15

On the value of reductionism and formal explicitness in phonological models: comments on Ohala's paper

JANET B. PIERREHUMBERT

In this paper, Ohala provides a nice case study in the style that he has become so well known for. He presents experimental data indicating that a regular phonological process, the direction of assimilation, is grounded in facts about speech production and perception. Such results are significant not only because of the light they shed on particular phenomena, but also as examples of research methodology. They point out the importance of seeking the right sphere of explanation for observed patterns in sound structure.

The type of explanation which is featured in this paper is phonetic explanation, or explanation based on the physics and physiology of speech. Phonetic explanations are especially attractive because of their reductionist character; it is very satisfying to reduce psychology to biology, and biology to physics. Ohala's comparison of phonetic and nonlinear phonological accounts of assimilation links reductionism ("None of the terms of the explanation are unfamiliar, other-worldly entities") with generality ("A few primitives go a long way"). It is not clear to me that this link is well-founded, especially with respect to ongoing research. Nonlinear phonology has identified a number of principles which have great generality, although their physical basis is unclear. In particular, the principle of hierarchical organization has been shown to be a factor in the lexical inventory, phrasal intonation, and allophony rules of many languages. On the other hand, some parts of phonetics are extremely particular, from a scientific point of view. For example, there is no reason to suppose that the specific nonlinear oscillator responsible for vocal fold vibration has any generality from the point of view of physics. At this point, physics has no general theory of nonlinear oscillators; each new system that has been studied has given rise to new analysis techniques and interpretations. Vocal fold vibration is more interesting than other systems of similar mathematical complexity chiefly because of its role in human language. A researcher who is trying to decide how to use his time often has a choice of whether to aim for reduction or for generality. In some cases, paradoxically, aiming for generality is best even if reduction is the aim; making the description more comprehensive and exact can narrow the class of possible underlying mechanisms. The field advances best if the judgement is made on a case-by-case basis, by assessing the feasibility and informativeness of the various methods that might be applied.

I think we need to consider, also, the possibility that higher-level domain-specific theories may incorporate scientific insights which are lost (to the human mind, at least) in the theories which "explain" them. I am reminded of a talk I heard a few years ago in which Julian Schwinger discussed his experiences designing microwave guides during the war. He at first viewed this assignment as a trivial and uninteresting one, since the physics of microwave guides is completely specified by Maxwell's equations. However, although the behaviour of microwave guides is indeed a solution to Maxwell's equations, their phenomenology proved to be so complex that an additional higher-level theory was needed to make it comprehensible. The interest of the assignment emerged in constructing this theory. In this case, the higher-level theory was desirable even though the explanation was already known. Such theories are doubly desirable when the explanation is being sought.

In his summary comparison of phonetic and nonlinear phonological representations of sound patterns, Ohala says that neither claims to represent sound structure in a computationally efficient manner. However, computational models, whether efficient or not have been very important in our progress towards explaining speech. So a brief review may be worthwhile.

On the phonetic side, the acoustic theory of speech production relied on calculations made using the first available digital computers. It demonstrated an approximation to speech production which can be computed efficiently, something recently emphasized in Fant (1985). This enabled it to support work on synthesis models, which have been so important to our understanding of speech perception, prosody, and allophony in continuous speech. It has also provided the basis for rigorous work on the limitations of the theory (cf. Fant, Lin, and Gobl 1985).

On the other hand, the formalism for phonological rules developed in Chomsky and Halle (1968) was grounded in earlier work by Chomsky and others on the theory of computation. The *Sound Pattern of English* (SPE) showed considerable descriptive flair, but so did much earlier work. It advanced over earlier work chiefly by its algorithmic approach, and the SPE formalism was successfully applied in implementing the phonological rules of text-to-speech systems, both for English and for other languages (see Allen, Hunnicutt and Klatt 1987; Carlson and Granström 1976; Carlson and Granström 1986, and references given there). Work in metrical and autosegmental phonology has built on the theory of trees and connected graphs, which are also computationally tractable. This tractability has made it possible to move quickly to implementations which incorporate theoretical advances in nonlinear phonology, for example Church's (1982) syllable parser and work by Pierrehumbert (1979), Anderson, Pierrehumbert and Liberman (1984) and Beckman and Pierrehumbert (1986) on synthesis of fundamental frequency contours. Such implementations of nonlinear representations have themselves led both to new observations and to theoretical innovations (see Pierrehumbert and Beckman 1988). Nonlinear phonology still presents one serious obstacle to computational implementations; it is not explicit about the interaction between derivational rules and well-formedness conditions. Compared to *The Sound Pattern of English*, the theory remains underformulated. This has been an obstacle to theoretical progress, too, since it has led to confusion about what claims are being made and what their consequences are for new data.

One lesson which emerges from reviewing computational work on sound structure is the value of formalization. Unlike Ohala, I feel that formalization pays off for all types of descriptions. Formalizing nonlinear phonological descriptions pays off because it clarifies the issues and assists systematic evaluation, in part by supporting the construction of computer programs. Some descriptions may be just as good as others, but not all are equally good; some are just plain wrong. There is no point in seeking a phonetic or cognitive basis for spurious generalizations. It is important to keep in mind, also the importance of formalizing phonetic descriptions. If we observe a parallel between some facts about speech and some physical law, we don't have an explanation, we have a conjecture. To have an explanation, it is necessary to write down the equations and determine the quantitative correspondence to the data. Only in this way we find out if additional physical or cognitive mechanisms are crucially involved. Exact modeling will be especially important for determining the interplay of phonetic and cognitive factors, since the cognitive system can apparently exaggerate and extend tendencies which arise in the phonetics.

can

A second lesson is the value of distinguishing levels of representation. For instance, one level in a synthesis system will have the job of specifying what linguistic contrasts in sound structure are possible. Another will specify how sounds are pronounced, in terms of the time course of acoustic or articulatory parameters. Yet another will specify a speech waveform. For the most part, different kinds of work are done at different levels, and so they are complementary rather than competing. However, competition does arise when the division of labor between levels is unclear. This is how most of the competition between nonlinear phonology and phonetics arises, in my opinion. For example, before Poser (1984), it was unclear whether Japanese had a phonological rule changing High to Mid after a pitch accent, or whether it had a phonetic rule reducing the pitch range after a pitch accent. Poser's experiments resolved this question. I feel optimistic that such issues will in general be resolvable by empirical investigation, and will not become mired in debates about philosophy and taste.

A third observation, suggested especially by work on speech synthesis, is that different degrees of explanation are possible at all levels of representation. Nonlinear phonology explains the sound patterns in the English lexicon better

[illegible]

than a word list does. This nonlinear explanation in turn requires further explanation; since (as Ohala points out) it is based on internal reconstruction, its cognitive basis is problematic and needs to be worked out. A similar gradation can be found within phonetics proper. The acoustic theory of speech production explains spectra of speech sounds by deriving them from the configurations of the articulators. But it too requires further explanation. Why does the linear approximation work as well as it does? Which articulatory configurations are possible in general? And why does one configuration rather than another occur in any particular case?

References

Allen, J., M. Sharon Hunnicutt, and D. Klatt. 1987. *From test to speech: The MITalk System*. Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.

Anderson, M., J. Pierrehumbert, and M. Y. Liberman. 1984. Synthesis by rule of English intonation patterns. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* I: 2.8.1–2.8.4.

Beckman, M. and J. Pierrehumbert. 1986. Japanese prosodic phrasing and intonation synthesis. *Proceedings of the 24th Meeting of the Association for Computational Linguistics. Association for Computational Linguistics*, 136–144.

Carlson, R. and B. Granström, 1976. A text-to-speech system based entirely on rules. Conference Record, 1976 IEEE International Conference on Acoustics, Speech, and Signal Processing 686–688.

Carlson, R. and B. Granström. 1986. Linguistic processing in the KTH multi-lingual text-to-speech system. Proceedings of the 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing 2403–2406.

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

Church, K. 1982. Phrase-structure parsing: a method for taking advantage of allophonic constraints Ph.D. dissertation, MIT.

Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton (2nd edition 1970).

1985. The vocal tract in your pocket calculator. In V. Fromkin (ed.) *Phonetic Linguistics*. London: Academic Press. 55–77.

Fant, G., Q. Lin, and C. Gobl. 1985. Notes on glottal flow interaction. *STL-QPSR* 2-3/85, Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, 21–41.

Hertz, S. 1982. From text to speech with SRS. *Journal of the Accoustical Society of America* 72: 1155–1170.

Pierrehumbert, J. 1979. Intonation Synthesis Based on Metrical Grids. *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, Acoustical Society of America, New York, 523–526.

Pierrehumbert, J. and M. Beckman. 1988. *Japanese Tone Structure*. Linguistic Inquiry Monograph Series. Cambridge: MIT Press.

Poser, W. 1984. The Phonetics and Phonology of Tone and Intonation~~al~~ in Japanese. Ph.D dissertation. MIT.