Title
Language, Audition and Rhythm

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 18(0)

Author
Port, Robert F.

Publication Date
1996

Peer reviewed

Language, Audition and Rhythm

Robert F. Port
Departments of Linguistics, Computer Science, Cognitive Science
Indiana University
Bloomington, IN 47401
port@indiana.edu

Oscillators and entrainment are just as much a part of the Basic Auditory Toolkit of cognition as they are of the Basic Motor Control Toolkit. The kinds of rhythm that are found in human languages may reflect either audition or motor control, or both. And it may be hard to tell which is the primary source. Language is both motor and perceptual, yet, fundamentally, neither. Language is an abstract, medium-independent structure – a system that is shared to varying degrees by members of a speech community. All of this is what is meant by saying that language is a cognitive system (van Gelder and Port, 1995).

What factors constrain the phonetic or phonological form of words and phrases in languages? The standard answer is that the physiology of the speech organs provides constraints (e.g., Hockett, 1955; Chomsky and Halle, 1968). For example, linguists usually describe speech in articulatory terms: 'labials', 'stops', 'high vowels.' Even prosodic patterns (over a longer time scale) may be described as in such articulatory terms as 'breath group'. But traditional articulatory descriptions, such as those of linguists, have typically been completely vague about what, say, a 'labial' gesture really is. Recent developments in the theory of motor control have improved the situation by showing how dynamical systems can model speech gestures as objects that are still flexible and abstract (Kelso et al., 1986; Saltzman, 1995). So articulation still provides the most widespread vocabulary for describing speech. We observe that when producing speech the lips jaw, tongue and larynx move quasi-periodically and exhibit strongly stereotyped behavior. For each language, we find that speakers master quite different yet apparently difficult articulatory skills – as evidenced by the persistence of foreign accent when adults learn a new language.

However, independent of the motor constraints, the perceptual system must also master difficult skills when learning a language. It may take years to learn language to asymptote, and, for an adult, some aspects of the perception (and production) of a second language may be effectively unlearnable (Logan, Lively, and Pisoni, 1991). So what is hearing that it could be so difficult?

But the problem of hearing must set in very general terms. What does the auditory system learn? It learns sound patterns. But what are the requirements for being a learnable pattern? Not every complex sound is an auditory pattern. Patterns must have particular kinds of symmetry - such as being a spatiotemporal event that recurs often in the environment. The frequency space can be modelled as linear summation of independent frequencies. But what kind of structures over time can their be? We think there are 3 basic types: slopes, periods and sequences. One limit is that patterns that are periodic at time scales between about .2 Hz ($T = 5$ sec) and 10 Hz ($T = .1$ sec) can be directly predicted. Some non-speech events exhibit oscillations in this frequency range: cricket chirp patterns, water waves, tree branches and leaves in the wind, animal gaits from mice to elephants, resonance of air-filled tubes (from the size of pen cap to a 10-ft tree trunk).

If perceptual systems are to recognize such time-distributed auditory objects, then they must predict them with oscillatory mechanisms (McAuley, 1995). Thus it is likely that oscillators and entrainment are essential components of the Basic Auditory Toolkit – the inventory of mechanisms that are exploited in learning auditory patterns that occur in time.

Language is permeated with rhythm - both because it involves motor control AND because it involves perceptual processes (Port, Cummins and Gasser, 1996). Both get lots of benefits from oscillation. So it may be difficult to be sure just why speech exhibits any rhythm one observes. One source of complexity in the problem is that languages do not just differ from each other in their rhythmic structure (e.g., Japanese mora timing versus English so-called 'stressed-timing'). Even WITHIN languages there is always great variety in rhythmic style. For example, compare baby-talk with the speech of macho athletes, or prose pronunciation with the sing-song recitation of a limerick; or the speech of an emotional preacher pounding his fist on the Bible. A wide range of speech rhythms are available to skilled speakers of any language.

In fact, in my lab we are studying what speakers do when asked simply to repeat a phrase over and over to a slow metronome that signals when they should start each repetition (Cummins, 1995; Cummins and Port, 1996). Surprisingly, a variety of unexpected constraints on the location of stressed syllable onsets could be observed such as that a stressed syllable within the phrase must begin at a simple harmonic fraction of the whole sentence repetition (that is, at 1/3 or 1/2 way through the cycle of the whole phrase).

This easily encouraged behavior of subjects reinforces our sense that *speech is always on the edge of rhythmicity*. It is certainly not highly periodic all the time, but it seems that speech is very susceptible to entrainment by any periodic source that happens to be available – by the beat of a drum, by a melody, by the legs of talking jogger, and by the click of our lab metronome. We do not yet know how to design models that exhibit this kind nearly periodic behavior in a useful way, but it seems important to work on.

Rhythm is widespread in nature. Most often, it results from mechanical constraints (e.g., of wave reflection in a uniform medium like a string or a bar or tube, etc) that result in harmonically related frequencies tending to co-occur. It may be that ultimately, this results in rhythmic, harmonic entrainment being one of the basic auditory mechanisms. The rhythm of speech probably results from some combination of mechanical constraints on the dynamics of the vocal tract combined with perceptual constraints on the oscillatory system that recognizes spoken language.

References

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper Row, New York.

Cummins, F. (1995). Identification of rhythmic forms of speech production. *Journal of the Acoustical Society of America*, 98:2894.

Cummins, F. and Port, R. (1996). Rhythmic constraints on English stress timing. In *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, PA.

Hockett, C. (1955). *The Manual of Phonology*. Waverly Press, Baltimore, Maryland.

Kelso, J. A., Saltzman, E., and Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics*, 14:29–60.

Logan, J., Lively, S., and Pisoni, D. (1991). Training japanese listeners to identify English /r/ and /l/. *Journal of the Acoustical Society of America*, 89:874–886.

McAuley, J. D. (1995). Perception of Time as Phase: Toward an Adaptive-Oscillator Model of Rhythmic pattern processing. Ph.D. Thesis, Indiana University.

Port, R., Cummins, F., and Gasser, M. (1996a). A dynamic approach to rhythm in language: Toward a temporal phonology. In Luka, B. and Needs, B., editors, *Proceedings of the Chicago Linguistics Society 1995*, pages 375–397. Department of Linguistics, University of Chicago.

Port, R., McAuley, J. D., and Anderson, S. (1996b). Toward simulated audition in an open environment. In Covey, E., Hawkins, H., and Port, R., editors, *Neural Representation of Temporal Patterns*, pages 77–106. Plenum, New York.

van Gelder, T. and Port, R. (1995). It's about time: Overview of the dynamical approach to cognition. In Port, R. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition*, pages 1–43. Bradford Books/MIT Press.