



HAL
open science

The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English)

Daniel J. Hirst, Caroline Bouzon

► **To cite this version:**

Daniel J. Hirst, Caroline Bouzon. The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English). Eurospeech (ISCA), Sep 2005, Lisbon, Portugal. pp.29-32. hal-00241549

HAL Id: hal-00241549

<https://hal.science/hal-00241549>

Submitted on 6 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English).

Daniel Hirst & Caroline Bouzon.

CNRS, UMR 6057 Laboratoire Parole et Langage,
Université de Provence, Aix en Provence, France
daniel.hirst@lpl.univ-aix.fr, bouzon@lpl.univ-aix.fr

Abstract

Research into the effect of stress and boundaries on segmental duration in speech has, for obvious reasons, most often been applied to carefully constructed sentences pronounced in laboratory conditions. The availability of a large labelled database of British English (Aix-Marsec) provides an opportunity to test different hypotheses concerning the factors influencing segmental duration from a corpus of authentic speech (defined as speech produced with the intent of communicating its meaning to the listener). In particular, in this paper, we look at the effect of stress and boundaries on prosodic structure in British English. Recent work has suggested that while word boundaries seem definitely to have a significant effect of the duration of segments, once the number of segments in the narrow rhythm unit is known, there is no orthogonal effect of word stress. In this study we look in particular at effects of word and intonation unit boundaries and at their possible interaction with stress and find that while intonation unit boundaries definitely affect segmental duration, no similar effect could be shown for word boundaries.

1. Introduction and background

It is generally accepted that, next to phonemic identity, prosodic structure is one of the major characteristics influencing segmental duration, where prosodic structure is the organisation of phonemes into such higher-level units as syllables, stress feet and intonation units. As Klatt [14] noted, nearly twenty years ago, however:

"One of the unsolved problems in the development of rule systems for speech timing is the size of the unit (segment, onset/rhyme, syllable, word) best employed to capture various timing phenomena." (p. 760)

There has been considerable speculation by phonologists as to the nature of prosodic representations, conceived of as an autonomous hierarchical structure [17],[9]. There has, however, been comparatively little experimental evidence to justify these theoretical constructs, so that their relationship to phonetic and acoustic data remains somewhat obscure and very much a subject for empirical investigation.

Among the units most frequently cited in phonetic studies as playing a role in the determination of segmental duration are

- the **syllable** [8] - with generally a binary distinction between stressed and unstressed syllables
- the **stress-foot** [1] - consisting of a stressed syllable and any following unstressed syllables up until (but not including) the next stressed syllable or until the boundary of an intonation unit
- the **word** [18]

- the **intonation unit** - (also known as phonological phrase, tone unit etc...) delimited by a full intonation boundary

To these, our own recent research [6], [7], presented in more detail below, has suggested that, following Wiktor Jassem's pioneering work [13], we should add:

- the **narrow rhythm unit** - consisting of a stressed syllable and any following unstressed syllables up until the end of the word. (cf also the "within word foot" [20])
- the **anacrusis** - consisting of any unstressed syllables not included in a narrow rhythm unit

Note that once the boundaries of the word and of the stressed syllable are given, the limits of the foot, narrow rhythm unit and anacrusis are automatically defined.

These potential phonological units are illustrated in Figure 1:

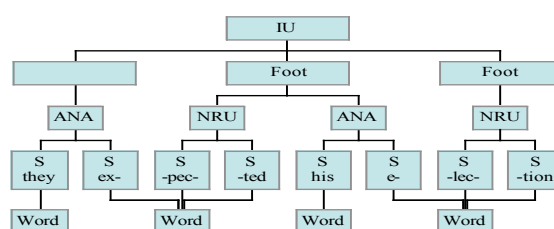


Figure 1. The Intonation Unit (IU) "They expected his election" parsed into syllables (S), narrow rhythm units (NRU), anacrusis (ANA), stress feet and words.

The status of the initial anacrusis is represented here as an empty node at the level of the foot, reflecting the fact that this has been analysed in different ways by different authors. For some, the initial anacrusis is assumed to be an immediate constituent of the superordinate intonation unit, for others it constitutes a defective foot, which Halliday [11] described as having a "silent ictus".

2. An analysis of a corpus of authentic speech

The majority of studies of segmental duration are based on corpora of carefully constructed sentences pronounced in laboratory conditions. For a recent study using this type of material, including an extensive bibliography see [21]. The motivation behind this restriction has been to systematically manipulate the prosodic structure of utterances in order to test different hypotheses. In recent work [6], [7] we have started analysing a large database of spoken English, the Aix-Marsec database, described in the following section, on the assumption that with a large enough database the different prosodic effects we wish to test should be sufficiently represented without needing to construct artificial utterances.

2.1. The Aix-Marsec database

The Aix-Marsec database, contains about five and a half hours of speech representing a number of different speaking styles, all of which can be characterised as what we have termed 'authentic speech', that is speech produced with the intent of communicating its meaning to listeners, unlike laboratory speech where the communicative function of the speech act is generally lacking.

The database derives originally from the SEC (Spoken English Corpus), a collection of BBC recordings from the 1980s, grouping eleven different radio speech styles ranging from news and interviews to poetry reading. The data consists of recordings of British English from 53 different speakers (17 male and 36 female) and includes about 55000 orthographically transcribed words as well as a manually transcribed prosodic annotation (G. Knowles and B. Williams) using a series of fourteen tonetic stress marks [15][16].

The SEC was subsequently adapted to facilitate computer use and was renamed the MARSEC (MACHINE READABLE Spoken English Corpus); changes consisted in manually aligning the word and (minor-major) intonation unit boundaries with the sound and changing some of the tonetic stress marks (TSM) into ASCII symbols in order to have a computer compatible annotation system [19].

Our contribution to the database [2] was to provide broad phonetic transcription for the corpus, which was then automatically aligned with the speech signal [3]. In addition, we added labels for prosodic categories, including syllable constituents (onset, nucleus and coda), syllables, anacruses and narrow rhythm units, stress feet and intonation units. All the labels were converted to the Praat TextGrid format [4][5]. The complete database is freely available for research (contact the first author) with the provision that users agree that any enhancements they make to the database shall also be made freely available.

2.2. Preliminary results obtained from the corpus

The first major finding of our analyses of the corpus was that while both word boundaries and stress seem to play a distinctive role in British English speech, no similar role could be shown for the syllable.

The effects of prosodic structure were demonstrated by looking at the linear correlation between the number of phonemes in a given unit and the duration of the corresponding phones. Strong negative correlations were found between the duration of segments and the number of phonemes in the stress-foot, in the narrow rhythm unit and in the word but not in the syllable or the anacrusis.

This confirms the findings of [12] who analysed the effect of prosodic structure on segmental duration in a passage of read speech and found a certain degree of compression on the level of the stress foot and the narrow rhythm unit but not on the level of the syllable, and concluded:

"there is no evidence in our data of any syllable timed feature of spoken British English". (p5)

The negative correlations were taken as evidence that the units in question play a role in planning the pronunciation of the utterance. A tendency to shorten segments on the level of suprasegmental units can be interpreted as a tendency to make these units more nearly equal in duration than they would be if there were no compression at all.

A second finding was that the degree of compression at the level of the narrow rhythm unit was greater than that at the level of the foot or of the word. This is understandable, since words can always be parsed into an anacrusis and/or a narrow rhythm unit and feet can always be parsed into a narrow rhythm unit followed by an optional anacrusis, as can be seen from the illustration in figure 1. It follows logically, then, that if there is a significant effect within the narrow rhythm unit but not in the anacrusis, the effect found at the level of the word as well as that at the level of the foot can be interpreted as the indirect effect of that of the narrow rhythm unit, diluted by the presence of any anacrusis.

A third major finding of these studies was that once the number of phonemes in the narrow rhythm unit was determined, no further orthogonal effect of stress was found. Phones in stressed syllables, in other words, are not intrinsically more lengthened (in our corpus) than phones in unstressed syllables in the same narrow rhythm unit, once the specific contribution of the identity of the phoneme is neutralised by a z-transformation of the raw segmental duration:

$$z_{i/p} = \frac{d_{i/p} - \mu_p}{\sigma_p} \quad (1)$$

where $z_{i/p}$ is the z-score of a given instance i of a phoneme p , $d_{i/p}$ is the corresponding raw duration, μ_p and σ_p are the mean and standard deviation for the duration of that phoneme.

In fact, if we plot the z-score of the duration of phones as a function of the number of phonemes in the narrow rhythm unit for stressed and unstressed phonemes, we see in Figure 2 (reproduced from [7]) that stressed phones are more lengthened than unstressed phones only when they occur in very short narrow rhythm units, containing no more than three phonemes. For longer units, the stressed phones are, in fact, most of the time actually relatively less lengthened than the unstressed phones.

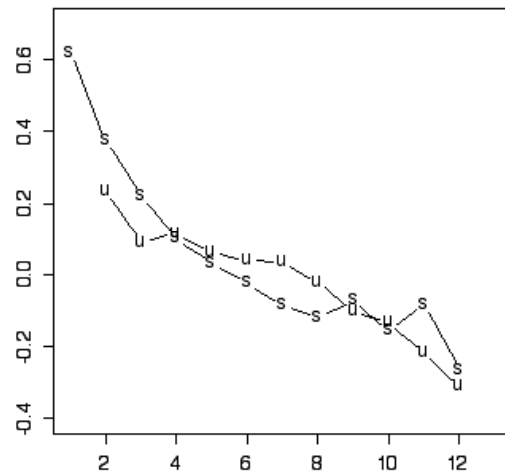


Figure 2. Duration of segments in z-score (y-axis) as a function of the number of phonemes in the narrow rhythm unit (x-axis) for stressed (s) and unstressed (u) phones.

3. New results from the Aix-Marsec database

A plausible explanation for this non-linear effect for stressed and unstressed phone duration is that these results are in fact conflating the effect of stress with the effect of final lengthening affecting the last few phonemes of an intonation unit. When the units are very short and occur at the end of an intonation unit, the stressed phones themselves undergo this lengthening effect; when the units are longer, only unstressed phones are affected.

Indeed, if we plot the relation between the distance of a phoneme from the end of the intonation unit against the corresponding z-transformed duration we can see clearly that the effect is very strong for the last two or three phonemes of the intonation unit but that it quickly tails off and then after about 30 phonemes becomes erratic as the distance from the end of the unit gets greater and the corresponding number of data points gets smaller.

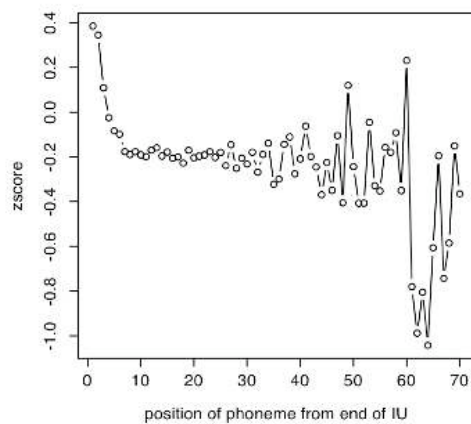


Figure 3. Z-score of duration of segments as a function of their position (in number of phonemes) as measured from the end of the intonation unit.

When the last two phonemes of the intonation unit are excluded from the analysis, an analysis of co-variance on the z-score of segmental duration as a factor of stress and of the number of phonemes in the narrow rhythm unit shows a very strong effect of the number of phonemes ($F(1,140503) = 529.703$; $p < 2e-16$) but a barely significant one for the factor stress ($F(1,14053) = 5.6124$; $p = 0.018$).

Although the interaction between stress and number of phonemes is highly significant ($F(1,14053) = 76.7915$; $p < 2e-16$), plotting the number of phonemes against z-score for stressed and unstressed phonemes, does not show any readily interpretable tendency for this interaction.

Some authors (eg [10]) have suggested that since phonemes may behave differently from others under the effect of stress - in particular, vowels might be more specifically affected by the presence of stress than consonants. In order to test this possibility, we carried out an analysis of co-variance using the z-score of vowel duration as dependent variable rather than the z-score of all phonemes. As in the previous analysis, the last two phonemes of each intonation unit were excluded from the analysis.

Both factors stress and number of phonemes in the narrow rhythm unit are highly significant ($F(1, 43080) = 1746.97$ and

1344.48, respectively $p < 2.2e-16$), as is the interaction between the two factors ($F(1, 43080) = 323.09$).

Plotting the number of phonemes against z-score for stressed and unstressed phonemes (Figure 4) seems, however, rather to show that if anything, unstressed vowels are more lengthened than stressed vowels within narrow rhythm units of the same size.

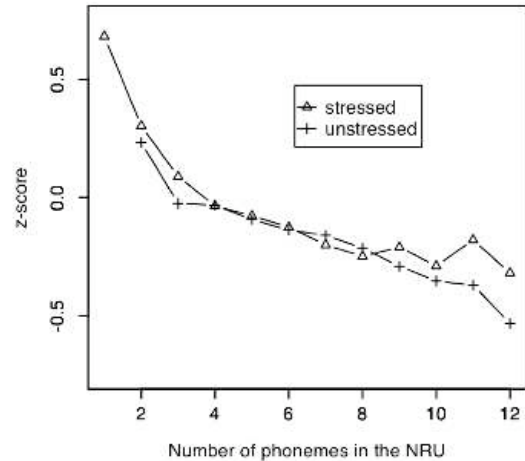


Figure 4. Z-score of segmental duration as a factor of number of phonemes in the narrow rhythm unit (NRU) for stressed and unstressed phones excluding the last two phonemes of each intonation unit.

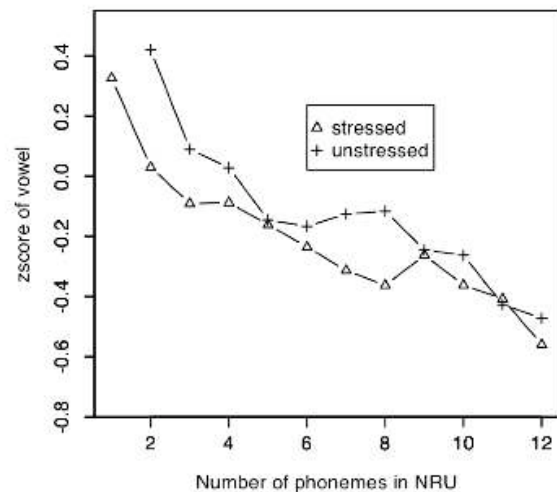


Figure 4. Z-score of duration of vowels as a function of the number of phonemes in the narrow rhythm unit (NRU) for stressed and unstressed phones excluding the last two phonemes of each intonation unit.

One other possibility, which our work had not so far taken into account, is that just as we showed that our results conflated final lengthening in the intonation unit and stress, so

perhaps there is a further effect of final lengthening in the word. Since the stressed syllable of a polysyllabic word is most frequently on the first syllable, any word final lengthening would most often be affecting final unstressed syllables and might be compensating for a specific lengthening effect of stress. To test this, we coded all phonemes for their position with values: *non-final*, *word-final* (last or penultimate phoneme of word) or *iu-final* (last or penultimate phoneme of intonation unit). An analysis of covariance with these factors showed both position and number of phonemes in the NRU to be highly significant ($F(2,140502) = 2270.4742$; $p < 2e-16$ and $F(1,140502) = 2057.8252$; $p < 2e-16$, respectively) while once again the factor stress itself was barely significant $F(1,140502) = 5.7906$; ($p = 0.016$). As can be seen from Figure 5, however, this effect of position appears to be exclusively one of intonation unit final lengthening - word final and non-final phonemes do not show any readily interpretable pattern of difference.

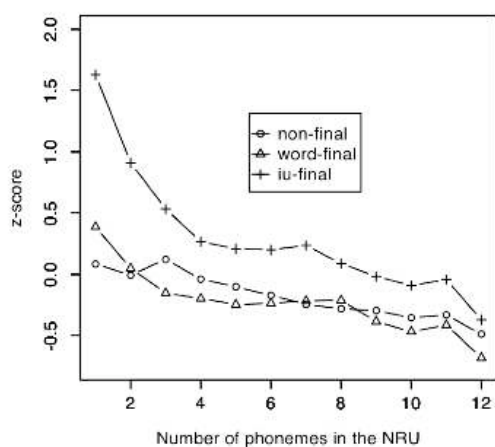


Figure 5. Z-score of the duration of of phones as a function of the number of phonemes in the narrow rhythm unit (NRU) for *iu-final* (last two phonemes of each intonation unit), *word-final* (last two phonemes of each word) and *non-final* phonemes.

4. Conclusions

The analyses reported here provide added support for our earlier conclusion that (as suggested by Wiktor Jassem in 1952 [13]) the stressed syllable has no special rhythmic status in British English other than that of occurring at the onset of a narrow rhythm unit. We also tested the possibility that it is the stressed vowel and not the consonants which carries the main effect of lengthening under stress, but the results were not at all compatible with this interpretation. Furthermore, while we found considerable evidence of lengthening of the final phonemes of intonation units, there was no evidence at all in our data that there is any similar systematic lengthening at the level of the word.

References

[1] Abercrombie, D. 1964. Syllable quantity and enclitics in English. In Abercrombie, D., Fry, P., MacCarthy, N. &

Trim, J. (eds): *In Honour of Daniel Jones*. London: Longman, 216-222.

[2] Auran, C., Bouzon, C. & Hirst, D.J. 2004. The Aix-MARSEC project: an evolutive database of spoken English. In Bel, B. & Marlien, I. (eds) *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan, 561-564.

[3] Auran, C., Bouzon, C., Hirst, D.J., Lévy, C., Nocéra, P. 2004. Algorithme de prédiction d'émissions de phonèmes et influence sur l'alignement automatique dans le cadre du projet Aix-MARSEC. in *Actes des XXVèmes Journées d'Etudes sur la Parole*, Fès, Maroc. 57-60.

[4] Boersma, P. 2001. Praat. A system for doing phonetics by computer. *Glott International*.5 (9/10), 341-345.

[5] Boersma, P. & Weenink, D. 2005. Praat. Doing phonetics by computer. [computer program]. Retrieved March 31, 2005 from <http://www.praat.org/>

[6] Bouzon, C. 2004. *Rythme et structuration prosodique en anglais britannique contemporain*. Doctoral dissertation, Université de Provence.

[7] Bouzon, C. & Hirst, D.J. 2004. Isochrony and prosodic structure in British English. in Bel, B. & Marlien, I. (eds): *Proceedings of the Second International Conference on Speech Prosody*, Nara, Japan, 223-226.

[8] Campbell, N. 1992. *Multi-level Timing in Speech*. Ph.D. thesis, University of Sussex.

[9] Goldsmith, J.A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell Ltd.

[10] Grabe, E. & Low, E.L. 2002. Durational variability in speech and the rhythm class hypothesis. In Gussenhoven, C. & Warner, N. (eds.): *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 515-546.

[11] Halliday, M.A.K. 1970. *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.

[12] Hill, D.R., Jassem, W. & Witten, I.H. 1978. A statistical approach to the problem of isochrony in spoken British English. *Research Report 78/27/6*, University of Calgary, Department of Computer Science.

[13] Jassem, W. 1952. *Intonation in Conversational English*. Warsaw, Polish Academy of Science.

[14] Klatt, D.H. 1987. Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82, 737-793.

[15] Knowles, G., Wichmann, A. & Alderson, P. 1996. *Working with Speech: perspectives on research into the Lancaster/IBM Spoken English Corpus*. London: Longman.

[16] Knowles, G., Williams, B. & Taylor, L. 1996. *A Corpus of Formal British English Speech*. London: Longman.

[17] Nespor, M. & Vogel, I. 1986. *Prosodic Phonology*. Dordrecht. Foris.

[18] Nootboom, S.G. 1991. Some observations on the temporal organisation and rhythm of speech. In *Proceedings ICPHS XII*, Aix-en-Provence, 228-237.

[19] Roach, P., Knowles, G., Varadi, T. & Amfield, S. 1993. MARSEC: A machine readable Spoken English corpus. *Journal of the International Phonetic Association* 23(2), 47-53.

[20] Turk, A. & White, L. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27, 171-206.

[21] White, L. 2002. *English speech timing. A domain and locus approach*. PhD thesis, University of Edinburgh.