ECOLOGICAL PSYCHOLOGY
2016, VOL. 28, NO. 4, 183–187
http://dx.doi.org/10.1080/10407413.2016.1230367

Routledge
Taylor & Francis Group

Direct Perceptions of Carol Fowler's Theoretical Perspective

D. H. Whalen [a,b,c]

[a]Program in Speech-Language-Hearing Sciences, City University of New York; [b]Haskins Laboratories; [c]Department of Linguistics, Yale University

ABSTRACT

Carol Fowler has had a tremendous impact on the field of speech perception, in part by having people disagree with her. The disagreements arise, as they often do, from 2 incompatible sources: her positions are often misunderstood and thus "disagreed" with only on the surface, and her positions are rejected because they challenge deeply held, intuitively appealing positions without being shown to be wrong. The misunderstandings center largely on the assertion that perception is "direct." This is often taken to mean that we have access to the speaker's vocal tract by some means other than the (largely acoustic) speech signal, when, in fact, it asserts that the signal is sufficient to directly specify that production. It is unclear why this misunderstanding persists; although there are still issues to be resolved in this regard, the stance is clear. The challenge to "acoustic" theories of speech perception remains, and thus direct perception is still controversial, as it seems that acoustic theories are held by a majority of researchers. Decades' worth of evidence showing the lack of usefulness of purely acoustic properties and the coherence gained by a production perspective have not changed this situation. Some attempts at combining the 2 perspectives have emerged, but they largely miss the Gibsonian challenge that Fowler has espoused: perception of speech is direct. It looks as though it will take some further decades of research and discussion to fully explore her position.

The means by which we recognize speech are complex enough that we are still arguing about fundamentals many decades after serious research on the topic began. The issues are quite fundamental, and the theoretical stance one takes depends in large part on the aspects of the biological worlds one is willing to come to grips with. Carol Fowler has been a consistent and articulate advocate for one approach, Direct Realism (DR), that places speech perception into an overarching Gibsonian theory of perception (e.g., Gibson, 1966). The position has been elaborated in an influential series of articles over the years (e.g., Fowler, 1990, 2010; Fowler, Rubin, Remez, & Turvey, 1980; Fowler & Smith, 1986). Her work has been stimulating but often in the sense of eliciting criticism rather than in convincing skeptics. Here I argue that her work is usually criticized for things it does not claim or treated as if it were not possible to claim what she claims. In the end, she may not be right, but she has often

CONTACT D. H. Whalen dwhalen@gc.cuny.edu Program in Speech-Language-Hearing Sciences, City University of New York, 365 Fifth Avenue, New York, NY 10016.

been misunderstood. Sorting out these issues would, in an ideal world, bring us a few steps further along the road to understanding speech perception.

Misunderstanding #1: "Direct" means "perfect"

Small variations in articulation that lead to the same phoneme are often taken to disprove DR. For example, Diehl, Lotto, and Holt (2004) state, "Even if one restricts the discussion to anatomically possible vocal tract shapes, there are many different ways to produce a given speech signal" (p. 171). This ignores the ability of speakers to recover aspects of these articulations that are indeed distinct and indexical of age, dialect, and so on. (And it allows the notion that the vocal tract normalization that acoustic theories depend on is "acoustic" and not articulatory.) Listeners certainly learn to ignore some recoverable aspects of articulation, as the extensive literature on second language acquisition attests. The vector analysis (Fowler & Smith, 1986) assumed for speech is also assumed for indexical features such as speaker identity. Thus when Nygaard, Sommers, and Pisoni (1994) claim that "it should be noted that independence between talker recognition and phonetic analysis is implicitly assumed by all current theoretical accounts of speech perception" (p. 42), they ignore the assumption of DR that vector analysis includes all aspects of perception, not simply speech, and therefore that interactions (i.e., influences of one vector on another) are indeed included in DR.

Misunderstanding #2: "Direct" means "the signal is irrelevant"

The second misunderstanding is that creatures using direct perception cannot end up with acoustically robust articulations. That is, if the signal is relevant, it is assumed that articulation is irrelevant. One aspect of this mistake is that the information conveyed by the signal is assumed to be specific articulatory shapes. For example, Guenther et al. (1999) state, "A major difference between the auditory target and vocal tract shape target computational model classes is that the former explicitly predict the existence of articulatory trading relations when producing the same phoneme in different contexts, whereas the latter do not. Because the current results show the existence of trading relations in all seven subjects, they appear to favor acoustic target models over vocal tract shape target models" (p. 2864). Part of this assertion depends on reducing the acoustic signal further than human listeners do, such as ignoring formant amplitudes (Iskarous, 2010). However, for purely linguistic purposes, the rough vocal tract shapes associated with higher level features are assumed by DR to be sufficient (Honorof, Weihing, & Fowler, 2011, p. 35). DR posits that listeners extract the trading relations because the acoustics specifies only part of the articulation. Changing one part of an articulation (say, larynx lowering) in apparent compensation for a change in another (say, lip spreading) can result in a signal that is compatible with the original articulation, at least as far as the function of speech is concerned.

Extremely detailed aspects of articulation are nonetheless clearly accessible to perception and learning. The multiple aspects of regional dialects that are tiny but attended to are extensive (Foulkes & Docherty, 2006), even to the point of being maintained by speakers who cannot perceive the difference themselves (Labov, Karen, & Miller, 1991). Thus when Johnson (2006, p. 485) assumes that DR places all variability into a "lawful" category, he assumes that fine detail cannot be lawful simply because it varies by dialect. This would seem to mean that only unlawful variation can be learned, but surely a phonetic difference must be under

control at some level for it to be maintained in a dialect. What is truly difficult to sort out is which aspects of speech directly specify minute details and which specify rough but usable vocal tract shapes. This is the continuing issue.

Misunderstanding #3: "Direct" doesn't really mean anything

A third misunderstanding is that there is no sense in which perception is "direct." This mostly derives from the perfectly correct observation that speech information reaches our brain via sense organs (for vision, see Ullman, 1980). A natural inclination is to take those sense stimuli as sufficient for perception. Thus, "whatever position one adopts about vowel representations, be they purely auditory, somehow reorganized relative to motoric constraints, or purely motoric, the input of the perceptual/decoding process is sensory, and a theory of vowel systems based on auditory patterns can be elaborated" (Schwartz, Boë, Vallée, & Abry, 1997, pp. 282–283). This dependence on sense organs has a long history (e.g., Berkeley, 1709), but it does not account for such phenomena as active versus passive touch (Gibson, 1962; Mace, 1980), sound localization, dyslexia, vocal imitation, and so on. But any animal in a real environment is not interested in the sense qualia; it is interested in the world structuring the signal. An account based solely on the signal, if it is meant to ignore the world, is inadequate, at least in DR. The immense challenges to such fields as computer vision, which begin by analyzing the signal, further support the DR position. The DR treatment of perception is intended to be universal, that is, this is the way biological, successful perceptual systems work. Thus Fowler, following Gibson, is adamant that perceiving speech is like perceiving any other aspect of the world. We are attuned to the functional, distal objects; the proximal stimulus is largely irrelevant. It is clear that the intuitive notion that a sound should be perceived as a sound will continue to hold sway in human thinking and scientific discourse, even if the less obvious notion that a sound can be a vocal tract is closer to the truth.

Motor theory

There are also many instances in which DR is equated with Motor Theory. In the best cases, the similarities between the two make the conflation a minor issue; in others, it is unfortunate. However, a catalog of individual instances would be exceedingly long, and so this discussion is instead short (but cf. Fowler, 1996).

Conclusion

Carol Fowler's version of Direct Realism in speech perception has generated a great deal of debate, not all of which has adequately addressed her position. The three main misunderstandings outlined here are still being made. Examples can still be found for perfection (Perkell, 2012, p. 383), signal irrelevance (Schwartz, Basirat, Ménard, & Sato, 2012, p. 339), and the nonutility of "direct" (Kang, Johnson, & Finley, 2016, p. 88). Pointing out these misunderstandings does not guarantee that Fowler's position is correct. It still lacks a certain level of specificity. This is necessary, given that learning affects the information that is available to any individual perceiver is unique, but it still presents challenges to hypothesis testing. Her position makes no direct claims about brain processes, even though they would

seem to be amenable to such predictions at some stage (e.g., Whalen et al., 2006). The way in which variation is learned is obscure. And it may be that at some future time, the meaning of "direct" will have been modified (in order to accommodate new findings) to such an extent as to be unrecognizable. However, Fowler's position has held up well in the experimental literature, despite these misunderstandings. We can hope that it will be correctly addressed and assessed in work going forward.

Funding

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-002717 to Haskins Laboratories.

ORCID

D. H. Whalen http://orcid.org/0000-0003-3974-0084

References

Berkeley, G. (1709). *An essay towards a new theory of vision*. Dublin, Ireland: Aaron Rhames.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*, 409–438. doi:10.1016/j.wocn.2005.08.002

Fowler, C. A. (1990). Calling a mirage a mirage: Direct perception of speech produced without a tongue. *Journal of Phonetics*, *18*, 529–541.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.

Fowler, C. A. (2010). Embodied, embedded language use. *Ecological Psychology*, *22*, 286–303.

Fowler, C. A., Rubin, P. E., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (pp. 373–420). New York, NY: Academic Press.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1962). Observations on active touch. *Psychological Review*, *69*, 477–491.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, *105*, 2854–2865.

Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, *39*, 18–38.

Iskarous, K. (2010). Vowel constrictions are recoverable from formants. *Journal of Phonetics*, *38*, 375–387.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499. doi:10.1016/j.wocn.2005.08.004

Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, *77*, 84–100. doi:10.1016/j.specom.2015.12.005

Labov, W., Karen, M., & Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, *3*, 33–74.

Mace, W. M. (1980). Perceptual activity and direct perception. *Behavioral and Brain Sciences*, *3*, 392–393.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*, 382–407. doi:10.1016/j.jneuroling.2010.02.011

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*, 336–354. doi:10.1016/j.jneuroling.2009.12.004

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*, 255–286. doi:10.1006/jpho.1997.0043

Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences*, *3*, 373–415.

Whalen, D. H., Benson, R. R., Richardson, M., Swainson, B., Clark, V., Lai, S., & Liberman, A. M. (2006). Differentiation for speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America*, *119*, 575–581.