

ENTRAINING SPEECH WITH SPEECH AND METRONOMES

FRED CUMMINS

(Department of Computer Science, University College Dublin)

RESUMO Duas novas abordagens experimentais para o estudo da fala são descritas. Na abordagem *Speech Cycling*, os locutores procuram casar o ritmo de sua fala com aquele de um metrônomo. Eles mostram fortes vieses que privilegiam a produção de padrões rítmicos simples baseados no encadeamento de um número inteiro de pés acentuais compreendidos no ciclo de repetição da frase. Na abordagem *Synchronous Speech*, dois locutores lêem um texto conjuntamente, tentando manter o sincronismo entre si. Novos resultados, fundamentados em vinte pares de locutores, são apresentados e demonstram que a prática não melhora de maneira acentuada o sincronismo, que os locutores usam informação visual para se sincronizarem entre si e que um índice independente de fluência de leitura é somente um fraco preditor de sincronismo. Tomados em seu conjunto, esses experimentos ilustram o uso de estímulos para induzir os locutores ao acoplamento, revelando assim tendências e limitações do sistema de produção de fala.

1. INTRODUCTION

Two novel experimental approaches to studying speech are to be outlined. One, documented better elsewhere in this report, is called *Speech Cycling*. The other, called *Synchronous Speech*, is of more recent vintage. What both approaches have in common is that they attempt to force the speech production system to produce speech which has been altered so as to make clear the underlying biases and proclivities of the generating system. Both try to influence the production system by entraining it to a stimulus, and they differ primarily in the means by which the speech production system is entrained.

Entrainment, or the yoking together of two dynamical systems, can take two principal forms. It can be a uni-directional forcing of a system, in which an external agent influences the dynamics of the system without being reciprocally affected, as for example when a cardiac pacemaker injects current into heart muscle, thereby significantly affecting its dynamics, or when a dancer's movements become entrained to the beat of a drum¹. Or two systems can become mutually entrained, if they affect and are in turn affected by each other. The two hands of a drummer are mutually

¹ Of course the notion of pure uni-directional forcing is often no more than a reasonable approximation. The pacemaker is affected by the heart's action to a small, but hopefully negligible, degree.

entrained, as are the limbs of the dancer or the undulating arms of the football crowd doing the “Mexican wave.”

Speech is susceptible to both forms of manipulation. In the first set of experiments, Speech Cycling, the speaker tries to match the spoken utterance to the temporal signature of an auditory metronome. The metronome is unforgiving, and the interesting data are those where speakers produce speech timing which differs in readily interpretable fashion from the given stimulus. In the second set, Synchronous Speech, two speakers speak together, and the resulting speech reflects their mutual influence on one another. The relationship is symmetrical.

2. FORCED COUPLING: SPEECH CYCLING WITH A METRONOME

2.1. Background: Finger cycling

A very substantial body of work by Scott Kelso and colleagues has investigated the properties of a movement system comprising one finger on each of the right and left hand, where the fingers are required to cycle or wag rhythmically, as illustrated in Figure 1. As each finger exhibits cyclic behavior, its position at any moment in time can be described by a phase measurement², and so the relative positions of the two fingers can be described by a simple phase difference. Without extensive training, two and only two stable patterns of behavior are found in this system; these are characterized by a stable phase difference of 0.0 and 0.5, and these two patterns can be called the in-phase and anti-phase patterns respectively. (This insight can be readily replicated by the interested reader by attempting to wag two fingers or hands in as many different ways as possible.)

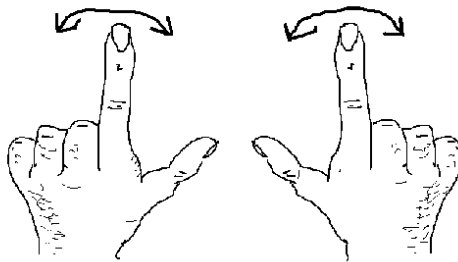


Figure 1: Movement system comprising two fingers wagging rhythmically. Other planes of movement or hand orientations are equally effective.

² I use a notation convention for phase in which a point within a cycle is described by a number in the range [0--1], with 0 equivalent to 1. Other conventions use 0..360 degrees, or 0..2 π radians.

The dynamical signatures of this system have been captured in an influential model, and the combination of model and experiment have been extended to several other systems (see Kelso (1995) for a summary, and Haken et al (1985) for the original model). Of particular interest here is a series of experiments done by Yamanishi and colleagues (1980) and by Tuller and Kelso (1989), in which subjects attempted to match a specified phase relation dictated by a pair of flashing lights (Figure 2) with that of their wagging fingers. When the lights flashed synchronously, the in-phase pattern was required and produced. When they alternated regularly, the anti-phase pattern was produced. Any other phase relationship between the lights resulted in performance which was skewed toward one or other of the two known stable patterns.

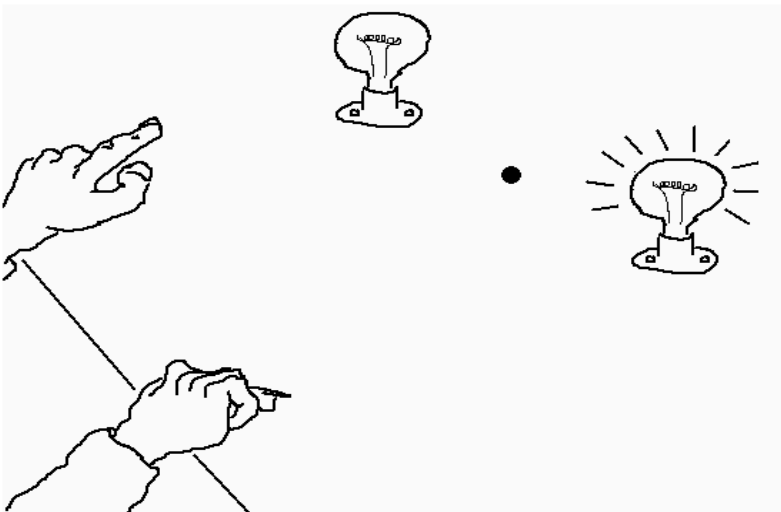


Figure 2: Movement system comprising two fingers wagging rhythmically. After Yamanishi (1980).

One aspect of this experimental procedure intrigued me, although it was of little interest for the original work studying the coupled fingers. If the two stable states of the movement system had not been known in advance, this experiment would have revealed them plainly. The original speech cycling experiments (Cummins, 1996; Cummins and Port, 1998) constitute an attempt to transfer some of this experimental approach to the speech domain in a repetitive, and hence cyclic, task. Unlike the finger wagging task, however, the existence of stable modes of production, and their form was not known a priori. The ability to specify target phase relations and identify systematic biases in the responses is the key result in what follows.

2.2. Identifying stable models of rhythmic production in English

In order to transfer some of the task structure found in the finger-wagging literature into the speech domain, a repetitive task needed to be designed. In all speech cycling experiments, the core of the task is to ask subjects to repeat a short phrase ('Examples: "big for a duck") in time with a metronome. The metronome typically emits a series of clicks or beeps with a basic period of about one second. Subjects are asked to repeat the phrase such that the onset of the first word occurs together with the metronome click. Various manipulations of the metronome signal are then possible (as outline elsewhere in this collection), which provide evidence for the various stable manners in which a phrase can be repeated.

A more direct investigation of the number and form of stable forms of repetition is possible, however, by taking a cue from the paradigm illustrated in Figure 2. In Targeted Speech Cycling, the metronome signal consists of two alternating tones, which provide targets for two distinct syllables in the phrase being repeated. Two examples are shown in Figure 3, in which a subject repeats the phrase "big for a duck" and attempts to align the onset of "big" with a high tone and the onset of "duck" with a low tone. In the upper example, the low tone occurs half way through the overall Phrase Repetition Cycle (PRC), and the stressed syllables ("big" and "duck") in the resulting speech form a regular series. Subjects find this pattern particularly easy to produce.

In the lower pattern of Figure 3, the low tone divides the PRC into a 2:1 ratio. Subjects can produce this pattern by introducing a nonce stress on the word "for", so that the phrase is stressed as "BIG..FOR a DUCK", and the stresses are again produced in a regularly timed series.

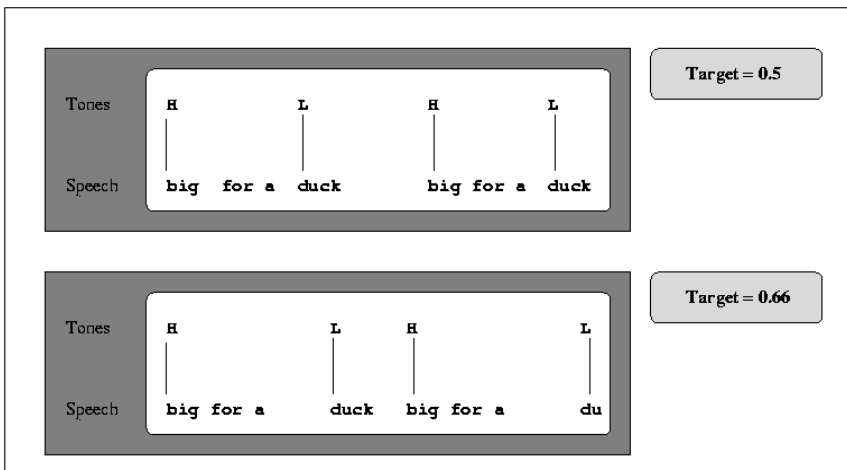


Figure 3: Targeted speech cycling. Subjects hear an alternating series of high and low tones, and attempt to align specific syllables with each. The target refers to the point at which the L-tone divides the overall H—H cycle.

In a series of experiments with native English speakers (Cummins and Port, 1998), we found that when asked to repeat short phrases similar to “big for a duck”, each with four syllables, the first and last being normally stressed, there were three and only three ways in which subjects could repeat the phrase. These three patterns each correspond to a simple rhythmic organization of the stress syllable onsets within the phrase. The three patterns are illustrated as musical rhythms in Figure 4, with the phrase “beg for a dime”. In the first pattern, the phrase is spoken as [BEG for a DIME <rest>], the second is [BEG for a DIME] and the third is [BEG FOR a DIME]. The reader is encouraged to try to mimic these three patterns, and then to explore the space of other possible patterns, both overtly rhythmic, and arrhythmic.

Each of these basic patterns is based on a nesting of two or three stress feet within the overall PRC. The importance of the stress foot (interval from the onset of one stress syllable to the next) is well documented in English phonology (Lieberman and Prince, 1977; Dauer, 1983) but it has hitherto proved difficult to link such supposed constraints to real time performance. That is, the term “rhythm” as used by phonologists typically refers to an alternation of strong and weak units, linearly ordered, and perhaps allowing of several different levels of strength (Hayes, 1995). It does not refer to the actual duration (in milliseconds) of speech units. Attempts to translate intuitions about rhythm into direct observations in real time have failed to establish a convincing connection (Lehiste, 1977; Dauer, 1983). In targeted speech cycling, we see that appropriate task constraints can reveal underlying principles of organization which are not observable in regular speech.



Figure 4: Musical notation for the three stable ways of repeating the phrase “beg for a dime”.

The existence of simple rhythmic forms of organization provided a direct link between findings about coordination among the limbs and coordination in the morphologically unrelated speech production system, suggesting common principles of coordination underlying the tasks accomplished by each system. Another feature which is common to both systems (limb control and speech production) is that both typically produce skilled movement which is not strictly regular, and the principles of elementary rhythmic organization are only revealed when the system “dumbs down” to do a simple repetitive task.

2.3. Difficulties in extending Targeted Speech Cycling to other languages

The stress foot has played such a large role in phonological typology that it was natural to extend the speech cycling paradigm to other languages, looking for commonalities and differences in rhythmic organization. Tajima conducted a series of experiments using native Japanese speakers, and uncovered several tantalizing phenomena reported

together in Tajima (1998). However a basic difficulty with creating experiments for Japanese speakers was the complete absence of stress in Japanese. While Japanese has phrasal accents that serve some functions similar to stress in English, their acoustic manifestation is strictly limited to intonation, and is therefore not as tightly bound to the temporal organization of speech as the stressed syllable in English. For this reason, Tajima used untargeted speech cycling, in which a stimulus tone is only provided for the phrase onset. This limitation means that one cannot scan directly for all stable forms of production as in Yamanishi et al. (1980) and Cummins and Port (1998).

In an unpublished pilot study, Cummins asked English, Italian and Spanish subjects to attempt to repeat short phrases, prosodically similar to those in Cummins and Port (1998), with temporal targets of 0.5, 0.6 and 0.66. A great amount of intersubject variability precluded final analysis of the results, but it was noteworthy that most of the Italian subjects and all of the Spanish subjects had great difficulty with the task. They differed from the English subjects, not so much in the patterns they produced, but in the facility with which they interpreted the task instructions. Spanish subjects in particular needed up to 45 minutes of explanation, demonstration and practice before they were willing to attempt the task, while English speaking subjects were typically ready to go after a brief explanation and a single demonstration.

The evidence from English speaking subjects has consistently been that they are coordinating a foot-level temporal structure within an overall phrasal unit. Although both Spanish and Italian have a stressed/non-stressed distinction, neither language has been described as having a stress foot which behaves as the English stress foot, and neither language displays the range of reduction phenomena found in English unstressed syllables, which Dauer and others have suggested contribute to the perception of English as stress-timed. It seems likely that the Spanish and Italian speakers did not have the stress foot as a stable production unit which could be harnessed in service of the task demands.

Both the finger wagging tasks and the above speech cycling tasks use a metronome to force the system of interest into its “natural” forms of organization. The metronome is not reciprocally affected. Cyclic properties of the underlying motor control systems are revealed, but we know that both motor systems are capable of precise aperiodic coordination as well. To further study natural modes of timing within the speech production system, the second set of experiments reported here uses speech to entrain speech.

3. MUTUAL ENTRAINMENT: SYNCHRONOUS SPEECH

The synchronous speech paradigm is illustrated in Figure 5. We adopt the simple expedient of having subjects read prepared texts in synchrony with one another. Typically, two subjects read a short text aloud together, each endeavoring to maintain synchrony with the other. We call speech elicited in this fashion *Synchronous Speech* (hereafter, SS).

A priori consideration of the SS elicitation condition suggests that the goal of speaking in tight synchrony with another speaker can be achieved only if the speakers manage to each make their speech temporally predictable for the other. This in turn means agreeing on common timing patterns for speech, and suggests that speakers must exploit their shared knowledge of speech timing, dispensing with idiosyncratic and unpredictable elements in their speech.

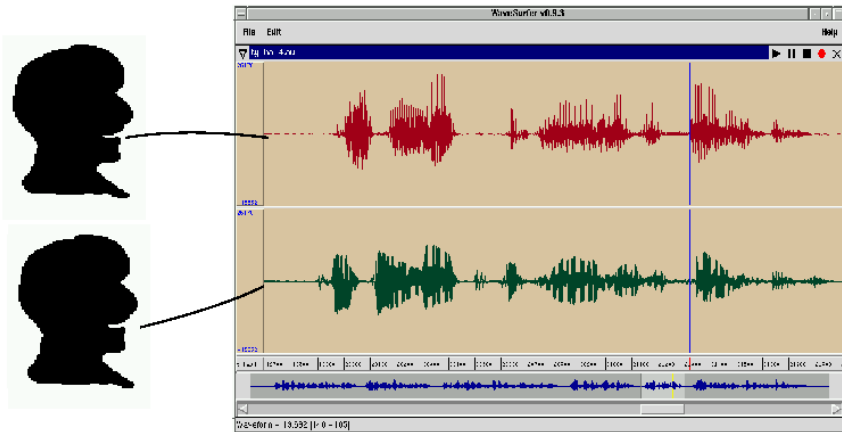


Figure 5: Elicitation of synchronous speech. Two subjects read a prepared text in synchrony. They are recorded onto two parallel channels.

An insightful analogy from the sphere of musical performance presents itself: The 14th violinist in an orchestra must play “closer to the score” than is required of a soloist. We strongly expect the playing of a competent ensemble player (the 14th violinist) to combine both nominal note durations as provided by the score, with conventional aspects of phrasing, such as phrase-final *rallentando*. By requiring subjects to read in synchrony, we similarly force them to restrict themselves to a common interpretation of speech timing³.

Initial investigation of SS confirmed the following (Cummins, 2001):

- Subjects can read in synchrony with little or no practice,
- Subjects agree on major phrasing (placement of silent pauses) in a synchronous reading,
- Synchrony is greater when reading together with a live speaker than when reading along with a recording of another speaker, and
- Synchrony is less marked at phrase onsets than at a late point in the phrase.

³ Just what the “score” is for speech is still an open question.

This set of results suggested many questions, some of which have recently been addressed in the work below. Specifically, we turn to the following questions about synchronous speech, recognizing fully that many other issues remain to be addressed:

- Does synchrony improve as subjects become more familiar with the task?
- Does synchrony improve as subjects become more familiar with a given text?
- Does visual information play a role in ensuring synchrony among speakers?
- Does reading fluency affect the degree of synchrony achieved?

4. METHODS

Twenty seven pairs of subjects participated. All subjects were from the province of Leinster in the East of Ireland, and were recruited from the population of University College Dublin. Subjects presented in pairs, so for the most part, speakers within a pair were familiar with each other. No attempt to control for familiarity was made.

Each subject pair took part in a single recording session lasting about one hour. In the session they were asked to read a series of texts as outlined below. For each text, the procedure followed was to allow subjects to read the text silently to themselves. Then the text was read by one of the subjects alone, by both subjects in synchrony, and finally by the second subject alone. The choice of first-reader alternated from one text to the next. Unless otherwise stated below, subjects were seated facing one another about 2 meters apart. Each wore a head mounted near-field unidirectional microphone (Shure SM10A) and recordings were made onto the right and left channels of a single stereo soundfile, via a portable DAT recorder.

The order in which texts were read, and the associated conditions, was as shown in the following table. Texts A,B,C and D were 4 simple versions of Aesop's fables, similar to and including a version of the North Wind and the Sun. The four texts are reproduced in the appendix. Closer discussion of each condition is given below.

Condition name	Text used	Comments
warmup	Aesop's fable	Data not analysed.
WITH_VIS_1	One of A,B,C, and D	Subjects could see one another
NO_VI_1	One of A,B,C, and D	Subjects could not see one another (seated back to back).
...	...	Other material intervened here which is not reported herein.
PRE_PRACT	Rainbow Text, 1 st paragraph	Seating etc as WITH_VIS_1
...	Rainbow Text, 1 st paragraph	Practice readings, alone and in synchrony
POST_PRACT	Rainbow Text, 1 st paragraph	Seating etc as WITH_VIS_1
WITH_VIS_2	One of A,B,C and D	Subjects could see one another
NO_VIS_2	One of A,B,C and D	Subjects could not see one another

4.1 Experiment 1: Synchrony and phrasal position

Readings from conditions WITH_VIS_1 and WITH_VIS_2 were analysed. Here subjects had read novel texts while facing one another. From each reading, asynchrony at points at the beginning, middle and end of major phrases within the text was measured (further details in Appendix A) by taking the absolute value of the temporal difference between corresponding vowel onsets. For each paired reading of a single text, median asynchrony in onset, medial and final position was computed for that trial. Medians, rather than means were used, as occasional dysfluencies are known to occasionally introduce outliers into the data. Table 1 gives the mean and standard deviations of these median values.

Table 1: Mean (s.d.) of the per-trial median asynchrony in msec broken down by phrasal position and condition. For each reading, a median value was computed. The table reports the means (s.d.s) of these medians.

Condition	Phrase onset	Phrase middle	Phrase end
WITH_VIS_1	62 (30)	40 (20)	44 (25)
WITH_VIS_2	61 (38)	45 (27)	40 (24)

From previous work, we had the expectation that synchrony would be less pronounced at phrase onsets than medially or finally, and so two comparisons were planned, each with a strong a priori hypothesis. Firstly, we predicted that onsets would be less synchronous than either medial or final measurements. Secondly, we predicted that medial and final positions would not differ in synchrony, as our previous experience had shown an effect which appeared to be restricted to phrase onset position.

The two conditions analyzed differed in that WITH_VIS_1 was obtained close to the start of the recording session, immediately after a single warmup reading. In contrast, WITH_VIS_2 was collected near the end of the session, at which point subjects had much more familiarity with the task of speaking in synchrony. The texts in the two conditions were novel and were selected from texts A,B,C and D (Appendix A).

A two factor ANOVA with practice (two levels) and position (three levels) was done, which showed a main effect of position, [$F(2,156)=8.81, p<0.001$], but not of practice, [$F(1,156)=0.00, n.s.$]. The interaction was not significant.

Because the effect of position in phrase was anticipated from previous results, planned comparisons of Phrase Onset versus Phase Middle and End, and of Phrase Middle with Phrase End were done. Lags at phrase onset were significantly different from lags at the other two positions [$F(1,156)=17.6, p<0.001$], while there was no significant difference between lags in middle and end positions [$F(1,104)=0.004, n.s.$]. In all subsequent analyses, therefore, we combine data in medial and end positions, but perform separate analysis in onset position, where considerable differences in degree of synchrony are to be expected.

4.2 Experiment 2: The role of visual information

In all previous experiments we had done, subjects were free to look at each other, though they did not give the impression of doing so, and they reported themselves that they were attending only to the written texts. In order to assess the possible role of visual information in synchronization, we followed the conditions of the previous experiment with readings in which subjects sat back-to-back, so that no visual contact was possible. An initial check, comparing phrasal position and degree of practice was made to verify that the no-vision conditions did not display a strong practice effect, and in fact there was no main effect of practice, nor was there an interaction. Therefore, in what follows we combine data from WITH_VIS_1 and WITH_VIS_2 into one set WITH_VIS, and likewise combine NO_VIS_1 and NO_VIS_2 into one set NO_VIS. Mean (s.d.) data summarizing trial medians is given in Table 2.

We conducted an ANOVA with visual information (present, absent) and position in phrase as factors. Main effects of both position in phrase and of visual information were found [vision: $F(1,307)=13.67$, $p<0.001$; position: $F(1,307)=13.67$, $p<0.001$], and the interaction was not significant [$F(1,307)=1.07$, n.s.]. Visual information was therefore of some help in synchronising with a co-speaker.

Table 2: Mean (s.d.) of the per-trial median asynchrony with and without visual information. Times are in ms.

Condition	Phrase onset	Phrase middle/end
WITH_VISION	63 (34)	42 (24)
WITHOUT_VISION	80 (30)	51 (30)

4.3. Experiment 3: Role of text familiarity

In Experiment 1 we found no effect whatsoever for practice at the task of reading in synchrony. In that case, however, a different, and novel, text was used in each condition. It is possible that a practice effect would be found if subjects read only a single text repeatedly. To this end, we had subjects read the first paragraph of the Rainbow Text (Appendix A) in the usual manner, i.e. first one speaker alone, then both in synchrony, and finally the second speaker alone. We then had them repeat this sequence 6 more times for practice, and recorded a final set of readings.

An ANOVA with practice (before, after) and position in phrase as factors showed main effects of both factors [practice: $F(1,158)=7.65$, $p<0.01$; position: $F(1,158)=44.54$, $p<0.001$], but no interaction [$F(1,158)=0.086$, n.s.]. Practice with a specific text, therefore, does indeed improve synchrony among speakers, in marked contrast to practice at the task alone, using novel texts for each reading. Mean lags, based on median values per reading are shown in Figure 6. It can be seen that the improvement in synchrony is modest, being of the order of 10 ms in each case.

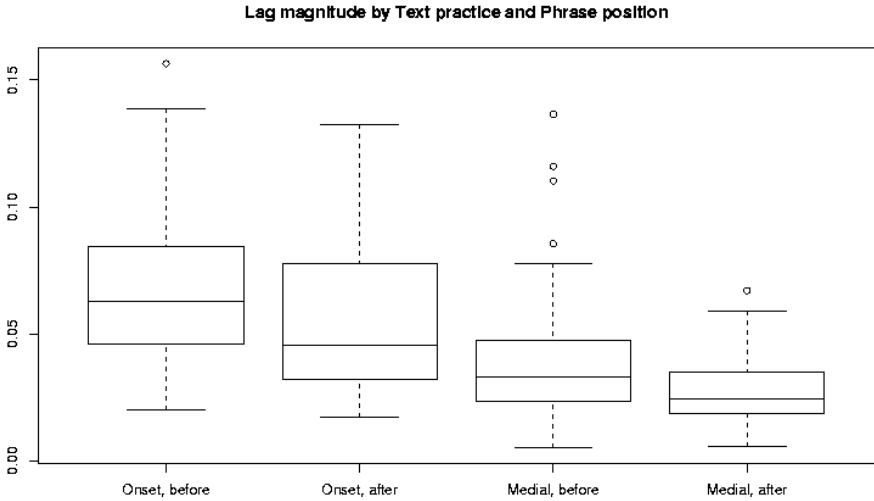


Figure 6: Effect of practice at phrase-onset and phrase-medial positions. Boxplots show medians (central horizontal line) and inter-quartile ranges (box outlines). The whiskers show the range of data points falling within 1.5 times the inter-quartile range, and any outliers falling outside this range are plotted with open circles.

4.4. Experiment 4: Reading fluency as a predictor of synchrony

It seems likely that the degree of synchrony attained by a pair might be determined in part by the facility with which each member of the pair reads aloud. In order to assess this, a coarse index of reading fluency was obtained for each speaker independently by considering those texts in which the speaker was the first to read a novel text alone, and simply counting obvious speech errors, without any attempt to subclassify those errors. For each pair, then, the error index was the sum of the error indices for the two speakers.

Figure 7 shows the relationship between errors for each pair and asynchrony. A positive correlation is evident in onset position [$r = 0.46$, $p < 0.05$], but not in medial position [$r = 0.24$, n.s.].

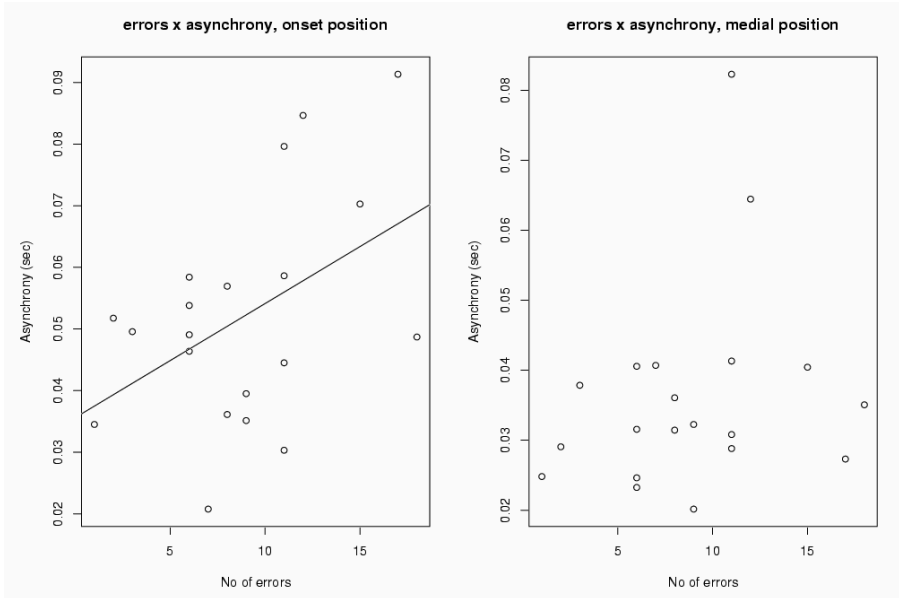


Figure 7: Asynchrony as a function of errors made when reading alone.

5. SYNCHRONOUS SPEECH: WHAT WE KNOW SO FAR

We now have initial answers to some of the questions raised in the previous section. Practice, it would appear, is not necessary for completing the task of speaking in synchrony, and even extensive practice on both task and text resulted in only slight improvement in synchrony. Practice at the task alone, using novel texts each time, did not produce any marked improvement.

The process of synchronization appears to be based on phrase-length units. We see this in several ways. Firstly, synchrony is not as good and is more variable at phrase onsets than either medially or finally, and there is no difference between these latter two positions (Expt 1). In addition, the combined fluency scores for a pair predicted synchrony at onsets, but not medially (Expt 4). Collectively these results suggest a process whereby synchrony extends over phrase-length units. It is established rapidly, over the

So subjects can read with a high degree of synchrony, agreeing on phrasing and matching the fine temporal detail of their co-speakers with considerable accuracy. These findings suggest more questions than they answer. In particular, we still need to uncover the informational basis for the strong entrainment observed: to what extent are speakers mutually entrained and to what extent are they relying on the production of unmarked or default temporal values in their own speech and that of their co-speakers? Is the process of entrainment a mutual one, or does one person predominate and the other follow?

As we discover more about synchronous speech, we are nearing the development of a novel tool for revealing the shared knowledge which speakers have about their language. For example, in a recent set of experiments, we have been using synchronous speech to investigate the degree to which speakers will modify their own dialect when speaking to someone from a markedly different dialect.

6. END NOTE

We have presented two different experimental frameworks, which have in common the entrainment of speech by a stimulus. In Speech Cycling, the stimulus is a regular metronome, and the relationship between stimulus and speaker is unidirectional and forcing. Speakers attempt to match the rhythm of their speech to that of the metronome, and the resulting speech provides a glimpse of the underlying rhythmic constraints operative in speech repetition. In Synchronous Speech, the stimulus is another person's speech and there is reciprocal adaptation. Here it is the proclivities of speech production which become apparent rather than its limitations. We induce speakers to reveal their common knowledge of the temporal organization of speech. Although this work is in its infancy, it promises to offer a wealth of insight into the dynamical structures underlying speech production, both their freedoms and their limits.

ACKNOWLEDGMENTS: This work was supported by a grant from the Irish Higher Education Authority. Thanks to Elena Zvonik for help with the measurements, and to Bob Port and Plinio Barbosa for helpful comments.

REFERENCES

- CUMMINS, F. (2001) On synchronous speech. *Acoustic Research Letters Online*. 3(1): 7-11.
- CUMMINS, F. & PORT, R.F. (1996). Rhythmic constraints on English stress timing. In BUNELL, H.-T. & IDSARDI, W. (editors). *Proceedings of the Fourth International Conference on Spoken Language Processing*, pp. 2036-2039. Alfred du Pont Institute, Wilmington, Delaware.
- CUMMINS, F. & 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.
- HAKEN, H.; KELSO, J.A.S. & BUNZ, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*. 51: 347-356.
- HAYES, B. (1995). *Metrical Stress Theory*. University of Chicago Press.
- KELSO, J.A.-S. (1995). *Dynamic Patterns*. MIT Press, Cambridge, MA.

- LEHISTE, I. (1977). Isochrony reconsidered. *Journal of Phonetics*. 5: 253-263.
- LIBERMAN, M. & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*. 8: 249-336.
- TAJIMA, K. (1998). *Speech Rhythm in English and Japanese: Experiments in Speech Cycling*. Unpublished PhD thesis. Bloomington, IN: Indiana University.
- TULLER, B. & KELSO, J.A.-S. (1989). Environmentally-specified patterns of movement coordination in normal and split-brain subjects. *Experimental Brain Research*. 75: 306-316.
- YAMANISHI, J.; KAWATO, M.; SUZUKI, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biological Cybernetics*. 37: 219-225.

APPENDIX A: TEXTS USED IN EXPERIMENTS

The following four texts were used in the experiments. Each text has been divided into major phrases. In each phrase, measurement points are the onset of voicing of the vowel (or sonorant) at the onsets of the syllables in square brackets; one each at the beginning, middle and end of the phrase. Measurements were made by hand using the acoustic waveform and spectrograms for guidance. No measurements are made on the first phrase. Texts A, B, C and D were randomly assigned across the 4 WITH_VIS and NO_VIS conditions, so that each text was read by each pair once only.

TEXT A

- (1) A farmer was sowing some hemp seeds in a field where a swallow and some other birds were busy collecting food. (2) “[Be]ware of that man” [sai]d the [swa]llow. (3) “[Why], what is he [do]ing?” said [the] others. (4) “That is hemp seed he is sowing; (5) be [car]eful to pick up every one of those [see]ds, or else you will re[gret] it.” (6) The [bir]ds did not list[en] to the Swallow’s ad[vi]ce, (7) [and] after some time the hemp grew [tall] and was made in[to] rope. (8) [Nets] were made from the rope and many birds that had ignored the [Swal]low’s advice were caught in the nets made [from] the hemp. (9) “What did I tell you?” said the swallow.

TEXT B

- (1) In the old days, men used to worship sticks and stones and idols, and prayed to them to give them luck. (2) [The]re was one man who had often prayed to a wooden [i]dol which he had been given by his [fa]ther. (3) Des[pi]te his praying, his luck [ne]ver seemed to [cha]nge. (4) He [pra]yed and he prayed, but [sti]ll he remained as unlucky as [ev]er. (5) [On]e day, in a rage, he went to the wooden [Go]d, and knocked it down from its pedestal with one [blo]w. (6) [The] idol broke in [two], and what did he [see]? (7) A [grea]t number of golden [coi]ns flying all over the [pla]ce.

TEXT C

- (1) One fine day it occurred to the Members of the Body that they were doing all the work and the Belly was having all the food. (2) [So] they held a meeting, and after a long discussion, they decided to go on [strike] until the Belly agreed to do its proper share of [the work]. (3) [So] for a day or two, the Hands refused to pick up food, the Mouth re[fus]ed to receive it, and the Teeth had no work to [do]. (4) [But] after a few days the Members began to find that they them[selv]es were not in a very active con[di]tion: (5) the [Hands] could hardly move, and the Mouth was all parched [and] dry, while the Legs were unable to sup[port] the rest. (6) And [so] they realized that even the Belly in its dull [qui]et way was doing necessary work for the [Bo]dy, (7) and [that] they must all work to[ge]ther or the Body will go to [piec]es.

TEXT D

- (1) The North Wind and the Sun were arguing one day about which of them was stronger, (2) [when] a traveler came along wrapped [up] in an over[coat]. (3) [They] agreed that the one who could make the traveler take his [coat] off would be considered stronger than [the] other one. (4) [Then] the North Wind [blew] as hard as he [could], (5) [but] the harder he blew, the tighter the [trav]eler wrapped his coat [a]round him; (6) and at last the North Wind gave up trying. (7) [Then] the Sun began to shine hot, and right [a]way the traveler took his [coat] off. (8) And [so] the North Wind had to admit that the [Sun] was stronger th[an] he was.

RAINBOW TEXT (1ST PARAGRAPH)

- (1) When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow. (2) [The] rainbow is a division of white [light] into many beautiful [co]lors. (3) [These] take the shape of a long round arch with its path high a[bove] and its two ends apparently beyond the hori[zon]. (4) [There] is, according to legend, a boiling [pot] of gold at one [end]. (5) [Peo]ple look, but [no] one ever finds [it]. (6) [When] a man looks for something beyond his reach, his friends [say] he is looking for the pot of gold at the end of the rain[bow].