Rhythmic grouping in word lists: competing roles of syllables, words and stress feet

Fred Cummins

Department of Computer Science
University College Dublin
fred.cummins@ucd.ie

ABSTRACT

The role of suprasyllabic constituents in determining speech timing remains hotly debated. We investigate the temporal organization of lists of words designed to reveal any competing influences on timing from syllables, lexical words and (Abercrombian) stress feet. Intervals between word onsets and stress foot onsets are observed in a baseline condition where stressed and unstressed syllables alternate, and in two conditions in which stresses are less regular. Evidence of competing influences on interval duration are found, suggesting that observed intervals are influenced both by syllabic or subsyllabic components, and also by suprasyllabic constituents.

1 INTRODUCTION

There has long been debate about the role of suprasyllabic prosodic constituents in determining the temporal organization of speech. In extreme form, the (thoroughly discredited) isochrony hypothesis claims that suprasyllabic units, such as the intervals bounded by successive stressed syllable onsets (the Abercrombian stress foot), are critical determinants of measured intervals [1]. The rejection of this simplistic and manifestly false hypothesis has led many to question the relevance of suprasyllabic units as determinants of speech timing [2]. A recent and comprehensive study which identified both the domain and the locus of various timing effects above the level of the syllable found evidence for lengthening at the beginning of the word, and also compression at the rightmost rhyme within a word, but no evidence for polysyllabic shortening [3]. Indeed, recent experimental studies have generally supported the role of lengthening and shortening effects at the edges of constituents, but have not demonstrated timing effects which operate throughout a suprasyllabic unit like the stress foot [4, 5, 6].

In many of the above studies, some effort is made to avoid "unnatural" timing patterns, as might be induced, e.g. by excessive repetition. In the present study, we look specifically at temporal structure where there is a repetition of stress patterns. We thus deliberately court the induction of what others have called "unnatural" patterns, and we suggest that temporal structure which arises in a repetitive context may provide important clues to the way in which the temporal organization of speech is dependent upon context. Repetition of stress patterns and foot structure may occur serendipitously in spontaneous speech, or may be invoked deliberately for rhetorical force. Regular stress patterns are, of course, exploited to the fullest in verse, but may be a characteristic of felicitous natural speech as well, whether by design or chance.

We examine the temporal structure of sequences of words, where subjects are encouraged to produce the words in "regular" sequence. In a **baseline** condition, the words are all trochees. In a second **iambic** condition, one word has the reverse, iambic, stress pattern, and in a third condition (**dactyl+iamb**), the word before the iamb is a dactyl. Subjects produce what they judge to be an isochronous reading, and by measuring the intervals between word onsets, and stressed syllable onsets, we look for systematic deviations from isochrony which may be related to suprasyllabic structures.

Figure 1 shows measurements made in each condition. We operationalize the onset time of a syllable as its estimated P-center [7] using the *bex* algorithm introduced in [8]. In doing so, we accept that our onset measurements will be somewhat noisy, and compensate as far as possible by employing only simple syllable structures, and a variety of segments. The P-center estimate is usually close to the vowel onset.

In the baseline condition, word onsets and stressed-syllable onsets coincide. We include this condition as, even in this case, subjects may deviate systematically from isochrony, e.g. by decelerating toward the list end [9], or by imposing a hierarchical patterning on the sequence of onsets [10]. We can then compare intervals of interest in the other two conditions with intervals from the baseline condition which are in the same serial position.

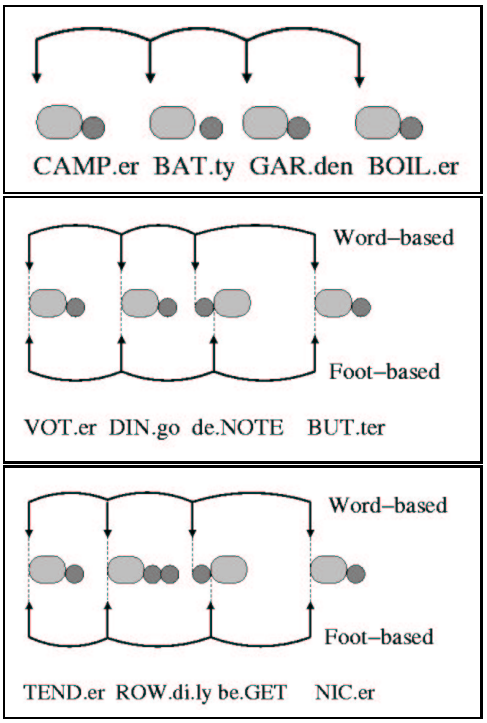

Figure 1: Top: onsets in the baseline condition. Middle and bottom: word- and stress-based onsets for the iambic (middle) and dactyl+iamb (bottom) conditions.

The comparisons which will be of most interest are those which test the possible role played by suprasyllabic units in determining interval timing. We distinguish three relevant hypotheses.

1. No Suprasyllabic Influence. This hypothesis, in the spirit of [11, 2], predicts that the duration of the syllables is primarily a function of the summed duration of their constituent segments, perhaps modified by the stress level of the syllable. By this hypothesis, all words in the baseline and iambic conditions should be of roughly equal duration (modulo segmental variation), while the dactyl of the third condition should be longer.

2. Lexical Word Influence. This hypothesis includes a role for the lexical word as a modulating influence. Intervals derived from the word-based measurements of the iambic and dactyl+iambic conditions should be more similar to baseline intervals than those based on the stress foot.

3. Stress Foot Influence. As before, but intervals based on foot-based measurements should more closely resemble those of the baseline.

2 METHODS

Fifty four Dublin students were recruited. Subjects presented in pairs and were recorded both reading lists alone, and together with a co-speaker. The lists were read as part of a larger data-gathering exercise in which speech produced alone was compared with speech produced when speaking with a co-speaker [12]. A randomized list of 18 word lists was prepared for each speaker pair (details below). After having each list read aloud by the experimenter, one subject then read each list in turn, with the instruction to read the list 'as regularly as possible'. The two subjects then read the list together, attempting to maintain synchrony with one another, and finally, the second subject read the list alone. For each subject, therefore, readings were obtained in a synchronous and in a solo condition. In the synchronous condition, subjects had had some practice at speaking synchronously, but had not read the word lists before. They were seated opposite one another and each could see the other.

Eighteen word lists were prepared. In the *baseline* condition, six lists each contain 8 trochees (example: "CAMPER BATTY GARDEN BOILER CARRY SHADOW TOASTER BULLY"). In the *iambic* condition, a stress reversal on one word led to successive feet having 3 and 1 syllables at a point which was either early or late within the sequence (Early example:"VOTER DINGO DENOTE BUTTER TINKER TELLING FARMER WEEPY"; Late: "VENDOR KICKER PITY TICKER KNOCKER BELOW BODY SUMMER"). In the *dactyl+iamb* condition, an additional unstressed syllable was added to one word to exaggerate the length difference between the long and the short stress feet. (Early: "TENDER ROWDILY BEGET NICER PACKER TALKER PASSING PEELING"; Late: "FINDER BEADY PETER CADDY LEGACY DEMEAN NUTTY CONGO"). Serial position of the perturbation (early/late) thus nests within list condition (baseline, iambic, dactyl+iamb).

For the baseline condition, data from all 54 speakers in both solo and synchronous conditions was analyzed. After examination of the variability in that data set, an initial analysis of 12 speakers (6 speaker pairs) from each of the other conditions was completed. Although results for these conditions are preliminary, the baseline results suggest that there is unlikely to be substantial change in the results as additional data is measured.

For each reading, P-centers were estimated for all syllables using the algorithm described in [8]. The intervals between the P-centers associated with the onsets of the vowels in the stressed syllables were measured. For lists in the *iambic* and *dactyl+iamb* conditions, interval measurements were done in two ways, using both word initial syllables, and stressed syllables as reference points.

3 RESULTS 1: BASELINE CONDITION

Figure 2 shows the duration of the first seven intervals for a representative word list (left) and for all the data combined (right). In the baseline condition, all words

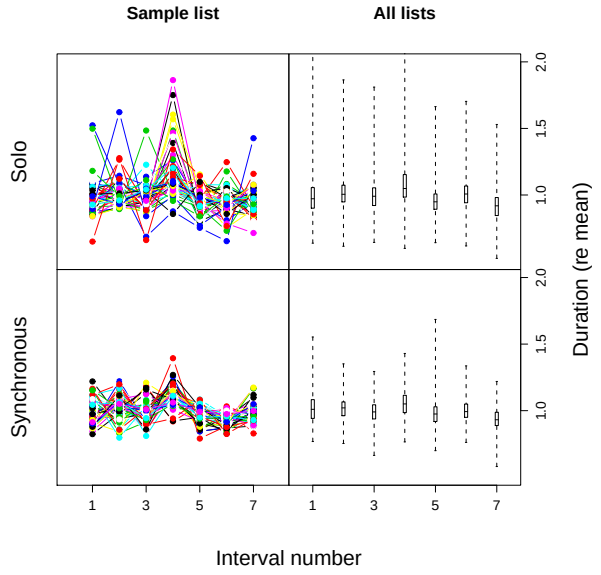

Figure 2: Reduction in inter-subject variability in the synchronous condition is evident both for a sample list (left) and for the summary data (right). Whiskers in the right panels span the entire data range ($n = 54$).

have the same trochaic stress pattern, and word onsets are thus identical to stressed syllable onsets. The eighth interval is not defined, as there is a break after each word list. Interval durations have been rate normalized by dividing each by the overall sequence duration, and then each interval has been divided by the mean of all points in its panel. The expedient of recording subjects in pairs, and asking them to read the lists in synchrony, resulted in a considerable reduction in inter-subject variability, as can be seen by comparing upper and lower panels. This is in line with other studies which have shown that synchronous speech is effective at reducing inter-subject variability in speech timing [13, 12, 14]. In what follows, we therefore report only the synchronous speech data. Analysis of the remaining data was carried out and the results were qualitatively similar, though noisier.

4 RESULTS 2: DACTYL+IAMB CONDITION

We compare intervals 2 and 3 (Early) and intervals 5 and 6 (Late) with the corresponding intervals from the baseline condition. Each of the three hypotheses above makes different predictions about the relative length of these intervals when we make measurements based on words or feet. In Table 1 we provide predictions and observed effects for Intervals 2 and 3, and 5 and 6, as measured in both ways. The observed effects are obtained by comparing median values, and corroborated by obtaining a non-parametric estimate of the difference in location parameters for the two distributions, using a Mann-Whitney test [15].

Intervals 2 and 3	Measurements taken			
	word-based		foot-based	
Hypothesis	int 2	int 3	int 2	int 3
No Suprasyll. Influence	+	0	+	–
Lexical Word Influence	0	0	+	–
Stress Foot Influence	–	+	0	0
Effect found:	0	+	+	0
Intervals 5 and 6	Measurements taken			
	word-based		foot-based	
Hypothesis	int 5	int 6	int 5	int 6
No Suprasyll. Influence	+	0	+	–
Lexical Word Influence	0	0	+	–
Stress Foot Influence	–	+	0	0
Effect found:	0	+	+	0

Table 1: Predicted lengthening (+) and shortening (–) of intervals 2 and 3, and 5 and 6 relative to baseline under each of 3 hypotheses and 2 types of measurement in the dactyl+iamb condition. The final row of each section lists the effects actually observed.

The observed effects are consistent across early and late positions, but do not accord squarely with any of the three hypotheses. The second and fifth intervals measured using word onsets are unchanged compared with the baseline data. This could be interpreted as supporting the Word Influence hypothesis, or could reflect a tension between the opposing predictions of the two other hypotheses, if we acknowledge that these hypotheses are not necessarily mutually exclusive, but could reflect independent contributions to the observed durations from sub- and supra-syllabic units. When foot onsets are used, the results are consistent with either the Word Influence or the No Suprasyllabic Influence hypotheses.

The third and sixth intervals, measured using word onsets, are long, relative to the baseline. This is predicted only by the Stress Foot Influence hypothesis. When foot onsets are used, no change is observed compared with baseline, which is again consistent only with the Stress Foot Influence hypotheses.

5 RESULTS 3: IAMBIC CONDITION

In the iambic condition, there are only two competing hypotheses, as the Word Influence and the No Syprasyllabic Influence hypotheses make the same predictions. For convenience, we will refer to them collectively as the No Stress Foot Influence Hypothesis. A table of predicted and observed effects is given in Table 2.

Intervals 2 and 3	Measurements taken word-based		foot-based	
Hypothesis	int 2	int 3	int 2	int 3
No Stress Foot Influence	0	0	+	–
Stress Foot Influence	–	+	0	0
Effect found:	–	+	+	–

Intervals 5 and 6	Measurements taken word-based		foot-based	
Hypothesis	int 5	int 6	int 5	int 6
No Stress Foot Influence	0	0	+	–
Stress Foot Influence	–	+	0	0
Effect found:	–	+	+	0

Table 2: Predicted lengthening (+) and shortening (–) of intervals 2 and 3, and 5 and 6 in the iambic condition under each of 2 hypotheses and 2 types of measurement. The final row of each section lists the effects actually observed.

With a single exception, these results appear, at first blush, to support both hypotheses! That is, when measurements are taken using word onsets, Intervals 2 and 5 are shorter than baseline, Intervals 3 and 6 are longer, as predicted by the Stress Foot Influence hypothesis. When measurements are taken using foot onsets, the reverse is found, as predicted by both the No Suprasyllabic Influence and Word Influence hypotheses. A sole exception is the sixth interval which is not lengthened relative to baseline when foot-based measures are employed.

6 Discussion

In both the iambic and the dactyl+iamb conditions, results obtained can be interpreted as arising from a combination of competing influences on interval durations. No one of the three hypotheses above can adequately account for the observed patterns. If the observed intervals result from a combination of influences, we could see the patterns observed here. This suggests that a gradient influence might be exerted by suprasyllabic constituents, which serves to quantitatively modulate durations.

The present work is motivated by the intuition that rhythmic context will affect the realization of suprasyllabic constituents in speech. The experiment was designed to exaggerate the role of rhythmic regularity, and to make manifest any role played by suprasyllabic constituents in determining speech timing. The evidence presented here supports an account in which sub- and suprasyllabic influences combine to determine observed intervals when the material being spoken is strongly conducive to rhythmic regularity.

ACKNOWLEDGMENTS

Keiichi Tajima (ATR) helped in designing the word lists. Rebecca Grimes (UCD) provided very helpful comments. Work funded by a Basic Research grant from Enterprise Ireland to the author.

REFERENCES

[1] David Abercrombie, *Elements of general phonetics*, Aldine Pub. Co., Chicago, IL, 1967.

[2] Jan P. H. van Santen and Chilin Shih, "Suprasegmental and segmental timing models in Mandarin Chinese and American English," *Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 1012–1026, 2000.

[3] Laurence White, *English Speech Timing: A Domain and Locus Approach*, Ph.D. thesis, The University of Edinburgh, 2002.

[4] Cécile Fougeron, "Articulatory strengthening at edges of prosodic domains," *Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3728–3740, 1997.

[5] Patricia A. Keating, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu, "Domain-initial articulatory strengthening in four languages," in *UCLA Working Papers in Lingusitics*, vol. 97, pp. 139–156. 1999, Also to appear in Broe, M. and Pierrehumbert, J. (Eds) Papers in Laboratory Phonology VI.

[6] Alice E. Turk, "Structural influences on boundary-related lengthening in English," in *Proceedings of ICPhS '99*. 1999.

[7] Sophie K. Scott, *P-centers in Speech: An Acoustic Analysis*, Ph.D. thesis, University College London, 1993.

[8] Fred Cummins and Robert F. Port, "Rhythmic constraints on stress timing in English," *Journal of Phonetics*, vol. 26, no. 2, pp. 145–171, 1998.

[9] André-Pierre Benguerel and Janet D'Arcy, "Time-warping and the perception of rhythm in speech," *Journal of Phonetics*, vol. 14, pp. 231–246, 1986.

[10] David A. Rosenbaum, Sandra B. Kenny, and Marcia A. Derr, "Hierarchical control of rapid movement sequences," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 9, no. 1, pp. 86–102, 1983.

[11] Jan P. H. van Santen, "Segmental duration and speech timing," in *Computing Prosody: Computational Models for Processing Spontaneous Speech*, Yoshinori Sagisaka, Nick Campbell, and Norio Higuchi, Eds., chapter 15. Springer Verlag, New York, 1996.

[12] Fred Cummins, "Practice and performance in speech produced synchronously," *Journal of Phonetics*, 2003, in press.

[13] Fred Cummins and Deb Roy, "Using synchronous speech to minimize variability in pause placement," in *Proceedings of the Institute of Acoustics*, Stratford-upon-Avon, 2001, vol. 23 (3), pp. 201–206.

[14] Elena Zvonik and Fred Cummins, "Pause duration and variability in read texts," in *Proc. ICSLP*, Denver, CO, 2002, pp. 1109–1112.

[15] Ross Ihaka and Robert Gentleman, "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 299–314, 1996.