenness and syntactic role in spontaneous speech. In *ICPhs99*, 1753–1756.

BECKMAN, M. E. (1986). *Stress and non-stress accent*. Dordrecht-Holland: Foris Publications.

BECKMAN, M. E., & AYERS, G. M. (1993). *Guideline for ToBI labelling*. First edition.

BECKMAN, M. E., & EDWARDS, J. (1990). Lengthening and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 152–178). Cambridge, U.K.: Cambridge University Press.

BECKMAN, M. E., & PIERREHUMBERT, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255–309.

BELL, A., GREGORY, M. L., JURAFSKY, D., GIRARD, C., BRENIER, J., & IKENO, A. (2002). Which predictability measures affect content word durations? In *PMLA ICSLP Workshop*.

BOOIJ, G. (1995). *The phonology of Dutch*. Clarendon Press.

CAMPBELL, W. N. (1992). *Multilevel timing in speech*. Ph.D. thesis, Sussex University.

CAMPBELL, W. N., & ISARD, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics*, **19**, 37–47.

CLARKSON, P., & ROSENFELD, R. (1997). Statistical language modeling using the CMUCambridge toolkit. In *Proceedings of Eurospeech 97*, 2707–2710.

CRUTTENDEN, A. (1986). *Intonation*. Cambridge, U.K.: Cambridge University Press.

CUTLER, A., & CARTER, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133–142.

FIDELHOLTZ, J. (1975). Word frequency and vowel reduction in English. In *CLS 75*, 200–213.

FOUGERON, C., & KEATING, P. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, **101**, 6, 3728–3740.

FOWLER, C. A., & HOUSUM, J. (1987). Talkers' signaling of "New" and "Old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, **26**, 489–504.

FOWLER, C. A., LEVY, E. T., & BROWN, J. M. (1997). Reduction of spoken words in certain discourse contexts. *Journal of Memory and Language*, **37**, 24–40.

FRY, D. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, **27**, 765–768.

GOLDINGER, S., & W., SUMMERS (1989). Lexical neighborhoods in speech production: A first report. *Research on Speech Perception Progress Report*. Psychology Department, Speech Research Laboratory, Indiana University, **15**, 331–342.

HUNNICUT, S. (1985). Intelligibility versus redundancy — conditions of dependency. *Language and Speech*, **28**, 45–56.

JURAFSKY, D., BELL, A., GREGORY, M., & RAYMOND, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam, The Netherlands: John Benjamins.

LADD, D. R. (1996). *Intonational phonology*. Cambridge, U.K.: Cambridge University Press.

LADD, D., & CAMPBELL, N. (1991). Theories of prosodic structure: Evidence from syllable duration. In *Proceedings of the XIIth International Conference on Phonetic Sciences*, **2**, 290–293.

LADEFOGED, P. (1982). *AUC course in phonetics*. New York: Harcourt, Brace, Jovanovich, 2nd edition.

LIEBERMAN, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, **6**, 172–187.

LINDBLOM, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory — Quarterly Progress and Status Report*, Research Laboratory of Electronics, MIT, Cambridge, **4**, 1–29.

MAYO, C., AYLETT, M., & LADD, D. (1997). Prosodic transcription of Glasgow English: An evaluation study of GlaToBI. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Proceedings of the ESCA Intonation Workshop* (pp. 231–234). ESCA.

NETER, J., WASSERMAN, W., & KUTNER, M. H. (1990). *Applied linear statistical models: Regression, analysis of variance and experimental design* (3rd ed.). Boston: Irwin.

PAN, S., & HIRSCHBERG, J. (2000). Modeling local context for pitch accent prediction. *Proceedings of the ACL 2000*, ACL, Hong Kong, 233–240.

PIERCE, J. R. (1961). *Symbols, signals and noise: The nature and process of communication*. New York: Harper.

PISONI, D., NUSBAUM, H., LUCE, P., & SLOWIACZEK, L. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, **4**, 75–95.

SAMUAL, A. G., & TROICKI, M. (1998). Articulation control is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language*, **39**, 175–194.

SANTEN, J. van (1998). Chapter 5: Timing. In R. Sproat (Ed.), *Multilingual text-to-speech synthesis* (pp. 115–139). Boston: Kluwer Academic Publishers.

SELKIRK, E. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic prosody II*. Trondheim: TAPIR.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423; 623–656.

SHATTUCK-HUFNAGEL, S., & TURK, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, **25**, 193–247.

SLUIJTER, A. (1995). *Phonetic correlates of stress and accent*. The Hague: Holland Academic Graphics.

SOTILLO, C. F. (1997). *Phonological reduction and intelligibility in task-oriented dialog*. Ph.D. thesis, University of Edinburgh.

TURK., A., & WHITE, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, **27**, 171–206.

TURK., A. E., & SAWUSCH, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics*, **25**, 25–41.

WIGHTMAN, C., SHATTUCK-HUFNAGEL, S., OSTENDORF, M., & PRICE, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, **91**, 1707–1717.

WRIGHT, R. (1997). Lexical competition and reduction in speech: A preliminary report. *Progress Report*, Indiana University, **21**, 471–485.