On a Supposed Dogma of Speech Perception Research: A Response to Appelbaum (1999)

Fernando Orphão de Carvalho
Universidade de Brasília

Abstract. In this paper we purport to qualify the claim, advanced by Appelbaum (1999) that speech perception research, in the last 70 years or so, has endorsed a view on the nature of speech for which no evidence can be adduced and which has resisted falsification through active ad hoc "theoretical repair" carried by speech scientists. We show that the author's qualms on the putative dogmatic status of speech research are utterly unwarranted, if not misconstrued as a whole. On more general grounds, the present article can be understood as a work on the rather underdeveloped area of the philosophy and history of Linguistics.

Keywords: Linguistics, epistemology, phoneme, speech perception.

1. Introduction

Appelbaum (1999) has identified what she claims to be a "dogma" underlying speech perception research. The so-called dogma seems to be a theoretical proposition granted special status by speech researchers, in the sense of not being amenable to empirical falsification or being subject to *ad hoc* "repair operations" to salvage its status in the face of apparently disconfirming evidence. The dogma is labeled "the dogma of isomorphism" by the author and it states that speech is organized, at some level, in the manner of strings that result of the linear concatenation of more elementary, discrete and invariant units. According to it, the units of speech, those marks or properties of the speech signal which allow for the identification of phonological segments, are organized as alphabetic symbols in strings: discrete, invariant units successively concatenated.

As a general conclusion to the paper, the author contends that "after half a century without empirical support, the persistence of the alphabetic conception of speech must be seen as a dogma, one which speech scientists would do well to give up" (S258). As I find this conclusion startling, it came as a surprise to learn that no rebuttal to this piece of reasoning has been published. The present paper aims at filling this gap.

2. The Stated Case for the Dogma of Isomorphism

In her statement on the nature of the dogma of isomorphism, Appelbaum chooses to approach the problem historically. She first presents what, she claims, was the 'stan-

Principia **13**(1): 93–103 (2009).
Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

dard' or 'received view' conception of linguists on the nature of the speech signal and its component parts during pre-World War II days. Around 1945, when many technological off-shots from the military efforts of the previous years started to become commercially or openly available, one particular gadget attracted the attention of linguists and speech scientists: the sound spectrograph (Joos 1948). The revolutionary aspect introduced by the use of this device is, according to the author, quite straightforward: It allowed researchers to literally 'see' what speech is like, simply by analyzing the devices' output or 'spectrograms'. Spectrograms displayed speech as variations in amplitude-by-frequency as a function of time.

The central accusation is built around this historical turn. As the author puts it, the advent of spectrograms laid in front of everyone's noses the "raw" and "concrete" evidence that speech is not organized in the manner people thought so, that is, there wasn't any longer room for an 'alphabetic conception' of speech: the acoustic segments or stretches that roughly correspond to phonological/phonetic segments were neither invariant nor strictly ordered, they varied as a function of other neighboring segments and cues to one particular segment could be found almost anywhere in a given utterance.

The dogmatic status of the 'alphabetic conception' of speech is so characterized: instead of changing their minds in face of this new body of evidence, linguists and speech scientists employed a seemingly *ad hoc* strategy to salvage their conception on how speech is organized. According to the author, researchers adopted the strategy of assigning this alphabet-like organization not to the acoustic stuff of speech, but rather to some other level of description. Cutting a few corners, what they held was that, whatever the nature of the speech signal itself, the "message" that gets encoded in the speech signal *is* organized according to the alphabetic principle. The message is organized in terms of the code (i.e., the particular language) shared by speaker and hearer. The former encodes it in acoustic form and the latter decodes it from the signal.

Appelbaum contends that this 'retreat strategy' is completely unjustified, being motivated by the researchers' implicit goal of keeping with the somehow comfortable assumption that the speech signal is organized in the manner of strings of letters. We now proceed to a detailed account of the flaws in Appelbaum's case for the dogma of isomorphism.

3. Criticism

3.1. The Received View and the History of Linguistics

According to the author, the so-called 'received view' on the structure of speech is "the assumption that phonetic segments — the individual consonant and vowel

sounds of a language - were embedded in the speech signal as discrete, interchangeable, serially arranged acoustic segments" (S251). To back up her statement, the author cites the famous assertion by Leonard Bloomfield that speech sounds organize themselves "like beads on a string". To begin with, it is necessary to point out a conceptual flaw in this presentation. The author seems to cluster two independent claims: the first one is that "speech sounds" are invariant and interchangeable while the second one asserts that these sound units are serially organized. Clearly, speech sounds could be context-dependent in their form but serially organized (or at least, deemed to be so). Alternatively, the units of speech could be invariant but "clustered" on top of one another to varying degrees (i.e., temporally simultaneous, as the pitch and timbre of an uttered vowel). This seemingly pedantic distinction actually helps us in cutting some ice.

An added problem refers the author's sloppy handling of technical terminology. At different times in the article the author refers to both 'phonemes' (S253) and 'phonetic segments' (S251) in a seemingly interchangeable manner. So, in the same page (S251) she offers two definitions of the alphabetic conception of speech: the one cited in the previous paragraph (where *phonetic segments* are organized "like beads on a string") and the other one: "the view that *phonemes* are discrete, linearly ordered segments which are neither multiply-realizable nor context-dependent" (italics are mine). At page S253, the author briefly describes the status of the historical turning point marked by the advent of spectrographic representations of speech and states that: "Before, it was assumed that different tokens of the same phonetic segment were interchangeable; afterwards it was acknowledged that the speech signal typically does not contain segments corresponding to ... commutable phonemes". The first is a statement about *tokens* of speech sounds, the other one about phonemes, which are, as we see below, more like *types*. This is deeply regretful since, especially to the discussion in question, the distinction between phonetic and phonological objects is crucial.

As it happens, linguists at Bloomfield's time, and specially those within the 'american structuralist' tradition, were well aware of the fact that speech sounds aren't acoustically invariant. The recognition of this phenomenon of invariance was central to the linguistic definition of "linguistically significant speech sound" or 'phoneme'. Any trained phonetician can easily spot the operation of systematic distortions that a given phoneme suffers when acoustically realized. As an example we could cite the textbook example that /g/ is pronounced with a slightly more anterior vocal tract constriction when close to anterior vowels (as in *geek*) but in a more posterior region of the vocal tract with posterior vowels (such as in *goose*). So, a phoneme was already recognized as something akin to a perceptual equivalence class of different *tokens* or 'allophones'. Leonard Bloomfield in his 1933 classic *Language* (the same work cited by Appelbaum) contends that:

> Non-distinctive features occur in all manners of distributions. In most types of American English, the t-phoneme in words like 'water' and 'butter' is often reduced to an instantaneous touch of the tongue-tip against the ridge behind the upper gums: in our habit, the sound produced suffices to represent the phoneme. (p. 81)

In the same vein but in a more dramatic rendering, Edward Sapir remarked in 1925 that:

> (...) what is felt by the speaker to be the "same" sound has perceptibly different forms as the [phonetic] conditions vary. (...) it is very necessary to understand that it is not because the objective difference is too slight to be readily perceptible that such variations stand outside of the proper phonetic pattern of the language (...) but that the objective difference is felt to be slight precisely because it corresponds to nothing significant in the inner structure of the phonetic pattern. (Sapir 1985, 37)

So, there are no grounds to believe that "before [the introduction of spectrographs] it was assumed that phonemes could be specified in context-independent acoustic terms" (S253).

We should approach the issue of serial organization with the difference between 'phonemes' and 'phonetic units' or 'allophones' right in hand. This conceptual distinction is akin to that between 'perceptual qualities or entities' and 'physical entities or stimulus properties' familiar from psychophysics. Phonemes are perceptual entities, part of the knowledge that an agent has of his language and is thus indispensable in explaining his linguistic behavior. When a linguist talks about phonetic primes, units or processes he usually means those "raw" acoustic and articulatory phenomena whereby phonemes are 'encoded' in a physical signal for communication (and the resulting effects of this encoding process) and those features that 'stand for' the phoneme in a speech signal or help in its identification. One is not reducible to the other, no more than perceptual entities such as 'pitch', 'intensity', 'timbre' or 'color' are reducible to physical objects such as fundamental frequencies of vibration or particular wavelengths. This point was forcefully exposed by Sapir in his 1925 paper sampled above.

Now, it is clear that the claim concerning the serial organization of speech primes was postulated to hold *at the level of phonemes* and not at the level of the acoustic signal. Another influential XX century linguist, Charles Hockett, defines the 'linearity assumption' as the claim that

> the distinctive sound-units or phonemes of a language are building-blocks which occur in a row, never one on top of another or overlapping. (Hockett 1947, 258)

Indeed, in the next period of the same paragraph, we learn from Hockett that, as long as other types of phonological features are concerned

> this assumption has been lifted (...) features of stress or tone, for example, which normally stretch over more than a single vowel or consonant, have been called 'non-linear' (...) in contrast to the linear vowels and consonants. (258)

So, not only does the hypothesis that segments are serially organized applies at the *phonological* level and not at the level of the acoustic description of the speech signal but, and this is a big 'but' in the face of claims of 'dogmatism', this hypothesis has been put into question whenever incoming evidence seems to call for it.

From the previous paragraphs it is clear, thus, that the picture painted by Appelbaum on the beliefs of linguists and speech scientists before the advent of spectrographic representations of speech is deeply flawed. Although not directly touched on in our discussion, the lack of invariance in the acoustic-articulatory implementation of phonemes was already understood by linguists such as Paul Passy, Henry Sweet and Baudoin de Courtenay back in the XIX century, the latter being credited as the actual inventor of the term *phoneme*. We based most of our exposition on American structuralism since it seems true that more than any other "school of thought", these researchers have stressed in rather explicit terms the importance of these phenomena. Of equal relevance to a truer historical picture of this stage in the early history of modern Linguistics, it is worthwhile to point out that the revisions and amendments to this conception on how phonological primes (phonemes or features) are organized, which were proposed at different times and places throughout the last century were, in almost all cases, couched and based on standard linguistic evidence and on the development of more inclusive theories on the nature of human language out of which the burgeoning amount of data from particular languages could be understood. This is true of the work of Charles Hockett, Kenneth Pike, J. R. Firth and the "London School" as well as traditional generative phonology and its "non-linear" descendants (cf. Goldsmith 1990).[1]

From all we know about acoustic phonetics and the nature of phonetic representations, Appelbaum's claims on the damaging consequences of the obvious absence of 'acoustic segments' corresponding to phonemes in spectrographic representations also seems to be at least an overstatement. Although no biunique correspondence can be observed, some generalizations can be made: It is well known that 'acoustic segments' identified on the basis of abrupt changes in some basic parameters describing the speech signal often outnumber the set of phonemes composing any given utterance (cf. Fant 1973). The reasons for this state of affairs are known in most cases: the acoustic theory of speech production allows one to predict, for example that formant transitions in CV utterances, as well as release burst spectra may provide information on stop place of articulation, and even that voicing spectra in the closure interval for voiced stops may also work that way (Barry 1984).

Also, acoustic representations are currently seen, on more general grounds, as

'weakly segmental' (cf. Stevens 1989, Pierrehumbert 1990, Ohala 1992) this meaning that the acoustic cues on which the identification of a given phonological unit are based tend to co-occur close to each other, rather than at arbitrary distances.[2] It has been proposed that these cues on which identifications or decision on phoneme identity rely tend to cluster near abrupt changes in some dimensions describing the speech signal, such as those resulting from changes in type of source (e.g., from glottal periodic to supra-glotal non-periodic or transient; Stevens 1989). This may be, in turn, an adaptive feature of the speech production mechanism, enforcing signal shapes in which the task recovering segmental information from the continuous speech signal is made easier (cf., e.g., Sussman *et al.* 1998 for some research on perception-production links in speech).

3.2. Evidence here, evidence there

Now, notwithstanding all the comments of the previous section, Appelbaum's case is still quite awkward. The author is basically claiming that since no evidence for phonemes and its properties could be found when scientists turned their attention to the physical nitty-gritty of the speech signal, then there is no evidence for the postulation of the phonemes as part of the way speech perception is organized. As I see it, this claim is on equal foot to some other possible, if somewhat unlike, claims such as: "look, retinal images are 2D, therefore, there's no hard evidence for your claims that there's such a thing as 3D perception". Or switching to my favorite modality: "this speech wave was high-pass filtered. There's no energy at the fundamental *ergo* you are not experiencing the pitch sensation you claim to experience".

That this parallel is sound can be seen from what Appelbaum tells us at page S256: "the assumption that speech was an acoustic alphabetic sequence was initially an empirical hypothesis (...) But the claim that the sequence of alphabetic segments existed at some other level of the speech chain was — and this is the point-remains, a purely stipulative claim". I see two basic problems with the ideas underlying this statement.

As psychophysics and more recent computational branches of psychology show us, perception is pretty much about 'finding one's way in a noisy and uncertain world' (Bennett *et al.* 2002, Hoffman 2000). Upon receiving a signal, a physical stimulus varying in magnitude as a function of time and according to the spatial distribution of their sensory organs, perceptual agents can, as a matter of course, "find" structure in the signal where actually there's none or completely "miss" some huge physical modulation of the signal which happens not to be important. In this latter category we have the case of phonemes: despite all the variations in the physical realizations of a phoneme (its many 'allophones') speakers ignore this variation and retrieve the relevant information, the identity of the phonological segment. Perceptual illusions

are the most salient examples of the former category of signal processing (Geisler & Kersten 2002). In fact, and as a matter of historical fairness, these points were made by traditional linguists well before the advent or widespread use of fancy Bayesian formulations of inference under uncertainty or issues of computational complexity in cognition. Palmer (1936, 82) for example, tells us that:

> Speech is no more than a series of rough hints, which the hearer must interpret in order to arrive at the meaning which the speaker wishes to convey.

So, the bottom line is: when talking about perception we are generally talking about two theoretically and evidentially distinct domains: the perceptual (often denoted in psychophysics as the ψ domain) and the physical (or φ). If there's independent, reliable evidence that some perceptual construct does exist (say, behavioral evidence) no amount of empirical hand-weaving of signal properties will discredit the independently established psychological or perceptual entities. We are obviously interested in understanding as well as possible how the physical structure of stimuli is organized and how it relates to the informational *milieu* surrounding perceptual agents (e.g., Gibson's *affordances*) but only to the extent that this informs us on how these agents use this information to make a living or accomplish tasks. Taking a bit further this ecological or functionalist view and depicting organisms as accomplishing certain tasks (such as deciding on which of many possible phonemes is present given some sampled stretch of a speech signal) we see how Appelbaum seems to confound the *proximate* object of perception (the acoustic signal reaching one's ears) and the *distal* object of perception (a phoneme or a 3D reconstruction of a retinal image). This is clear when the author discusses the supposed "shift" theories of speech perception have experienced after the advent of spectrograms (S254–S255) and claims that researcher have changed the "locus of speech" from the acoustic signal to other levels such as "invariant neural structures" or "articulatory structures":

> (...) this change in what is going on at the acoustic level does not correspond to a change in the underlying conception of speech, because the acoustic signal is no longer being treated as the locus of speech. That is the objects of speech perception are no longer taken to be acoustic in nature. (S254)

One could legitimately ask then: what kind of object are you talking about? There is not such a thing as an exclusive place called "the locus of speech". As we showed in section 3.1., the acoustic signal was taken as the *proximal* and not the *distal* object of perception well before the spectrogram. Already in 1934, we have the following statement by American linguist Morris Swadesh:

> The phonemes of a language are, in a sense, percepts to the native speaker of the given language, who ordinarily hears speech in terms of these percepts. (Swadesh 1934, 117)

Serial organization and categorical structure are properties of the distal objects of speech perception. Showing that the proximal object has such and such properties fulfils the need to understand how is it that one goes from 'here' to 'there', and not whether 'there' exists or not. Put another way, describing physical properties of signals is a way to state in full detail the problems our theories of information processing are asked to solve and not to evaluate the other part of the problem (i.e., the nature of the *output* of this information processing).

Within this perspective, we are able to discern Appelbaum's last mistake: the lack of evidence claim. As it seems, the entire history of modern linguistics, which has successfully worked out the grammars of hundreds of languages, describing and explaining a vast array of grammatical processes and patterns of historical change, not to mention the multidisciplinary realms of language acquisition (Werker & Tees 1984), processing and even its neural basis (Poeppel *et al.* 1997); for more recent work done under the same assumptions, cf. Phillips 2001, Kuhl 2004, all of which employ the concept of serially organized phonemes conceptualized as equivalence classes of context-dependent allophones, produces, after all, no evidential support to the theoretical construct *phoneme*.

Noam Chomsky has called attention to this strange reasoning whereby people rate some "types of evidence" as more reliable or more important than others on an absolute basis. When discussing the positions of philosopher Willard Quine on matters of language and human psychology, he describes Quine's double standard:

> If the evidence derives from psycholinguistic experiments on perceived displacement of clicks, then it counts; if the evidence derives from conditions on referential dependence in Japanese or on the formation of causative constructions in numerous languages, then it does not count — tough this is evidence interpreted in the normal manner of the natural sciences. (Chomsky 2001, 55)

This last clause almost sums up the case for now: *phonemes* much like other theoretical entities in science (pick up as you wish: atoms, genes, double-helixes, curved spaces, chemical bonds) are justified for their explanatory role in theories intended to hold of a certain set of phenomena[3] and not for their relation to extraneous or fashionable standards on quality of evidence.

4. Conclusion

From the previous discussion we conclude that the case for there being a dogma in speech perception research is not warranted, at least not in the form exposed by Appelbaum (1999). Besides poor historiography and a seemingly confuse understanding of what perceptual problems are, we have identified in the paper subject to

criticism an utterly arbitrary disregard for evidence from Linguistics, evidence which bears crucially on the phonological structuring of language. This latter flaw stems in all certainty from a naïve belief on the "hardness" of some bodies of evidence as opposed to others.[4]

References

Appelbaum, I. 1999. The Dogma of Isomorphism: A Case Study from Speech Perception. *Philosophy of Science* **66**: S250–59.

Barry, W. 1984. Place-of-articulation Information in the Closure Voicing of Plosives. *Journal of the Acoustical Society of America* **76**(4): 1245–47.

Bennett, B. M.; Hoffman, D. D.; Prakash, C. 2002. Perception and Evolution. In Heyer, D. & Mausfeld, R. (eds.) *Perception and the Physical World.* Chichester: John Wiley & Sons, 229–46.

Bloomfield, L. 1933. *Language.* Chicago: University of Chicago Press.

Broselow, E. 1996. Skeletal Positions and Moras. In Goldsmith, J. *Handbook of Phonological Theory.* Oxford: Blackwell Publishers, 175–205.

Chomsky, N. 2001. Language and Interpretation: Philosophical Reflections and Empirical Inquiry. In Chomsky, N. *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press, 46–74.

Fant, G. 1973. *Speech Sounds and Features.* Cambridge, MA: MIT Press.

Geisler, W. & Kersten, D. 2002. Illusions, Perception and Bayes. *Nature Neuroscience* **5**(6): 508–10.

Goldsmith, J. 1990. *Autosegmental and Metrical Phonology.* Oxford: Blackwell Publishers.

Hockett, C. 1947. Componential Analysis of Sierra Popoluca. *International Journal of American Linguistics* **13**(4): 258–67.

Hoffman, D. 2000. *Visual Intelligence: How we create what we see.* Oxford: W. W. Norton & Co.

Joos, M. 1948. Acoustic Phonetics. *Language* **24**: 1–136.

Kuhl, P. 2004. Early Language Acquisition: Cracking the Speech Code. *Nature Reviews Neuroscience* **5**: 831–43.

Ohala, J. 1992. The Segment: Primitive or Derived? In Docherty, G. J. & Ladd, D. R. (eds.) *Papers in Laboratory Phonoloy II: Gesture, segment, prosody.* Cambridge: Cambridge University Press, 166–83.

Palmer, L. R. 1936. *Introduction to Modern Linguistics.* London: Macmillan.

Phillips, C. 2001. Levels of Representation and the Electrophysiology of Speech Perception. *Cognitive Science* **25**: 711–31.

Pierrehumbert, J. 1990. Phonological and Phonetic Representation. *Journal of Phonetics* **18**: 375–94.

Poeppel, D. 1997. Mind Over Chatter: Review of T. Deacon, The Symbolic Species: The Co-evolution of Language and the Brain. *Nature* **388**: 734.

Poeppel, D. *et al.* 1997. Processing of Vowels in Supratemporal Auditory Cortex. *Neuroscience Letters* **211**: 145–8.

Robins, R. H. 1997. *A Short History of Linguistics.* London: Longman.

Sapir E. 1985/1925. Sound Patterns in Language. In: Mandelbaum, D. (ed.) *Selected Writings of Edward Sapir.* Berkeley: University of California Press.
Stevens, K. 1989. On the Quantal Nature of Speech. *Journal of Phonetics* **17**: 3-45.
Sussman, H.; Fruchter, D.; Hilbert, J.; Sirosh, J. 1998. Linear Correlates in the Speech Signal: The Orderly Output Constraint. *Behavioral and Brain Sciences* **21**(2): 241–99.
Swadesh, M. 1934. The Phonemic Principle. *Language* **10**: 117–29.
Werker, J. & Tees, R. 1984) Cross-Language Speech Perception — Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development* **7**: 49–63.

FERNANDO ORPHÃO DE CARVALHO
Programa de Pós-Graduação em Lingüística (PPGL)
e
Laboratório de Línguas Indígenas (LaLI)
Universidade de Brasília/UnB
Campus Darcy Ribeiro
Brasília, Brazil
fernaoorphao@gmail.com

Resumo. Neste trabalho temos como objetivo criticar o trabalho de Appelbaum (1999) no qual a autora argumenta a favor da posição de que os pesquisadores da área de Percepção de Fala têm defendido uma hipótese acerca da natureza da fala para a qual não há justificativa empírica, mas que tem sido mantida através de "reparos teóricos" efetuados pelos pesquisadores. Nós demonstramos que as considerações da autora acerca do status putativamente dogmático da pesquisa em Percepção de Fala são injustificados. Em termos mais gerais, o presente trabalho pode ser compreendido como um esforço particular de elucidação epistemológica básica dos fundamentos da pesquisa sobre a linguagem.

Palavras-chave: Linguística, epistemologia, fonema, percepção de fala.

Notes

[1] The lack of decisive influence of instrumental phonetics on these proposals is not without parallel in the history of linguistics. From the second half of the XIX century on, Linguistics ('structural linguistics') has developed itself with little concern for much of instrumental phonetics and similar studies on speech (cf. Robins 1997).
[2] An anonymous reviewer pointed out that contemporary phonologists mostly work with the concept of 'feature' instead of the notion of 'phoneme'. This being true, it is also true that in all theoretical models employing independent ('autosegmental') features for the purposes of phonological description and explanation some notion of a segment with the discrete status of the phoneme is given formal representation and somehow maintained in the theoretical machinery (as CV-tier slots, timing tier slots, place nodes or something else; cf. e.g. Goldsmith 1990; Broselow 1996). Also, in several instances both in synchronic pattern and diachronic reorganization of phonological systems, their action in relation to independent features can be seen (compensatory lengthening; anchoring of floating features interacting

with feature co-occurrence constraints, etc). As it seems, the most revolutionary aspect of modern feature theory has been the realization that features aren't unordered with these timing slots or segments, as it was the case with the phoneme matrixes or features bundles of classical SPE-like phonological theory (cf. Goldsmith 1990).

[3] David Poeppel also puts it nicely when commenting on the general skepticism concerning evidence adduced from the study of the structure of languages: "One of the troubling things about studying language is that almost anyone is willing to contribute an opinion on how language works, including scientists from other fields who should know better. Language is viewed as an area that licenses unconstrained speculation. Despite all lay intuitions, however, the scientific study of language requires technical expertise." (Poeppel 1997)

[4] I gratefully acknowledge the previous comments made on this paper by professor Paulo Abrantes of the Philosophy Department at the University of Brasilia (UnB) and by an anonymous reviewer. All remaining shortcomings or faults are mine.