Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Phonetics 31 (2003) 139–148
Journal of Phonetics
www.elsevier.com/locate/phonetics

Practice and performance in speech produced synchronously

Fred Cummins*
Department of Computer Science, University College Dublin, Dublin 4, Ireland
Received 24 April 2002; received in revised form 17 October 2002; accepted 21 October 2002

Abstract

Synchronous Speech is speech elicited by asking two readers to read a prepared text aloud and in synchrony. Under these conditions a remarkable degree of synchrony is maintained without the need for extensive practice. It is demonstrated that a typical inter-speaker lag is no more than about 40 ms, although this figure is somewhat higher at phrase onsets. Practice at the task of reading in synchrony does not improve performance, although practice at a specific text does improve synchrony somewhat. Speakers are shown to make use of visual information about their co-speaker, even though they are reading from a held text. Finally, it is shown that Synchronous Speech is a collaborative form of speech in which both speakers modify the timing of their speech about equally, compared with control readings.

1. Introduction

When reading a known text, speakers possess a remarkable ability to speak in synchrony with one another (Cummins, 2002). With no more instruction or practice than being asked to "read this (known) text in time with your co-speaker. 3..2..1..(go)", speakers manage to synchronize such that corresponding points in the two parallel speech channels are no more than about 40 ms apart on average. This ability, first reported in Cummins (2002), is all the more remarkable, as utterance-to-utterance variability, even within speaker, would suggest that two speakers might have great difficulty in precisely timing their spoken durations to match those of another speaker (e.g. Klatt, 1986 and many others).

There are several reasons why this ability might strike us as particularly interesting. These can be seen more clearly if we consider a close analogy from the realm of music. Both an ensemble player, such as the 14th violinist in the string section of a large orchestra, and the soloist, may play from a score. The musical score contains nominal note durations, based on the regular subdivision

*Tel.: +353-1-716-2902; fax: +353-1-269-7262.
E-mail address: fred.cummins@ucd.ie (F. Cummins).

0095-4470/02/$ - see front matter
PII: S0095-4470(02)00082-7

of the bar, e.g. into 4 quarter notes, and based also on a specified tempo, which may vary throughout a piece. We should expect the ensemble player to closely reproduce these nominal durations, and to vary from them only in predictable fashion. For example, a gradual slowing down, or *rallentendo*, at the end of a musical phrase is an established convention which may not be noted explicitly in the score, but which is part and parcel of conventional timing, so that all violinists in the ensemble might be expected to display it.

The soloist, on the other hand, has considerably more freedom to depart from the nominal durations of the score for the purpose of expressive variation (Repp, 1996). We should not expect to find as tight a correspondence between performance and underlying score if measurements were taken from the playing of a soloist (especially in a musical tradition such as romanticism, which demands a maximum of expressive interpretation). If we set ourselves the task of deducing the nominal durations given in an unknown score, we should be better off attending to the less imaginative, but more informative, timings produced by the solid 14th violinist, than trying to disentangle the score and the expression in the performance of a soloist.

The job of deducing nominal durations given in an unknown score is remarkably like the task of the experimental phonetician, endeavouring to identify underlying structure based on noisy performance data. Yet experimental phonetics has hitherto focused on the analysis of the soloist (speakers reading texts alone), rather than that of the ensemble speaker. Of course the analogy is not complete. In particular, we remain agnostic about the nature and existence of any underlying score for speech production, though something similar is posited by virtually every theory of language and speech (and is even sometimes called a 'score', (Browman & Goldstein 1986)). The analogy is sufficiently compelling to warrant further investigation of speech produced in ensemble, or to coin a term, Synchronous Speech.

As the foregoing argument suggests, Synchronous Speech is of interest, not merely because it is a well-defined speaking style (though this alone would merit attention), but because it may prove to become a useful tool for studying phenomena, especially temporal effects, ordinarily tackled by other means. Before that can come about, however, it will be necessary to adequately characterize Synchronous Speech (hereafter SS) and to investigate its properties in some detail.

In an initial set of experiments using only 4 subjects, Cummins (2002) confirmed that subjects could read with a high degree of synchrony, without extensive practice, and with relative ease. In a SS condition, the median lag between co-speakers was 30 ms (upper quartile: 55 ms). This was shorter than that found when subjects tried to synchronize with recordings of other speakers, (median lag: 56 ms, upper quartile: 95 ms), suggesting that SS is produced by an active process of accommodation to the co-speaker, rather than by one speaker imitating or close shadowing another. Another finding was that speakers appeared to agree on pause placement in the SS condition, placing silent pauses only at sentence ends, whereas solo readings elicited much greater variability in pause placement. This suggests, indeed, that expressive variability is reduced in SS, and that speakers are producing something more akin to an expressionless but uncontroversial default when producing SS. A limitation of this study was the small number of subjects (4, yielding 6 pairs). None of these speakers exhibited any difficulty with the task, but with such small numbers, no strong inference about the general population is warranted.

To avoid confusion, it should be pointed out that SS is quite different from Shadowed Speech (Marslen-Wilson, 1973, 1985), in which one speaker attempts to reproduce the speech of another, without prior knowledge of the upcoming utterance. In SS, subjects read through a text first to

become familiar with it, and then read in synchrony while holding the text in front of them. Lags of about 30 ms in SS are much shorter than the shortest lags reported for shadowed speech (ca 200 ms).

In the present study, the approach taken in Cummins (2002) is greatly extended, by recording from 27 subject pairs, using a large range of texts and conditions. Our main goal is to arrive at a sounder understanding of the SS condition by asking the following questions:

1. Is practice at the task necessary or helpful?
2. Is practice at a specific text helpful?
3. Is visual information used in synchronizing with another speaker?
4. Is the resulting speech a collaborative effort in which both speakers accommodate approximately equally, or does one speaker lead and the other follow?

These questions are of immediate theoretical interest. If the initial characterization of SS as a form of collaboration among speakers, in which idiosyncratic variability is reduced as far as possible, is correct, then we need to know whether the collaboration is based on shared knowledge or is instead a skill which is learnt. If the former, then SS may be a means of probing speakers' implicit knowledge of speech production. If the latter, the ability is of much less interest, as the skill may not tell us anything about speech beyond this particular elicitation condition. The first two questions address these issues.

Given that speakers can synchronize rapidly and well, we need to ask what the informational basis for this feat is. Depending on the control theory employed, an account of synchronization might appeal to innate knowledge of durations, closed-loop feedback, entrainment among parameters of dynamic systems, or similar (Kelso & Kay, 1987, Chapter 1; Adams, Weismer, & Kent, 1993; Bingham, 1995). Any such account will need to take into consideration empirical evidence that one or other potential information source is or is not used. While we cannot exclude auditory feedback or innate knowledge, we can control the presence of visual information. If visual information is found to aid synchronization, then an account which relies principally on entrainment among speakers will be favored, and any account will have to include some role for online accommodation based on perceptual feedback.

Finally, the synchronization process might conceivably come about by having one speaker lead and the other follow. Or each speaker might alter their speech to accommodate the other. This question of 'leading' versus 'collaboration' was not addressed in the previous study, but is clearly essential in understanding the nature of synchronization. Different classes of models would suggest themselves, depending on whether a mutual entrainment, or a leading dynamic were operative (Strogatz & Stewart, 1993).

While definitive answers cannot yet be provided to all these questions, the present work goes a long way towards establishing initial answers.

2. Methods

27 subject pairs from the Eastern counties of Ireland were self-selected by public advertisement among a university population. No attempt was made to control for degree of familiarity within

pairs, or for gender, so pairs include both co-habiting couples and absolute strangers. Subjects were paid for a single recording session lasting about 50 min. In this time, a series of texts were recorded. For each text, the following procedure was adopted:

1. Both speakers read the text through silently once.
2. One (randomly selected) speaker read the text aloud, in the presence of the other.
3. Both speakers then attempted to read the text in synchrony, after a verbal signal from the experimenter ("start after 1: 3...2...1..."),
4. Finally, the remaining speaker read the text alone, but in the presence of the other.

In this way order effects were randomized, and both solo and synchronous readings were available for each text.

Recording were made onto the right and left channels of a single stereo file, using head-mounted near-field unidirectional microphones (Shure SM10A). Except where noted below, speakers sat opposite one another, and within sight of each other, about 2 m apart. Each speaker had a sheaf of texts to hold as well as the speaking task to attend to.

The order in which texts were read, and the associated conditions, was as shown in the Table 1. 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.

3. Analysis and results

3.1. Synchrony and phrasal position

Readings from conditions WITH_VIS_1 and WITH_VIS_2 were analyzed, to establish the effects of phrasal position and of practice with the task. In both conditions, subjects had read novel texts while facing one another. The first readings were obtained early in the recording session, the second almost at the end. 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

Table 1
Chronological order in which texts were read, with associated conditions

Condition	Text	Comments
Warmup	Aesop's fable	Data not analyzed. This condition was simply to introduce the task.
WITH_VIS_1	One of A, B, C and D	Subjects could see one another.
NO_VIS_1	One of A, B, C and D	Subjects could not see one another (seated back to back)
...	Word lists, poems	not analyzed herein, but recorded using the same procedures
PRE_PRACT	Rainbow Text, 1st paragraph	Seating, etc., as WITH_VIS_1.
...	Rainbow Text, 1st paragraph	Practice readings, alone and in synchrony.
POST_PRACT	Rainbow Text, 1st 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 (seated back to back)

Each session lasted approximately 50 min.

Table 2
Mean (S.D.) of the per-trial median asynchrony in ms broken down by phrasal position and condition

Condition	Phrase onset	Phrase middle	Phrase end
WITH_VIS1	62 (30)	40 (20)	44 (25)
WITH_VIS2	61 (38)	45 (27)	40 (24)

For each reading, a median value was computed. The table reports the means (standard deviations) of these medians.

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. Trial medians, rather than means were used, as occasional dysfluencies are known to occasionally introduce outliers into the data. Table 2 gives the mean and standard deviations of these per-trial median values.

From previous work, it was to be expected 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, it was predicted that onsets would be less synchronous than either medial or final measurements. Secondly, it was 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 Phrase 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, data in medial and end positions were combined, but position (onset/non-onset) was retained as a factor in our analysis.

3.2. The role of visual information

In all previous experiments on Synchronous Speech, 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, the conditions of the previous experiment were followed 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

Table 3
Mean (S.D.) of the per-trial median asynchrony with and without visual information

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

Times are in ms.

display a strong practice effect, and in fact there was no main effect of practice, nor was there an interaction. Therefore, in what follows data is combined from WITH_VIS_1 and WITH_VIS_2 into one set WITH_VIS, and likewise from NO_VIS_1 and NO_VIS_2 into one set NO_VIS. Mean (S.D.) data summarizing trial medians is given in Table 3.

An ANOVA was conducted 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: $F(1,307) = 49.63$, $p < 0.001$], 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.

3.3. Role of text familiarity

In the first experimental condition no effect whatsoever was found 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, subjects were asked to 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. They were then asked to repeat this sequence 6 more times for practice, and a final set of readings was recorded.

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 Table 4. It can be seen that the improvement in synchrony is modest, being of the order of 10 ms in each case.

3.4. Collaboration and accommodation

In investigating how synchrony is achieved and maintained across speakers, some measure of the degree of accommodation shown by each speaker is necessary. This requires the development of a similarity metric based on temporal structure with which readings by one speaker in the solo and synchronous condition can be compared. The following procedure was adopted, based on pilot studies: One phrase (the final sentence of the first paragraph of the Rainbow Text) was excised from readings in the PRE_PRACT and POST_PRACT conditions. Sixteen well-defined points in the waveform were identified by hand. These points correspond to reliably recognizable

Table 4
Effect of practice with a fixed text, as revealed at phrase-onset and phrase-medial positions

Condition	Phrase onset	Phrase middle/end
PRE PRACT	68 (33)	40 (27)
POST PRACT	60 (34)	27 (14)

Mean (S.D.) of trial medians are given in ms.

events such as stop releases, vowel onsets, etc., and together they divided the utterance into 15 sub-intervals of approximately 2–4 syllables each. Analysis was performed only on utterances for which all 16 points could be reliably identified.

Comparisons between solo and synchronous readings from the same speaker were of particular interest. To this end, pairs of phrases from the same speaker in the two conditions were identified. The Euclidean distance between the two corresponding vectors of 15 interval durations was calculated. For each speaker, the degree of change between solo and synchronous readings was calculated,

$$d_a = \sqrt{\sum_{i=1}^{n}(a_i - A_i)^2}, \tag{1}$$

where a_i refers to an interval duration from the solo reading, and A_i refers to the corresponding interval from the synchronous reading. Distances were only calculated for speaker pairs where each of the 4 phrases (2 speakers, 2 conditions) provided a full set of 16 measurement points. Nineteen speaker pairs provided complete interval measurements.

Fig. 1 plots the value of d for matched speakers. Exactly equal accommodation among speakers would result in equal values of d, and hence data points along the line $y = x$ (dotted line). A linear regression on the actual values obtained provided a slope of 0.6 and an intercept of 0.08, which supports a claim that both speakers are, in general, exhibiting similar degrees of change in response to the demands of speaking synchronously.

In general, speakers will tend to slow down in the synchronous condition. For the present speakers, the duration of the phrase was reliably longer in the synchronous condition compared with the solo condition ($t(37) = -7.62, p < 0.01$). Furthermore, the difference in overall phrase duration was greater between matched pairs of solo recordings than between matched pairs of synchronous recordings (Wilcoxon signed rank test, $V = 159, p < 0.01$).

4. Discussion

The present set of experiments confirms and extends the findings of Cummins (2002) about the effect of speaking synchronously. Speakers can read synchronously with little effort, and there is no advantage of practice unless a specific text is repeatedly read. Synchrony is not as strong at phrase onsets, but rapidly reaches a median value of about 40 ms. The degree of synchrony achieved seems to rule out any simple reactive account, in which one speaker listens and then acts.

The first questions to be addressed here dealt with the effect of practice. Although it was already known that speakers could synchronize effectively with no practice, the finding that considerable

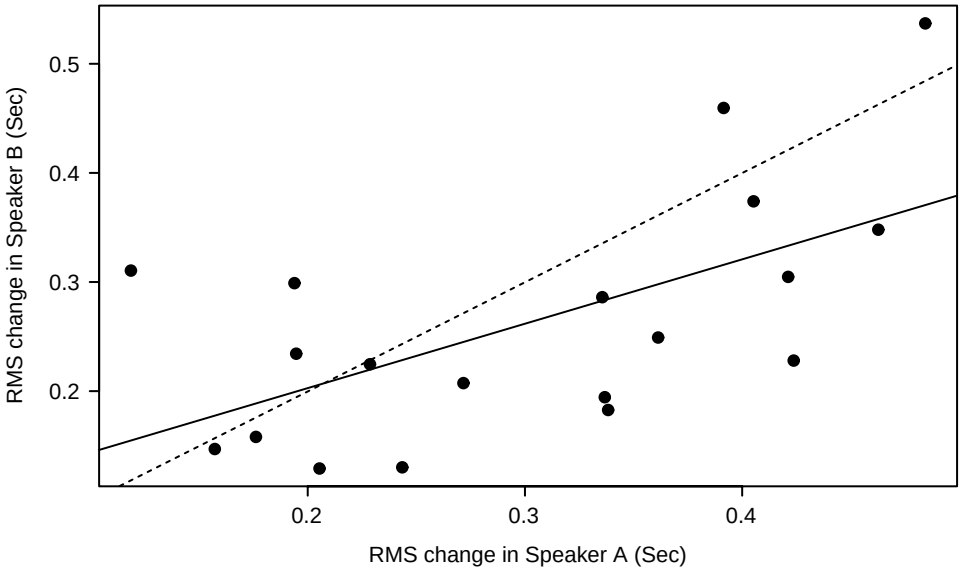

Fig. 1. Degree of change between solo and synchronous readings for matched pairs of speakers. The dotted line is $y = x$. Points on this line would suggest exactly equal amounts of accommodation within a pair, and thus no leading. The solid line is the best linear fit (slope: 0.6, intercept: 0.08, $r^2 = 0.33, p < 0.01$).

practice at the task did not yield improved performance was striking. Indeed, the effect of practice on the same text, and with the same speaker, was only of the order of 10 ms. This strongly supports the claim that the basis for synchronization is shared knowledge of what it is to speak in an unmarked manner, without idiosyncratic or unpredictable variation.

Visual information is of some use in synchronizing with another speaker' although speakers' own reports suggest that they are not aware that they are using vision in accomplishing task goals, and in spite of the fact that speakers are holding a text from which they are reading. The further investigation of the nature of the visual information used will be of considerable interest, but the fact that subjects were simultaneously reading from a sheaf of texts is a strong indicator that peripheral and not foveal vision is being used here.

Finally, the resulting synchronous speech is, indeed, a collaborative effort in which each speaker accommodates to the other. Undoubtedly dyads will show some degree of variability here. A louder speaker may be more effective in altering the speech of a quieter speaker. No attempt was made here to volume-match the speakers. Nor was there any control for speaker familiarity. The informal impression has been gained that speakers who know each other better have an easier time with the task and produce less dysfluencies, but this remains to be tested in a controlled environment.

Already enough is known about Synchronous Speech to warrant its use in phonetics experiments in which intersubject temporal variability is high. For example, ongoing work compares inter-sentential pause duration in synchronous and solo conditions Zvonik & Cummins (2002). Initial results suggest that pauses produced in a SS condition are considerably less variable than those elicited from solo readings. Many other experimental agendas could profit from an experimental tool which constrains variability in a principled manner, by exploiting a speaker's

own implicit knowledge of what is extraneous and what essential in speech timing. It is to be hoped that Synchronous Speech will prove to be a generally useful tool in the study of timing and variability in speech.

Acknowledgements

Work supported by a grant from the Irish Higher Education Authority to the author for collaborative work with Media Lab Europe. My thanks extend to Doug Whalen and to three anonymous reviewers who greatly improved this paper.

Appendix A. Texts used in recordings

The following four texts were used in the experiments. Each text has been divided into major phrases. In each phrase, measurement points are vowel (or sonorant) onsets of the syllables in square brackets; one each at the beginning, middle and end of the phrase. 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 [Swall]ow'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].

References

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research*, *36*, 41–54.

Bingham, G. P. (1995). The role of perception in timing: feedback control in motor programming and task dynamics. In E. Covey, H. Hawkings, T. McMullen, & R. Port (Eds.), *Neural representation of temporal patterns*. New York, NY: Plenum Press.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.

Cummins, F. (2002). On synchronous speech. *Acoustic Research Letters Online*, *3*(1), 7–11.

Kelso, J. A. S., & Kay, B. A. (1987). Information and control: a macroscopic analysis of perception-action coupling. In H. Heuer, & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 3–32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Klatt, D. (1986). Problem of variability in speech recognition and in models of speech perception. In J. Perkell, & D. Klatt (Eds.), *Invariance and variability in the speech processes* (pp. 300–320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.

Marslen-Wilson, W. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*, 55–73.

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America*, *100*(6), 3917–3932.

Strogatz, S. H., & Stewart, I. (1993). Coupled oscillators and biological synchronization. *Scientific American*, 102–109.

Zvonik, E., & Cummins, F. (2002). Pause duration and variability in read texts. In *Proceedings of the International Congress of Spoken Language Processing*, Denver, CO (pp. 1109–1112).