Implications of rhythmic discreteness in speech

Robert F. Port

Linguistics and Cognitive Science, Indiana University, U.S.A.

port@indiana.edu

Are there temporal constraints on speech timing? Some have argued against them (Chomsky and Halle, 1968; van Santen, 1996) while others have claimed, for example, that languages come in two possible rhythm types, stressed-timed and syllable-timed (Abercrombie, 1967; Pike, 1943). We prefer to avoid for the moment, the question of a typology of rhythm types. Instead we will focus on styles of speech that are explicitly rhythmical. In recent studies we have verified what laymen already understand—that human speech is easily and naturally spoken in a rhythmical way. But hearing a rhythmic speaking style and demonstrating this objectively are quite different things. The empirical verification we offer for rhythmic structure in speech is based on the realization that vowel onsets (i.e., approximately P-centers) are the most important event determining perceived speech rhythm (Allen, 1972; Morton et al., 1976). That is, when speaking rhythmically, English speakers (and very likely speakers of other languages too) adjust overall timing so that vowel onsets occur near certain privileged temporal locations. This is important for research since it means that if we measure vowel onset locations, we do not need to pay much attention to other aspects of phonetic events in order to characterize speech rhythm.

Rhythmic effects should be of interest to students of language development since in some form it is exhibited quite early. Various kinds of rhythm are found in children's speech from well before first-words. Examples include the reduplication of syllables in babbling and the observation that children differentiate the prosody of their mother's language from other languages shortly after birth. The cognitive skills I explore in this paper are ones that children probably acquire very early in language acquisition. We have assumed that adults already have significant experience and skill—even if they may be largely unaware of their own skills in this regard.

We have found (Cummins and Port, 1998; Tajima and Port, in press) that, if asked to repeat a short phrase like *Buy the boy a cake* once per second or so, there are three temporal patterns that English speakers find simple and natural. Treating the interval between each *Buy* as the basic time unit, the vowel onset of *cake* most often divides that interval in half. At a faster rate speakers may use waltz-time reading where *buy, boy* and *cake* are on the three beats. The third (slowest) timing pattern also divides the unit interval into three, but *cake* occurs on the second beat and there is a musical rest on the third. Experiments show that these preferences or constraints are powerful. Speakers find it very difficult to locate onsets at arbitrary positions in the cycle, but can produce only these 3 patterns easily, accurately and with no training. What mechanism could account for these constraints?

Following reasoning resembling that of Haken, Kelso and Bunz (1986; Kelso, 1995), we propose that there are attractors at the preferred phase angles. These attractor locations may be caused by coupled oscillators that are phase locked in harmonically related frequencies. Thus the waltz meter might reflect a system of two oscillators in the frequency-ratio of 1:3 that are phase locked so that certain of their pulses occur synchronously. Then the locations of pulses of the two oscillators act as attractors, drawing either the perception of acoustic onsets or the motor gestures of speech production, toward the pulse. This would account for the bias for vowel onsets to occur only near harmonic fractions of the cycle.

In a more recent experiment (Port et al., 2002) we tested two predictions: that, first, if there is a pulse at 2/3 of the way through a cycle (in synchrony with a tone onset), there should also be a similar pulse at 1/3. This
follows from the notion that an oscillator cannot simply skip a beat; it goes through phase zero on every cycle. Second, it was predicted that a meter defined by two oscillators (e.g., at frequency ratio 1:3 or 1:2) would yield ‘stronger’ attractors than a meter specified by three oscillators (e.g., at frequency ratios 1:2:4). These predictions were supported by data based on 10 speakers with enough repetitions to provide over 3000 measurements per condition. These results further support the hypothesis that the metrical patterns of speech are maintained by oscillatory metrical structures.

These results have important consequences for our understanding of speech.

1. Methods using rhythmically spoken speech are easy to design and are probably easily adapted to young speakers. Study of the developmental origin of these skills will not be difficult.

2. This periodic linguistic behavior displays some properties of speech that allow comparison and interaction with other motor systems, like those involving fingers, feet and jaw, as well as with perceptual meter systems (see McAuley and Kidd, 1998; Large and Jones, 1999).

3. These powerful temporal constraints—e.g., harmonically related periods and phase locking at 0 phase—show that there are global timing constraints on speech that can be observed under these conditions at least. Other experiments (Tajima and Port, in press) show that these constraints differ from language to language, as Abercrombie and Pike might have predicted.

4. Not everything in the speech signal is equally relevant to these temporal patterns: it’s the acoustic onsets that have the most relevance for rhythmic patterns.

5. These constraints seem to demand interpretation as resulting from oscillators and structures of coupled oscillators. The metrical systems underlying most music seem to be essentially similar. Thus it turns out that language and music are intimately related in ways not traditionally suspected. Speech can easily be warped into some form of chant or music.

6. More generally, these results reaffirm that language is an embodied, physical system that is not isolated in a Platonic space of pure ‘competence’ where serial order is the only kind of time possible.