Interval Timing in Spoken Lists of Words

Fred Cummins

Department of Computer Science

University College Dublin

fred.cummins@ucd.ie

August 26, 2004

Abstract

Isochronous interval production with discrete motor responses has been studied most intensively using well-practiced tapping in a synchronization/continuation paradigm. Somewhat fewer studies have examined the timing of shorter sequences of intervals which may be susceptible to metrical grouping. I here look at the attempted isochronous production of lists of eight trochees, and examine the resulting interval patterns both in terms of the Wing-Kristofferson model (Wing and Kristofferson, 1973) and Rosenbaum's hierarchical timing model (Rosenbaum et al., 1983). In Experiment 1, readings are self paced, and done under conditions of uncertainty. The Wing-Kristofferson model is not applicable to the data, which more closely resemble data exhibiting hierarchical control of intervals. In Experiment 2, readings are paced, well-practiced and done without uncertainty. Neither Wing and Kristofferson's, nor Rosenbaum's model adequately capture the serial dependencies observed, though under these conditions variability is greatly reduced. In both experiments, there are dependencies between non-adjacent intervals which neither model can yet account for.

1 Introduction

Serial order in behaviour has been the focus of intense study, at least since Karl Lashley's seminal paper of 1951 (Lashley, 1951). In that paper, Lashley identified the sequencing of action as a fundamental problem for which nervous systems have evolved solutions, and presented the understanding of temporally coordinated behaviour as one of the greatest challenges in understanding the neural control of action. Since then, a significant body of literature has arisen with a narrower focus on the temporal control (and perception) of isochronous sequences, most typically in tapping tasks (Chen et al., 2002; Ivry and Hazeltine, 1995; Wing et al., 1989). One guiding intuition has been that timing and coordination might reasonably be studied independently of the mode of behaviour. There is now ample experimental and neurological evidence that at least some aspects of timing and coordination are regulated by dedicated mechanisms that are not bound to specific action systems such as the hand/arm linkage used in tapping, or the active articulators of the vocal tract, as used in speaking. Indeed, a wealth of studies suggests that central timing mechanisms may be common to both production and perception (Meegan et al., 2000; Ivry and Hazeltine, 1995; Ivry, 1996; Rosenbaum, 2001; Hazeltine et al., 1997).

The Wing and Kristofferson model (1973) has frequently been applied to tapping data obtained under well-practiced, paced conditions. This model allows the interval variability to be decomposed into two parts: one related to the variability of a central clock, the other associated with motor or peripheral variability. This decomposition rests on several important independence assumptions. First, the clock and motor variabilities are assumed to be independent. Secondly, variabilities among successive intervals are assumed to be independent. This latter requirement is something of a leap of faith, and many studies have adopted the questionable expedient of excluding a subset of data from analysis if it is found to violate these assumptions (Max and Yudman, 2003; Ivry and

Keele, 1987).

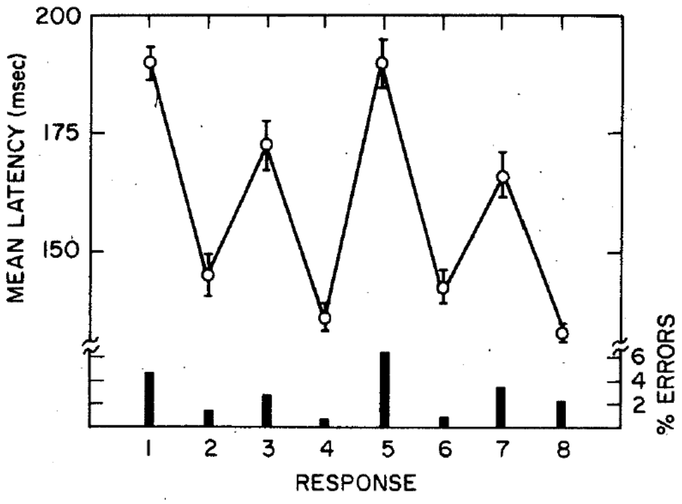

Figure 1: Mean latencies for responses in errorless trials (points) and percentage of trials with at least one error (bars). Reproduced (permission requested from APA) from Rosenbaum et al., (1983). The first interval plotted is that between the end of one sequence and the onset of the next, and does not have a counterpart in the data to be presented herein.

A rather different view of tapping data was presented in Rosenbaum et al., (1983), in which subjects were instructed to produce rapid continuous streams of sequences of 8 isochronous taps, using alternating fingers on both hands. The task was well practiced, and self-paced, with the instruction to complete the sequences "as quickly as possible". The resulting interval sequence, shown in Figure 1 was not isochronous; rather, interval duration was systematically related to serial position. Rosenbaum et al. interpreted this as strong evidence for hierarchical execution of the movement pattern. It is certainly consistent with a metrical-like organisation in which production units group recursively in pairs, with the second element of a pair thereby acquiring some final lengthening (final lengthening is a ubiquitous and familiar phenomenon in serial production of action, including speech). Some of the principal elements which distinguish this study from other isochrony studies are the use of alternating effectors for tapping, and the elicitation of short sequences of eight responses. Either of these factors may contribute to the global, metrical like organization observed.

Although the tapping literature is rich and varied, the investigation of deliberately produced isochrony in speech has attracted much less attention. Intuitions that isochrony might be a regularly occurring feature of naturally produced speech have failed to produce supporting evidence (Dauer, 1983; Lehiste, 1977). The speech cycling task, introduced in Cummins and Port (1998), required subjects to repeat short phrases in time with a series of tones that specified the relative timing of stressed syllable onsets. Strong constraints on inter-stress intervals were observed, such that an integral number of stress feet[1] of equal duration were nested within an overall phrase repetition cycle. Under these constrained experimental conditions, at least, isochrony of stress onsets can be produced by English speaking subjects.

Isochronous speech production has been investigated within the framework of the Wing and Kristofferson model. Max and Yudman (2003) had stutterers and non-stutterers complete speech, nonspeech orofacial, and tapping tasks, and analyzed all the data using the Wing and Kristofferson model. They report that lag 1 autocorrelations, averaged by participant and condition, were generally in the predicted range of 0 to -0.5, but nonetheless, some 13 of 60 cases (20 participants, 3 conditions) in the speech task, 12 of 60 cases in the orofacial nonspeech task and 10 of 54 in the tapping task violated this prediction (they do not report on the possible significance of autocorrelations at lags greater than 1). Their analysis continued by applying a "correction" to individual trials that violated the model's assumptions, as had been done before, e.g. by Ivry and Keele (1987), who simply recorded a motor delay estimate of zero when the lag 1 autocorrelation

[1] The Abercrombian notion of stress foot, defined as the interval between stressed syllable onsets, irrespective of word boundaries, was used (Abercrombie, 1967).

was positive.

The repercussions of this form of data manipulation are not entirely clear, and a typical claim is that the manipulation does not influence the “tenor of the conclusions” (Ivry and Keele, 1987). Caution seems advisable in applying a model, where a substantial proportion of the data clearly violate assumtions of the model. A principal motivation for the present study was to see what serial dependencies might exist among intervals beyond those predicted by the Wing-Kristofferson model.

The present study examines a specific kind of isochronous speech production: the production of lists of eight trochees. The intervals between successive stressed syllable provide the primary data. The task of repeating eight words in a regular fashion brings with it the possibility of hierarchical organization, as evidenced in the data of Rosenbaum et al., (1983). There is, however, no equivalent to the alternating effectors employed therein. It is also a serial interval production task, and so the potential suitability of the Wing-Kristofferson model will also be investigated.

The data collected in this study were part of a larger data gathering exercise which studied the effects on speech timing of constraining speakers to read prepared texts in synchrony (Cummins, 2003a). In previous experiments, we had found this constraint to be effective at reducing inter-speaker variability for such factors as speaking rate, pause duration and accent placement (Cummins, 2004). To date, it has not been observed that synchronous speech introduces any artifacts other than a reduction in temporal variability. Synchronous speech was used, as well as conventional ‘solo’ speech in the present study, partly so that any artifacts introduced by speaking in synchrony might be identified. As will become apparent, there were few observed differences between synchronous and conventional speech.

2 Methods

Fifty four Dublin students were recruited for payment. Subjects presented in pairs and were recorded reading a series of texts, both alone and in synchrony wth one another. The word lists which are discussed here were elicited midway through a recording which lasted approximately 45 minutes per pair. For each pair of speakers, a randomized sequence of 18 word lists was prepared. Of these, 6 were all trochaic (discussed here) while the remaining 12 contained some words with different stress patterns (iambs and dactyls. For details see Cummins, 2003b.). The entire set of 18 lists was read aloud by the experimenter once, to ensure that any uncertainties about pronounciation and stress patterns were addressed. Then one of the two subjects read each list, with the instruction that the list be read '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 reading in synchrony with their co-speaker, but had not read the word lists before. Subjects were seated opposite one another and could see each other (Cummins, 2003a). Recordings were done using head-mounted near-field dynamic microphones (Shure SM10A), and CSL recording software.

2.1 Measurements

For each series of eight words, the onsets of the stressed syllables were identified using the Bex algorithm first introduced in Cummins and Port (1998). This algorithm provides a rough estimate of the P-centre of a syllable, which is located near to the vowel onset, but varies systematically as a function primarily of the consonants in the syllable onset (Morton et al., 1976; Scott, 1993). Although this can be no more than a very approximate estimate of a P-centre, the use of an automatic procedure ensured maximal consistency in onset identification, and allowed a correction algorithm to be employed as described below. Each series of eight words thus provided seven complete interval measurements.

In looking at interval durations, the primary interest here is concerned with patterns of relative duration within a series. Each interval between two stressed syllable onsets was therefore first expressed as a proportion of the interval from the first to the last stressed syllable onset.

Two kinds of systematic influences on interval duration were then considered. Firstly, it may be found that intervals vary systematically as a function of their position within the sequence. This is precisely the variability of interest here, so that any possible serial dependencies in the intervals produced may be indentified. However there is a second possible source of systematic variation which needs to be corrected for, as it is not of interest to the present study. This is variation caused by consistent misestimation of the actual P-centre of a syllable. In ignorance of the actual P-centres, the word-specific effects of misestimation can be at least evened out across lists by subtracting an estimate of the word-specific effect on interval duration from each interval. Let $\overline{I}_p$ denote the mean interval in position p, and $\overline{I}_{wp}$ be the mean interval for the subset of those intervals containing word w, then for each such interval I_{iwp}, a correction factor is applied:

$$I_{iwp} = I_{iwp} - (\overline{I}_{wp} - \overline{I}_p) \tag{1}$$

3 Results

A repeated measures ANOVA was conducted with serial position and speaking condition as within-subject factors. There was a significant main effect of serial position [F(6,4420)=99, p<0.001] and the interaction of serial position and speaking condition was also significant [F(6,4420)=7.2,

$p<0.001$]. There was no main effect of speaking condition. Examination of linear trends in each trial revealed that synchronous trials showed a tendency for intervals to get shorter as the trial progressed, while rate increases and decreases were evenly divided in the solo condition. Both data sets were therefore detrended by removing the linear component from each trial. A repeated measures ANOVA on the resulting matrices showed a continued main effect of position [F(6,4420)=88, $p<0.001$] and no other effects or interaction. F-tests comparing variability in each serial position across speaking conditions revealed no significant differences except in position 3, for which the variability in the synchronous condition was less. Because of the difference in linear trends in the original data, solo and synchronous data will be looked at separately.

Figure 2 (top panels) shows relative interval durations as a function of serial position in list for both Solo and Synchronous conditions. Median interval values vary systematically as a function of position in list. Median interval durations alternate as short-long-short-longest-short-long-short, similar to the tapping data from Rosenbaum et al., (or, rather, intervals 2–8 therein). Variation within a given interval is, however, much greater than the variation observed as a function of position in list.

The lower panels in the same figure show the standard deviation of interval durations. There is considerably higher variability in the fourth interval in each condition. This is the interval which separates the first half of the list from the second.

An analysis of interval data in terms of the Wing and Kristofferson model requires computation of the lag one autocorrelation. The assumptions of the model predict a negative lag one autocorrelation, bounded between zero and -0.5 (see Wing and Kristofferson, 1973 for details). The assumptions of the model are deemed to be violated in the case of positive autocorrelations, negative lag one correlations below -0.5, or significant autocorrelation at lags other than one. In many

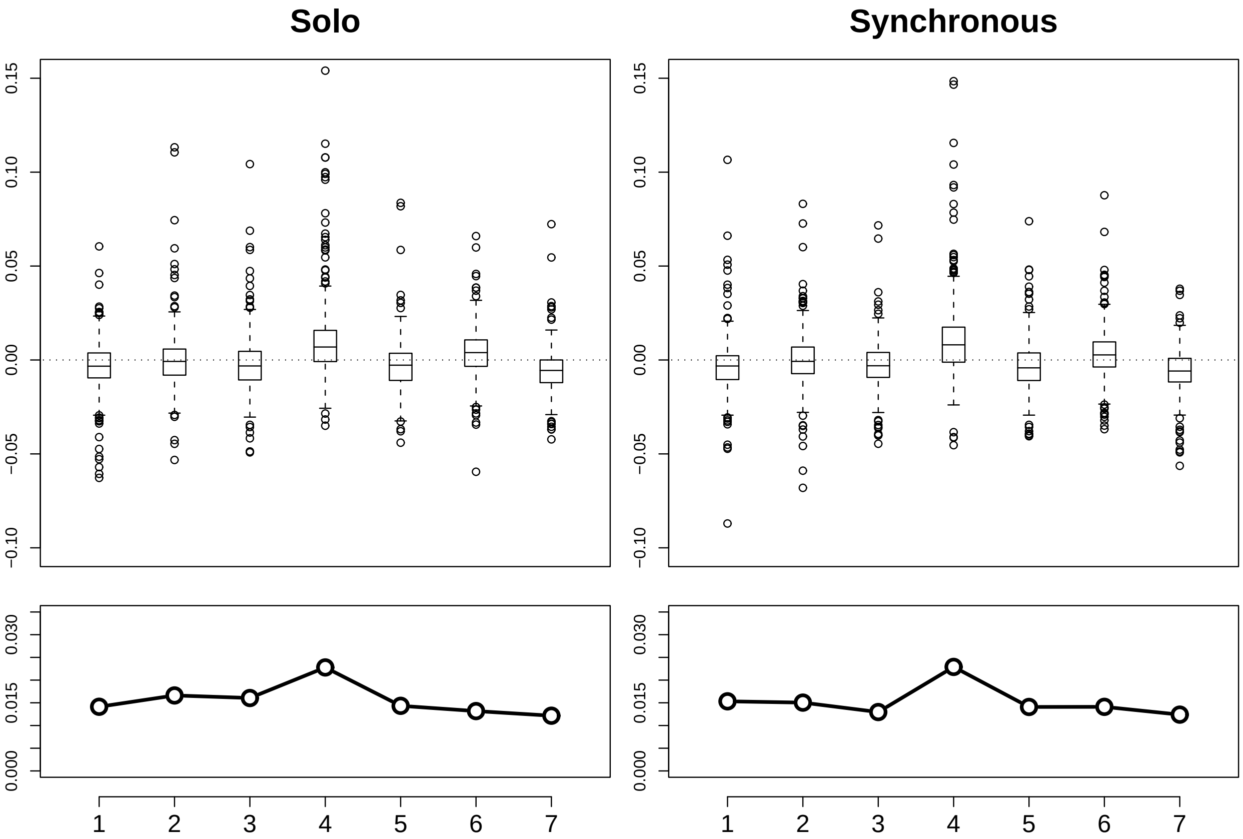

Figure 2: Top: Relative interval duration for all lists after correction and detrending. Bottom: Standard deviation of relative interval duration.

studies, a small subset of tapping data has been found to violate the assumptions of the model, and this data has typically simply been discarded (usually with dual statistical analysis of the data both with and without the problematic points, to ensure sanity). For the present data, the lag one autocorrelation was calculated for each of 641 7-interval sequences, following the procedure of Wing and Kristofferson (1973):

$$\rho_1(1) = \frac{\text{cov}(I_j, I_{j-1})}{\text{var}(I)} \quad (2)$$

where

$$\operatorname{cov}(I_j, I_{j-1}) = \frac{\sum\limits_{j=2}^{N} (I_j - \overline{I})(I_{j-1} - \overline{I})}{N-1} \tag{3}$$

and

$$\operatorname{var}(I) = \frac{\sum\limits_{j=1}^{N} (I_j - \overline{I})^2}{N} \tag{4}$$

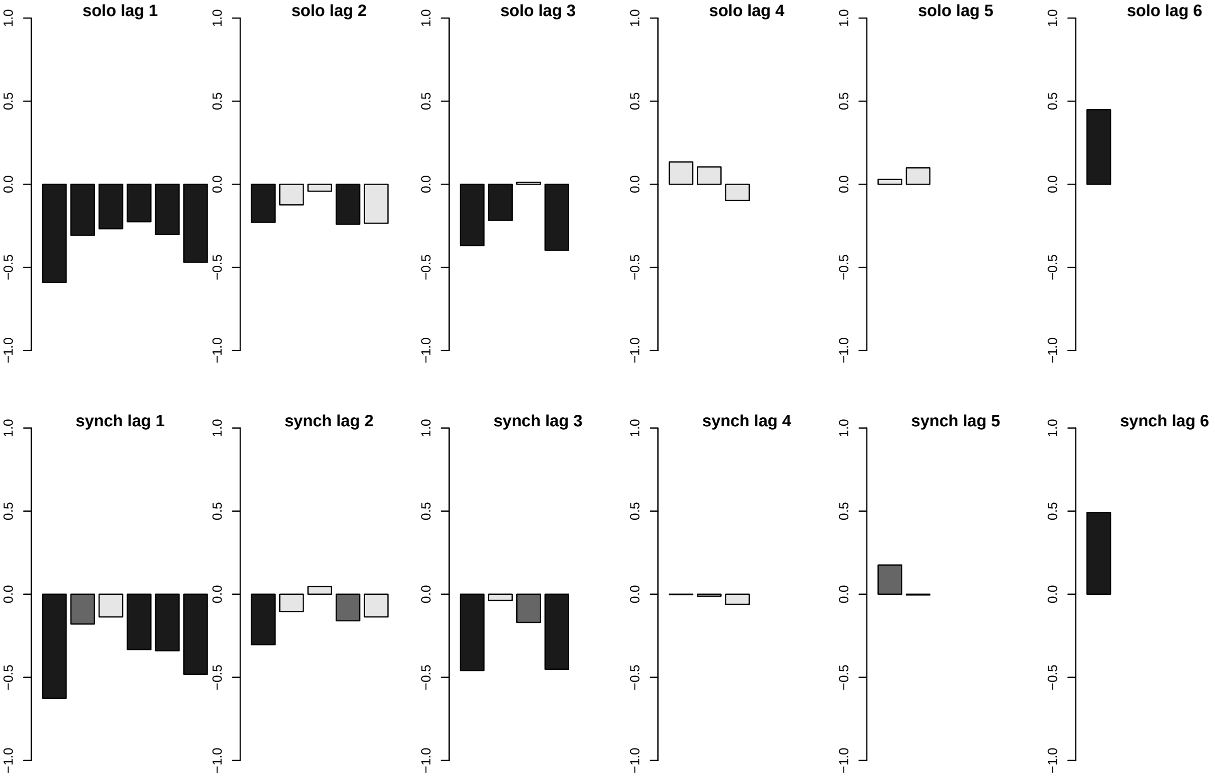

Figure 3: All pairwise correlations in solo (top) and synchronous (bottom) conditions. The top left panel shows correlations between intervals 1 and 2, 2 and 3, 3 and 4 etc. Significant correlations at p<0.001 are shown in black, and at p<0.01 in grey.

Some 45% of the coefficients lie either below -0.5, or above 0, either of which represents a violation of the assumptions of the Wing and Kristofferson model. A decomposition into hypothetical clock and motor variances is not justifiable in this case. It was therefore decided to examine the serial dependencies among intervals somewhat more closely.

For each pair of intervals within each condition, the Pearson product moment correlation was calculated, as shown in Figure 3, in which the top row shows all pairwise correlations for the solo data, and the lower row shows correlations for the synchronous data. Thus the top left panel shows correlations between intervals 1 and 2, 2 and 3,. . . ,1 and 7, while the top right shows correlations between intervals 1 and 7. The black bars show correlations which are significant at $p < 0.001$, and dark grey for those significant at $p < 0.01$.

As expected, there are negative correlations in almost all positions at a lag of one. Correlations at other lags are not insignificant, however. One striking feature is a yoking of intervals 1, 4 and 7. The strong positive correlation between intervals 1 and 7 is wholly unexpected, and may be seen in conjunction with the negative correlations between intervals 1 and 4 and 4 and 7. This pattern suggests that there is non-trivial global structure to the sequence timing, but a full account of this does not yet suggest itself. From this analysis, it seems clear that serial dependencies exist at a variety of lags in these data. Interval durations are not independent, and the sequence must be regarded as being timed as a complex whole, rather than being a succession of independent intervals.

3.1 Discussion of Experiment 1

When quasi-isochronous sequences of eight trochees are produced, the resulting interval pattern deviates somewhat from isochrony. The long-short alternation is much less pronounced than that found by Rosenbaum et al (1983), but did appear in both speaking conditions. Furthermore, there were considerable dependencies among non-adjacent intervals which are not readily accounted for

Several factors make the quasi-isochronous interval production here different from that normally collected in tapping experiments. Subjects were asked to produce regular productions, but no metronome or synchronization phase was provided. The sequences produced were short (as will in general be the case for speech data, due to breathing requirements). The data were collected under conditions in which there was a certain amount of uncertainty about the upcoming sequences, because lists which incorporated words of other stress patterns were admixed with the regular lists analyzed here. Finally, subjects did not have extensive practice at the isochronous interval production task.

Before jumping to conclusions about apparent global timing structure in such sequences, it seems natural to ask to what extent the apparent global constraints on timing seen here might stem from either the lack of a pacing signal, or the high degree of rhythmic uncertainty under which the lists were produced. In a follow-up experiment, some steps towards creating conditions more akin to those of well-known tapping studies were undertaken. A pacing signal was introduced, and three comfortable speaking rates were specified. Furthermore, the experiment was structured so as to maximize subjects' certainty about rhythmic patterning.

4 Experiment 2: Low Uncertainty, Paced, Well-Practiced Production

4.1 Methods

Six subjects took part, and each was recorded with every possible co-speaker, producing 14 dyads in all (scheduling problems precluded one data set from being collected: speakers 2 and 6). A revised set of lists was prepared (see Appendix for details), comprising three lists, each of which was used in all 8 possible rotations, so that each word appeared in each position an equal number of times. The net effect of this should be to average out any effects due to systematic misestimation of P-centres, obviating the need for the correction factor used in Experiment 1.

Subjects were recorded in separate sessions reading lists alone, or with each of the 5 co-speakers. Within each session, subjects read each of the three lists in each of 8 possible rotations at three specified rates, yielding 72 readings per session. List order was randomized.

Because of the presence of a pacing signal, it was desirable to include some variety in the pacing tempo. Speaking rate was therefore controlled by providing a metronome signal consisting of 4 evenly spaced beeps at one of 300, 400 or 500 ms separation. Subjects were instructed to listen to the 4-beep sequence, and then to begin reading the list isochronously at the same rate. All lists were pure trochaic lists, so there was no uncertainty about stress patterns, and subjects were given practice beforehand until they decided they were familiar and comfortable with the task.

4.2 Experiment 2: Results

This experiment combined factors of speaking condition and rate. As detrending had had a significant effect on the data in the previous experiment, each interval sequence once more had any linear component removed before further analysis. A repeated measures ANOVA showed only a main effect of serial position [F(6,15277)=64, $p<0.001$]. As neither the ANOVA nor visual inspection of the data revealed any differences between speaking rates or speaking conditions, the data were then analyzed as a whole, with a focus on serial position effects.

Figure 4 shows intervals and standard deviations for the entire data set. The most obvious difference between these data and Experiment 1 is the marked reduction in variability for all

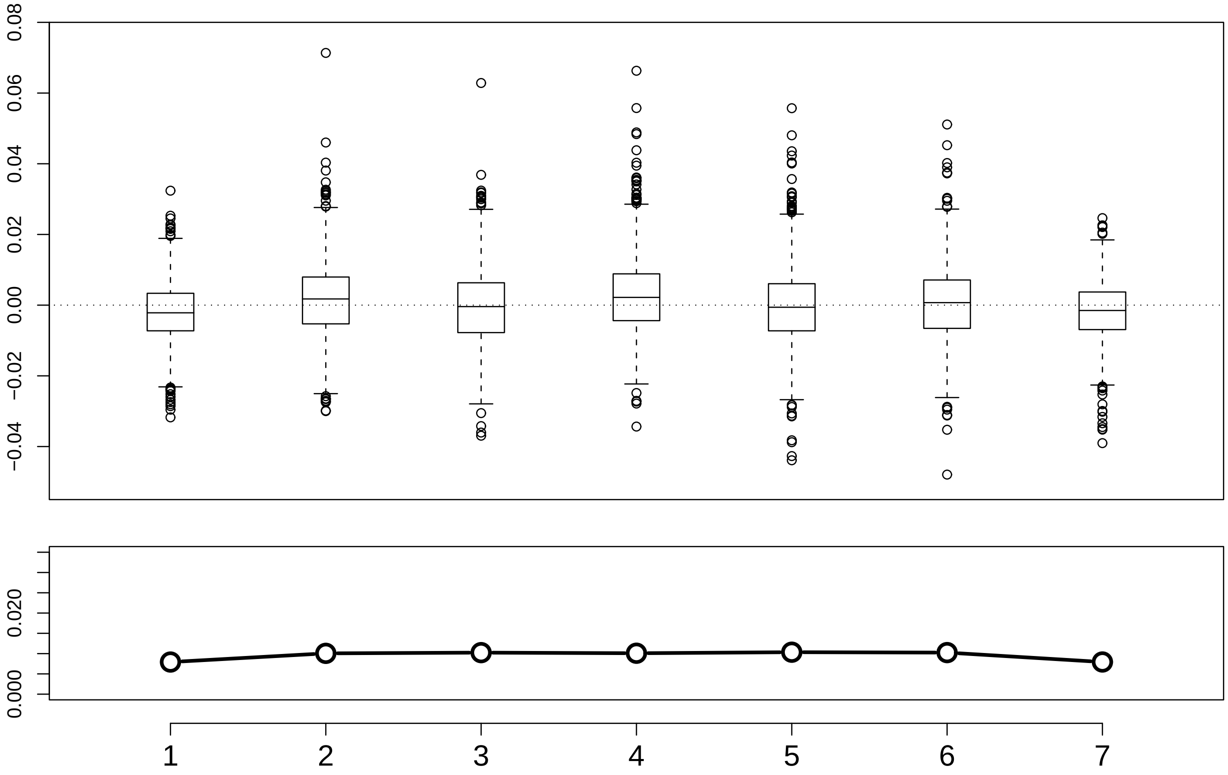

Figure 4: Experiment 2. Interval duration (top) and standard deviation (bottom). The scale of the lower panel is the same as in Figure 2, while that of the upper panel is different.

intervals, including interval 4, which previously showed much greater variability than all other intervals.

The lag one autocorrelations were again examined to test the applicability of the Wing and Kristofferson model. Again, some 45% of autocorrelation coefficients violated the assumptions of the model.

All possible within-sequence correlations were again explored, as shown in Figure 5. Negative lag one correlations predominate, as in the synchronous condition of Experiment 1, but once more

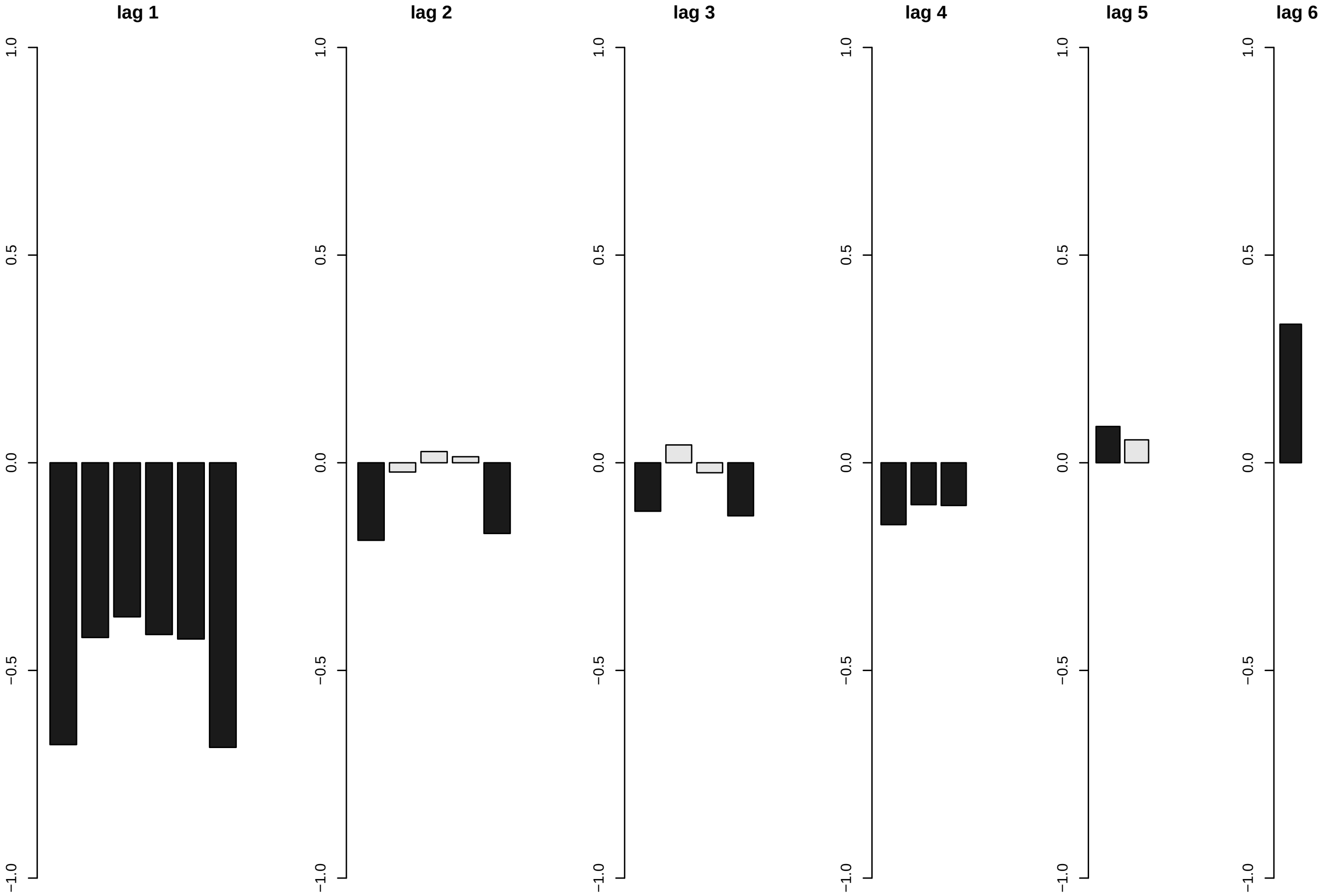

Figure 5: Experiment 2. Correlations at all possible lags (all rates and speaking conditions).

there is a strong positive correlation between intervals 1 and 7 which is not predicted by any current model. Among the other long-range dependencies, correlations between intervals 1 and 4 and 4 and 7 are once more prominent, while of the remainder, only intervals 1 and 3 were also highly significant in Experiment 1.

5 Discussion

Sequences of eight words, each with a strong-weak rhythm, were produced under a range of circumstances. Somewhat surprisingly, the patterns obtained did not seem to depend on speaking condition (solo or synchronous) or on speaking rate (with IOIs of 300, 400 and 500 ms). The only difference found in speaking conditions was a slight tendency for speakers to speed up within a sequence in the synchronous condition in Experiment 1. This was not found in Experiment 2.

Although subjects were attempting to produce an isochronous sequence, there were non-trivial and systematic deviations from isochrony found in both experiments. While adjacent intervals were, by and large, anti-correlated, as expected, there were other significant correlations between non-adjacent intervals. In both experiments, intervals 1 and 4, and 4 and 7 were negatively correlated, while intervals 1 and 7 were positively correlated. There was also a negative correlation between intervals 1 and 3 in both experiments. Although correlations with a significance of $p < 0.001$ were examined in each case, it would seem incautious to read too much into relatively weak (though significant) correlations, especially when they were not consistent across the two experiments. The very strong correlation between intervals 1 and 7 cannot be overlooked though, and is not satisfactorily explained by any model the author is aware of.

Both the original data of Rosenbaum et al (1983) and the present data suggest that short sequences which lend themselves to hierarchical grouping, will be produced with alternating long and short intervals. Both studies used sequences of eight responses, which might be particularly well suited to inducing this kind of metrical pattern. A systematic study of inter-interval dependencies in short sequences at a variety of lengths, and in both the spoken and manual domains, would now seem to be strongly motivated.

Acknowledgements

This research was funded in part by an Enterprise Ireland Basic Research Grant to the author. Particular thanks are due to three anonymous reviewers, who provided very insightful and helpful ciriticism. Thanks also to Rebecca Grimes, who collected the data in Experiment 2.

References

Abercrombie, D. (1967). *Elements of general phonetics.* Aldine Pub. Co., Chicago, IL.

Chen, Y., Repp, B. H., and Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual feedback conditions. *Human Movement Science*, 21:515–532.

Cummins, F. (2003a). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2):139–148.

Cummins, F. (2003b). Rhythmic grouping in word lists: competing roles of syllables, words and stress feet. In *Proceedings of 15th ICPhS 2003*, pages 325–328, Barcelona.

Cummins, F. (2004). Synchronization among speakers reduces macroscopic temporal variability. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

Cummins, F. and Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, 26(2):145–171.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11:51–62.

Hazeltine, E., Helmuth, L. L., and Ivry, R. B. (1997). Neural mechanisms of timing. *Trends in Cognitive Sciences*, 1:163–169.

Howell, P. (in press (2001)). The EXPLAN theory of fluency control applied to the diagnosis of stuttering.

In Fava, E., editor, *Clinical Linguistics: Language Pathology, Speech Therapy, and Linguistic Theory*, Clinical Issues in Linguistic Theory. John Benjamins.

Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology*, 6:851–857.

Ivry, R. B. and Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):3–18.

Ivry, R. B. and Keele, S. W. (1987). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience*, 1(2):136–152.

Lashley, K. S. (1951). The problem of serial order in behavior. In Jefress, L. A., editor, *Cerebral Mechanisms in Behavior*, pages 112–136. John Wiley and Sons, New York, NY.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5:253–263.

Max, L. and Yudman, E. M. (2003). Acuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *Journal of Speech, Language, and Hearing Reaearch*, 46:146–163.

Meegan, D. V., Aslin, R. N., and Jacobs, R. A. (2000). Motor timing learned without motor training. *Nature Neuroscience*, 3(9):860–862.

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83:405–408.

Rosenbaum, D. A. (2001). Time, space and short-term memory. *Brain and Cognition*, 48:52–65.

Rosenbaum, D. A., Kenny, S. B., and Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 9(1):86–102.

Scott, S. K. (1993). *P-centers in Speech: An Acoustic Analysis*. PhD thesis, University College London.

Wing, A. M., Church, R. M., and Gentner, D. R. (1989). Variability in the timing of responses during repetitive tapping with alternate hands. *Psychological Research*, 51:28–37.

Wing, A. M. and Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics*, 14(1):5–12.

Appendix: Word Lists Used

The following word lists were used in the experiments reported here:

Experiment 1

- camper batty garden boiler carry shadow toaster bully
- shining taming taper season mucky barter weedy funny
- tango lighter daddy wiper pony cutter pinky mango
- sender batter dagger dinky leper shepherd pity larger
- shady final patter folly bingo banker many teapot
- poster sheila rapid poking gutter needy dinner shadow

Experiment 2 The following three lists were used in all possible rotations, thus along with “mitten never turtle...”, there was “never turtle... mitten”, “turtle....mitten never” etc. Each list contains the same set of initial consonants. Complex syllable structures were avoided, as are were medial consonant clusters.

- mitten never turtle damage buggy gossip palate kennel
- bobby copper ticking gallop nodded middle doughnut parrot
- beggar dollop tacky codded gaiter mutton panel knitting