Rhythmic Commonalities between Hand Gestures and Speech

Fred Cummins and Robert F. Port
Departments of Linguistics and Cognitive Science
Indiana University
Bloomington, IN 47405
fcummins@cs.indiana.edu, port@cs.indiana.edu

Abstract

Studies of coordination in rhythmic limb movement have established that certain phase relationships among cycling limbs are preferred, i.e. patterns such as synchrony and anti-synchrony are produced more often and more reliably than arbitrary relations. A speech experiment in which subjects attempt to place a phrase-medial stress at a range of phases within an overall phrase repetition cycle is presented, and analogous results are found. Certain phase relations occur more frequently and exhibit greater stability than others. To a first approximation, these phases are predicted by a simple harmonic model. The observed commonalities between limb movements and spoken rhythm support Lashley's conjecture that a common control strategy underlies the coordination of all rhythmic activity.

Introduction

In his famous paper on the problem of serial order, Lashley (1951) emphasized the importance of rhythmic coordination in all integrated movement, suggesting that speech and other forms of coordinated action must share common organizational principles. There has been a good deal of research into the rhythmic principles that facilitate and constrain coordination among the limbs and hands (Kelso and Scholz, 1985; Kugler et al., 1980; Bernstein, 1967). Recent work has shown that the relative timing of repeated limb movements can be well modeled by low-dimensional oscillator dynamics. However, there has been little effort to link these findings to the production of speech.

One line of research on finger motion finds that when subjects are asked to wag two fingers, or both hands, cyclically toward and away from the body's midline, subjects have a strong preference for a synchronous phase relation between the fingers or hands, where synchrony means that the limbs move toward and away from the midline simultaneously. The anti-synchrony phase relation, where both move left and then both move right, is less stable but is much more stable (small variance, insensitivity to perturbation) than other arbitrary phase angles between the limbs (Kay et al., 1991; Kelso and Kay, 1987). Furthermore, while both synchrony and anti-synchrony are stable at slower tempos, an increase in tempo eventually leads to a control regime in which only synchrony is stable. Study of the stability properties of each production mode and of the phase transition between stable modes suggests the existence of an underlying dynamic which is parameterized by rate. The system exhibits two competing attractors at slower rates, a single attractor at fast rates, and hysteresis observed between the two cases. That is, if a subject tries to maintain anti-synchrony between effectors, increasing tempo will eventually lead to a switch to synchrony, and on tempo reduction, synchrony will be maintained beyond the tempo at which the switch previously occurred.

In describing rhythmic coordination in this paper, we will mark the timing of events relative to an overall cycle using phase, with a range of 0 to 1. By arbitrarily taking one of the effectors in the above studies as defining the cycle, the stable patterns observed have relative phases of 0 (synchrony) and 0.5 (anti-synchrony) between the hands.

In the literature on the performance of rhythmic patterns, it is well established that subjects perceive and produce patterns in which the intervals are related as simple integers (1:1, 2:1, 3:1 etc.) with much greater facility than patterns in which the component intervals have arbitrary or complex ratios (e.g. 2.72:1). Fraisse (1982) gives an overview of older work, while Collier and Wright (1995) give a more recent summary. This is true, whether subjects spontaneously tap out groups of, say, 2 to 4 elements, in which case the intergroup intervals tend to relate to the intragroup intervals as 2:1 (Essens and Povel, 1985; Fraisse, 1956) or whether they try to reproduce specific interval ratios (Collier and Wright, 1995; Summers et al., 1989; Tuller and Kelso, 1989), where the intervals produced gravitate towards simple harmonic ratios. Expressed using our phase convention, and taking the largest repeating unit as the cycle, subjects in both these cases are showing strong preferences for events at phases of $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$ etc.

The investigation of global speech rhythm has had much less tangible results. Pike (1945) was the first to classify languages as being stress-timed or syllable-timed. Although some data has supported versions of the syllable-timing hypothesis for languages like Japanese (Port et al., 1996; Port et al., 1987), attempts

to find evidence for isochrony of stressed syllables for English have been largely fruitless (Dauer, 1983; Lehiste, 1977). A weak tendency towards equally spaced stresses is documented for some English speech (Jassem et al., 1984), but these data do not satisfactorily account for speaker/listener impressions of rhythmicity (Dauer, 1983; Lehiste, 1977). To look just for isochrony, however, is analogous to looking only for 1:1 interval ratios in spontaneous tapping. Little attention has been paid to interstress intervals which might be related as 2:1, 1:2 or other simple ratios. Both Jones (1960, 1st ed. 1918) and Martin (1972) make an explicit connection between the rhythms of music and speech by applying standard Western musical rhythm notation to English phrases. They thus imply that interstress intervals show hierarchical organization, with long and short intervals in speech organized into temporal structures based on harmonic fractions. While neither Jones nor Martin specifically suggest that this hierarchical organization can be identified directly from the speech signal itself, this is certainly implied in these approaches.

In music, events are notated at harmonic phases of the measure cycle. If the description of speech as having a somewhat music-like rhythm is accurate, there should be some events which occur at simple phases within an overall cycle. That is, global speech timing should be harmonically or rhythmically constrained, and should exhibit simple interval ratios, much as are found in the literature on the manual production of rhythmic patterns. We conducted an experiment to see if stress placement within a repeated phrase might act like taps in a repeated tapping task by exhibiting a bias toward the occurrence of stressed syllable onsets at harmonic fractions (eg, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$) of the period of a repeated text. In order to parallel experiments on the manual production of rhythm (Collier and Wright, 1995; Summers et al., 1989, etc.), subjects were instructed to place a phrase-medial stress at specified phase angles.

Stress placement: An experiment

Methods

In our experiment 6 subjects were asked to repeat the simple phrase "Take a pack of cards." The period from the onset of "take" to the next onset of "take" defined the basic phase cycle from 0 to 1. This interval was fixed at 1.5 sec. Auditorily, the subjects were presented with just the words "take" and "cards" with "cards" located at one of 8 phase angles between .3 and .65 relative to the basic cycle.

In a single trial, subjects were asked to continually repeat the phrase "Take a pack of cards" and to align the words "take" and "cards" with those of the stimulus signal, which provided a target phase for the placement of the phrase-medial stress. Each trial contained three sets of phrase repetitions. After speaking along with the stimulus 7 times, the stimulus was turned off, and subjects continued speaking, attempting to maintain the target timing pattern. After another 7 repetitions, they paused for 3 seconds and then performed a third set of 7 repetitions. All 8 target phase angles were tested within a block of 8 trials. There were three blocks, within which the target phase given by the stimulus was either increased from trial to trial within the block, or was decreased, or the order was randomized.

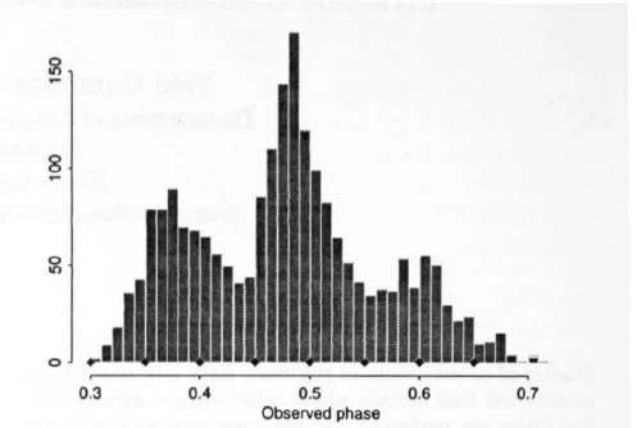

Figure 1: Histogram of all observed phases of the medial stressed syllable "cards" in repetitions of the phrase "Take a pack of cards." Target phases are marked in black on the abscissa. Although there were 8 target phases, the overall distribution is clearly trimodal.

The time of onset of the initial and medial stressed syllables was measured automatically from audio recordings by an onset detector that picks out an increase in the smoothed signal energy envelope, restricted to a frequency range of about 300–2000 Hz. This locates a "beat" very close to the vowel onset of each syllable, and is thus similar to algorithms for locating "P-centers" (Scott, 1993; Marcus, 1981).

Results

The main results pooled across speakers, repetition sets and trial blocks are shown in Figure 1 as a frequency histogram of the measured phase angle of "cards." The first main finding was that subjects could not produce all target phases equally well. Although the target phase angles for the medial stress were equally probable at 8 different phase angles, the produced phase angles exhibit a strong preference for phases close to 0.5, and somewhat weaker preferences for phases near 0.36 and 0.6. These values are close to $\frac{1}{3}$ and $\frac{2}{3}$, predicted by a simple harmonic model for stress location (although the consistent deviation away from actual harmonic predictions merits further attention). Contrary to our expectations, subjects' attempts to reproduce the target phase were no more accurate when speaking simultaneously with the

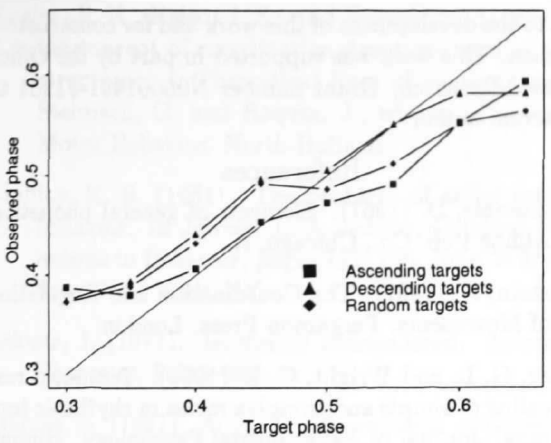

Figure 2: Observed phase of the vowel onset for "cards" as a function of target phase when target phases are increasing, decreasing and randomized across trials. The line $y = x$ is also included for reference. Hysteresis is apparent in that a sequence of trials in which phase increases from trial to trial produces smaller mean phase values than a sequence with decreasing phase.

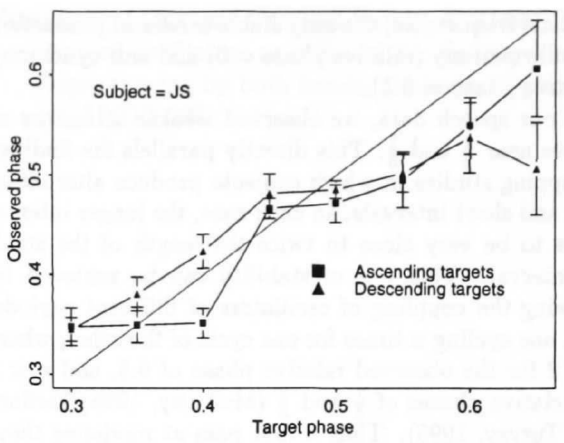

Figure 3: Sample data for one subject showing the mean and standard deviation for each trial. The two conditions plotted are the phase increasing block and the phase decreasing block.

prompt than they were immediately after its cessation or after the 3 second delay.

A second important result is that the produced phase was influenced by that produced on immediately preceding trials, as shown in Figure 2. For example, the target phases of 0.4 and 0.45 tend to be produced at values close to 0.5 when performed immediately after targets at 0.5 (that is, on descending) yet they tend to stay close to 0.333 when performed after targets at 0.333 (in the ascending condition). Random target ordering yields intermediate values of mean phase. The effect of recent targets on the distribution of observed phases is important information for inferring the symmetry properties of the underlying dynamic.

The data shown so far is averaged across subjects, which obscures considerable intersubject variability. The bias for harmonic fractions can be seen more clearly when individual speakers are examined. By way of example, Figure 3 shows some results for a single subject. Each data point shows the mean observed phase and standard deviation for repetitions within a single trial. As in the previous figure, trials are grouped by block (targets are either ascending or descending across trials). First, essentially the same patterns as in the pooled data are evident: for most target values, the ascending function lies to the right of the descending function, i.e. smaller mean phases are produced. A clear example of bias for targets at harmonic fractions can be seen in the ascending curve for targets at 0.3, 0.35 and 0.4. All were imitated with the same output phase at about 0.35. Then targets of 0.45, 0.5 and 0.55 were all produced very close to 0.5. As the target changed from 0.4 to 0.45, the subject's productions jumped discretely from $\frac{1}{3}$ to $\frac{1}{2}$. It appears that the actual phases produced by subjects are dependent, not only on the combination of target phase and nearby harmonic attractors, but they are also influenced by context (i.e. productions and targets on previous trials).

Discussion: Speech as harmonically timed

In tapping and limb movement studies, the dependent variable is usually the phase of one limb with respect to a cycle defined by the other. In this study we measured the phase of a phrase-medial stressed syllable with respect to a phase cycle defined by the repetition of the phrase as a whole. Our hypothesis was that harmonic fractions of the larger cycle should serve as attractors for the timing of stressed syllables just as simple harmonic fractions (often expressed as integral interval ratios such as 2:1, 3:1 etc.) are found to be attractors in studies of manually produced rhythmic patterns. The data provide strong support for this.

Our core finding is that subjects produced certain phases much more often than others although target phases are equally distributed. In the finger wagging studies of Kelso, Kay and colleagues, only relative phases of zero and 0.5 were observed to be stable (Kay et al., 1991). We, too, found a phase of 0.5 to be strongly preferred by subjects. Kelso and others have modeled the dynamics of this behavior with oscillatory systems described by second-order differential equations. Yamanishi et al (1980) suggested using two coupled oscillators of identical periods to model the control of two fingers in a bimanual tapping task. Kay et al (1991) likewise consider coupled oscillators as a possible underlying control mechanism. In each case, as the oscillators beat at

identical frequencies, the only stable modes of production are at synchrony (relative phase = 0) and anti-synchrony (relative phase = 0.5).

In our speech data, we observed weaker attractors at phases near $\frac{1}{3}$ and $\frac{2}{3}$. This directly parallels the findings of tapping studies in which subjects produce alternating long and short intervals. In each case, the longer interval tends to be very close to twice the length of the smaller interval. This sort of stability can be achieved by allowing the coupling of oscillators of different periods, with one cycling n times for one cycle of the other, where $n = 2$ for the observed relative phase of 0.5, and $n = 3$ for relative phases of $\frac{1}{3}$ and $\frac{2}{3}$ (McAuley, 1995; Treffner and Turvey, 1993). Thus a first pass at modeling these data will refer to two endogenous oscillators of different periods. The slower oscillator is identified with the timing of the repeated phrase as a whole, while the faster corresponds to the control of the metrical foot.

Unpublished data from pilot studies we have conducted suggest that in the absence of a target temporal pattern, subjects will exhibit a very strong tendency to produce approximately harmonic phases. The strong preference for a phase of 0.5 may well figure in a full account of impressions of isochrony. The evidence for "silent beats," in which one inter-stress interval is observed to be twice as long as neighboring intervals, also suggests that harmonic phases are preferred and produced by speakers (Abercrombie, 1967). In the present study, we set task demands which are at odds with these intrinsic tendencies by asking the subjects to produce patterns in which the two oscillators could no longer couple. Their tendency to gravitate towards simple integer ratios was, however, clearly evident in the resulting data.

Conclusions

Together, these results support Lashley's conjecture that there is a common control strategy underlying the global coordination of speech as well other coordinated rhythmic activity. They suggest that 'isochronous timing' for stressed syllables is only one of many temporal relationships that can be supported by a rhythm-based model for speech timing. They also suggest the potential utility for speech researchers of a variety of research methods originally developed for work on limbs. Of course, there remain many issues which need attention, for example, the influence of phonetic content on the observed phases. The stability properties of the stable modes in speech production likewise remain to be investigated. Notwithstanding, the present study appears to be the first to offer direct phonetic timing data in support of music-like rhythms in human speech, and establishes clear links between rhythm in speech and limb movements.

Acknowledgments

We are indebted to Keiichi Tajima, Michael Gasser, Richard Shiffrin and Wendy Goldberg for their contributions to the development of this work and for constructive criticism. This work was supported in part by the Office of Naval Research, Grant number N00001491-J1261 to the second author.

References

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

Bernstein, N. (1967). The Coordination and Regulation of Movements. Pergamon Press, London.

Collier, G. L. and Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. Journal of Experimental Psychology: Human Perception and Performance, 21(3):602–627.

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

Essens, P. J. and Povel, D. (1985). Metrical and nonmetrical representations of temporal patterns. Perception and Psychophysics, 37(1):1–7.

Fraisse, P. (1956). Les Structures Rhythmique. Érasme, Paris.

Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D., editor, The Psychology of Music, pages 149–180. Academic Press, New York.

Jassem, W., Hill, D., and Witten, I. (1984). Isochrony in English speech: its statistical validity and linguistic relevance. In Gibbon, D. and Richter, H., editors, Intonation, Accent and Rhythm, volume 8 of Research in Text Theory, pages 203–225. Walter de Gruyter, Berlin.

Jones, D. (1960). An Outline of English Phonetics. Cambridge University Press, Cambridge, England, 9th edition. 1st edition published 1918.

Kay, B., Saltzman, E., and Kelso, J. A. S. (1991). Steady-state and perturbed rhythmical movements: Dynamical modeling using a variety of analytical tools. Journal of Experimental Psychology: Human Perception and Performance, 17:183–197.

Kelso, J. and Scholz, J. (1985). Cooperative phenomena in biological motion. In Haken, H., editor, Complex Systems: Operational Approaches in Neurobiology, Physics and Computers, pages 124–149. Springer Verlag.

Kelso, J. A. S. and Kay, B. A. (1987). Information and control: a macroscopic analysis of perception-action coupling. In Heuer, H. and Sanders, A. F., editors, Perspectives on Perception and Action, chapter 1, pages 3–32. Lawrence Erlbaum Associates, Hillsdale, NJ.

Kugler, P. N., Kelso, J. S., and Turvey, M. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In Stelmach, G. and Requin, J., editors, Tutorials in Motor Behavior. North-Holland.

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.

Marcus, S. (1981). Acoustic determinants of perceptual center (P-center) location. Perception and Psychophysics, 30:247–256.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. Psychological Review, 79(6):487–509.

McAuley, J. D. (1995). On the Perception of Time as Phase: Toward an Adaptive-Oscillator Model of Rhythm. PhD thesis, Indiana University, Bloomington, IN. Available as Cognitive Science Technical Report No. 151, Cognitive Science Program, Indiana University, Bloomington, IN.

Pike, K. (1945). The Intonation of American English. University of Michigan Press, Ann Arbor, MI.

Port, R., Cummins, F., and Gasser, M. (1996). A dynamic approach to rhythm in language: Toward a temporal phonology. In Luka, B. and Need, B., editors, Proceedings of the Chicago Linguistics Society, pages 375–397. Department of Linguistics, University of Chicago.

Port, R. F., Dalby, J., and O'Dell, M. (1987). Evidence for mora timing in Japanese. Journal of the Acoustical Society of America, 81(5):1574–1585.

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

Summers, J. J., Bell, R., and Burns, B. D. (1989). Perceptual and motor factors in the imitation of simple temporal patterns. Psychological Research, 50:23–27.

Treffner, P. J. and Turvey, M. T. (1993). Resonance constraints on rhythmic movement. Journal of Experimental Psychology: Human Perception and Performance, 19(6):1221–1237.

Tuller, B. and Kelso, J. (1989). Environmentally-specified patterns of movement coordination in normal and split-brain subjects. Experimental Brain Research, 75:306–316.

Yamanishi, J., Kawato, M., and Suzuki, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. Biological Cybernetics, 37:219–225.