The Phonetics-Phonology Interface

John J. Ohala
University of California, Berkeley
E-mail: ohala@socrates.berkeley.edu

ABSTRACT

To bridge the gap between the dissimilar mechanisms and entities of the phonological grammar and those of the phonetic mechanisms for producing and perceiving speech, the 'phonetics-phonology interface' (PPI) has been invoked. In this brief presentation, an introduction to the ICPhS 2003 Symposium on The Phonetics-Phonology Interface, I present an exegesis and taxonomy of the PPI and some citations to relevant literature.

1. INTRODUCTION

The expression 'phonetics-phonology interface' (PPI) is subject to about as many imaginative interpretations as there are variants of the (in)famous "light bulb" jokes [7]. Even though there are only three words in the expression, there are more than 3^3 interpretations and opinions about its meaning, its reality, its implications for theoretical and practical treatments of speech. In this short paper, a preface to the symposium of the same name, I will explore some of these interpretations.

2. EXEGESIS

The first thing to establish is the domain in which the PPI is supposed to operate. Among the candidates are the conditions under which researchers attempt to discover how things work in the phonetic and phonological mechanisms. Although one may certainly encounter cases where phonological speculation is carried out with little attention to or awareness of findings in phonetics[1] as well as cases where phonetic research seems to ignore the phonological function of speech sounds, it can be agreed, presumably, that the practice of 'autonomous phonology' and 'autonomous phonetics' is to be avoided. [10]. Ideally, whether trying to understand sound patterns in language or how speech production and speech perception work, both domains (and others) must be integrated. However, the usual domain in which the PPI is invoked is in accounting for how language users mediate between the mechanisms and entities in their phonological grammar and those evident in the more directly observable phonetic realm, whether it is in the act of speaking or perceiving that is involved. In this context we can ask what the term interface might imply.

An ***interface*** is the site where or the mechanism by which two largely independent mechanisms meet, interact, exchange substances or information. Familiar uses include, e.g., a computer-printer interface, or the interface between a living cell and its environment. The term implies two fundamentally autonomous mechanisms that either function in a different way or produce different kinds of information. If the term 'interface' hadn't been chosen, other acceptable terms might have been 'translator' or 'go-between'. Is there a need for translation between the actions, symbols, or entities that processed in phonology and phonetics? There are a priori reasons to think that some translation or interface is necessary. The principal reasons given is the discreteness of the elements in a phonological representation but the continuous nature of the speech. Another reason that may be adduced is that the phonological representation, indeed, all stored knowledge, is in some sense latent (like potential energy) whereas at least one stage of speech production and perception is dynamic and has a physical embodiment (like kinetic energy). None of these translation problems, of course, is unique to speech. They arise when almost any motor skill is considered: typing, playing musical instruments, dancing, etc. But there is an added reason for positing a translation process: the units and representations at the phonological and phonetic level are *so different* (even aside from the discrete vs. continuous dichotomy). Phonological features,

[1] I differentiate between what I call *taxonomic* phonetics and *scientific* phonetics. Taxonomic phonetics has given us the descriptive labels for differentiating speech sounds. It has remained largely unchanged since the late 19th century when Sweet, Passy, Jespersen, Viëtor, and others established it. Scientific phonetics has a quite ancient history, going back at least to Galen, the 2nd c. anatomist, to the phonetic studies of the Arabs and Persians in the 7th to the 11th centuries, and up to modern phonetic studies. Unlike taxonomic phonetics, scientific phonetics continues to change, as all sciences do, by incorporating new data, methods, and theories. It gives outdated theories a decent burial. It is scientific phonetics that has often received too little attention by phonologists.

ISBN 1-876346-48-5

for example, under most conceptions of them, are viewed as occupying time slots but are themselves timeless or at least ambiguous as to how they are to be played out in real time. Most phonological representations are ambiguous as to how features interact with each other. Certain proposed phonological features are difficult to interpret in phonetic terms or have little or no empirical support, e.g., [tense], [heightened subglottal pressure]. The phonetic interpretation of segments phonologically underspecified for a given feature still occupies phoneticians.

On the other side, to make distinctions to describe regular patterns in speech and especially to characterize the difference in speech patterns between one language another phoneticians need a finer set of distinctions than phonologists are prepared to acknowledge. In general, phonologists have sought the simplest or minimal set of phonological specifications for speech sounds.

It is from such apparent mismatches that most of the work regarding the PPI stems, in many cases without explicitly invoking that label and, in fact, much of it predating the coinage of that term.

3. THREE APPROACHES TO THE *PPI*

Abandon it. One suggestion is that there will be no way to derive phonetic events from phonological representations. Ladefoged [8] concluded "[d]escriptions of phonological patterns in languages involve features that are not in a one-to-one relationship with … phonetic parameters … There is no necessity for most phonological features to be part of mental representations."

Get on with the job. Much energy has been directed at taking phonological representations of various sorts and developing algorithms that convert them to phonetic parameters. A huge literature of this sort exists, e.g., [4][5]. In the same vein, there have been many algorithms developed that produce synthetic speech from an input that is some sort of phonological representation [9] [14].

In contrast to the success of speech synthesis by rule from phonological input, automatic speech recognition is mostly done by retrieving whole words or phrases, although there are some exceptions [3].

Reduce the Mismatch at the *PPI*. Many other approaches take the position that the conception of either phonological representations or the phonetic representations are wrong and that a correct reformulation of one or the other will ameliorate the mismatch.

Articulatory phonology [1] holds that the phonological units of speech are discrete gestures roughly corresponding to consonants and vowels. These gestures are inherently dynamic and much of what is regarded as 'phonology' (assimilation, epenthesis, syncope, etc.) comes about due to variations in the execution and phasing of these gestures. With a similar aim, there has been an abundance of efforts to show that apparent phonetic complexity can be greatly simplified—and thus reduce its mismatch with phonological simplicity [2] [3].

Addressing the assumption that phonological and phonetic entities differ in the discrete vs. continuous dichotomy, Ohala [10] pointed out that there are obvious discontinuities in the speech in the acoustic (and aerodynamic) domains, manifested as rapid modulations in speech amplitude, periodicity, spectrum, and F0. He suggested that at an intermediate stage between phonemes and the level encoded by speakers and decoded by listeners, another unit, the diphone, might have some validity. This idea suggests that the mismatch between phonetic complexity and phonological simplicity can be ameliorated by an intermediate level of complexity: the inventory of diphones in a language would be on the order of N^2, where N is the number of phonemes.

Another approach to reduce the dissimilarities between phonetic and phonological representations is to re-examine phonological representations. From one point of view phonological representations are simple in their inventory of elements but complex in the kind of processing that that entail, whether this is accomplished by rules or by constraints. Some researchers have questioned the fundamental assumptions of mainstream phonology.

Ohala [12] suggests that much of the mechanisms of mainstream phonology, from markedness to feature geometry, is designed to reflect the naturalness and therefore the generality and simplicity of rules of grammar. But, he argues, true generality in the statement of sound patterns in language can only be obtained from the application of principles and constraints that are clearly extra-grammatical, i.e., physical and language-independent psychological constraints. These latter include same the pressure-flow constraints that hold in automobile carburetors and the same psychological principles that make visual camouflage so effective in the animal world. Removing the burden of naturalness from psychological grammars could make them simpler and much closer to the phonetic structure of speech. Similar views are expressed by Harris and colleagues [6].

ISBN 1-876346-48-5

4. Conclusion

Even if we succeed in reducing the dissimilarities between what have been conceived to be the operation of the phonological grammar and phonetic mechanisms there will continue to be a problem of bridging the gap between the latent and the dynamic, between the psychological and the physical. But we can better solve these problems if we have an realistic idea of what it is that is on either side of the interface.

REFERENCES

[1] C. Browman and L. Goldstein. "Articulatory phonology: An overview." *Phonetica,* vol. 49, pp. 155-180. 1992.

[2] C.A. Fowler. "An event approach to the study of speech perception from a direct-realist perspective." J. Phonetics vol. 14, pp. 3-28. 1986.

[3] O. Fujimura. "C/D model: A computational model of phonetic implementation." In *Language Computations*, E. S. Ristad, Ed., Providence, RI: Am. Math. Soc., pp. 1-20, 1994.

[4] C.-W Kim. The linguistic specification of speech, PhD Diss., UCLA. 1966.

[5] J. Pierrehumbert and M. Beckman. *Japanese tone structure.* Cambridge, MA: MIT Press. 1988.

[6] J. Harris and G. Lindsey. "Vowel patterns in mind and sound." In *Phonological knowledge: conceptual and empirical issues,* N. Burton-Roberts, P. Carr and G. Docherty, Eds. Oxford University Press, pp. 185-205. 2000.

[7] http://www.eyrie.org/~thad/strange/lightbulbs.htm.

[8] P. Ladefoged. "What are linguistic sounds made of?", *Language*, vol. 56. pp. 485-502. 1980.

[9] I. Mattingly. "Synthesis by rule of prosodic features." *Language and Speech*, vol. 9, pp. 1-13. 1966.

[10] J. J. Ohala. "The integration of phonetics and phonology," *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence, 19-24 Aug 1991.* vol. 1, pp. 1-16. 1991.

[11] J. J. Ohala. "The segment: primitive or derived?" In *Papers in Laboratory Phonology II: Gestures, segments, prosody.,* G. J. Docherty and D. R. Ladd, Eds. Cambridge, UK: Cambridge University Press. pp. 166-183. 1992.

[12] J. J. Ohala. "Phonetic explanations for sound patterns: implications for grammars of competence," In *Proc. 13th Int. Congr. Phonetic Sciences, Stockholm, 13-19 August 1995.* K. Elenius & P. Branderud Eds., vol. 2, pp. 52-59. 1995.

[13] H. Reetz. "Converting speech signals to phonological features". *The XIVth International Congress of Phonetic Sciences,* San Francisco. vol. 3, pp. 1733-1736. 1999.

[14] N.Umeda and R. Teranishi. "The parsing program for automatic text-to-speech synthesis developed at the Electro-technical Laboratory in 1968," IEEE Trans. A, vol. 23, 183-188. 1968.

ISBN 1-876346-48-5

ISBN 1-876346-48-5