

Auditory perception is not special: We see the world, we feel the world, we hear the world

Carol A. Fowler

Dartmouth College, Hanover, New Hampshire 03755 and Haskins Laboratories, New Haven, Connecticut 06511

755

(Received 5 December 1990; accepted for publication 14 February 1991)

The literature on "contrast" provides no evidence that durational contrast should occur in the speech and nonsense signals used in research cited by Diehl *et al.* [J. Acoust. Soc. Am. 89, 2905–2909 (1991)]. Moreover, there is evidence that, in comparable signals, it does not occur. Accordingly, their own account of the collection of findings on rate normalization is not viable. Their comments on my research do not imperil my interpretation of it or challenge my criticism that classification judgments of acoustically analogous speech and nonsense signals do not permit interpretation, by themselves, in terms of underlying auditory-system mechanisms. Their arguments that in auditory perception, uniquely, we hear proximal stimulation, not its physical causal sources, is implausible. Their theoretical perspective generally, I argue, is unrealistic.

PACS numbers: 43.71.An, 43.66.Ba, 43.71.Es, 43.66.Mk

INTRODUCTION

I will first clarify the intended scope of the criticisms I raised in Fowler (1990a). Diehl *et al.* (1991) overextend them in both places where they are mentioned (in the first sentence and last paragraph of their commentary), and I would not like them to be generally misunderstood. I did not intend to "challenge the general interpretability of experiments in which the perception of speech sounds is compared to the perception of acoustically analogous nonspeech sounds." Rather, I challenged the interpretability of experiments in which *classification* of speech sounds is compared with the classification of acoustically analogous *nonsense* signals. In fact, I expect that investigations of the perception of real acoustic-signal-causing events will prove quite telling in helping to set findings on speech perception in context. [For example, I expect them to show that the motor theorists (Liberman and Mattingly, 1985) are mistaken when they conclude that speech perception is special *because* listeners to speech recover the causal source of the acoustic signal. I expect research on other sound sources to show that that fact about speech perception is wholly unspecial.] My argument was, and is, that response patterns in identification tasks such as the one used by Diehl and Walsh (1989) do not immediately reflect the character of auditory system processes applied to acoustic signals. Accordingly, they should not be interpreted as if they do. Because perceptual systems have evolved to recover the world of physical events, perceivers will strive to hear a real world cause of any acoustic signal they intercept. That will happen because it must, even when stimuli are nonsense acoustic signals, and subjects' response patterns will reflect that attempt at event recovery.

Having studied the comments of Diehl *et al.*, I stand by my story. I will address several of their comments below. In the section immediately following, however, I consider the account of the findings of Diehl and Walsh (1989) that they continue to put forward.

I. AUDITORY DURATIONAL CONTRAST

Diehl *et al.* (1991) characterize the explanation offered by Diehl and Walsh (1989) of their own and Miller and Liberman's (1979) findings as an explanation "based on a general sensory factor that we refer to as 'durational contrast.'" Diehl and Walsh themselves (1989, p. 2154, italics in the original) refer to "the general auditory principle of *durational contrast*," and, in a report applying the same hypothesis to the ostensible effect of vowel duration on perceived following voiced and voiceless obstruent closure, Kluender *et al.* (1988, p. 161) write: "we suggest that the principle of durational contrast provides a natural explanation of both the speech and the nonspeech results." When I first read the latter two papers, I assumed that there was such a principle, since the authors had written that there was. However, their papers provide clues that should have suggested otherwise. Despite these references to a "principle" of auditory durational contrast, the authors do not provide a single citation either to discussions of the principle itself in the literature on auditory perception or to findings of durational contrast. A search of the literature (reported in Fowler, in press a) revealed no mention of such a principle and a very small number of reports of durational contrast, always for signals longer than those under investigation in the research by Diehl and Walsh and Kluender *et al.*; the search also revealed findings of assimilation—that is, the reverse of contrast. As far as I can determine, there is no principle of auditory durational contrast, and there is no body of research establishing the conditions under which durational contrast will be found. Accordingly, when Diehl *et al.* lay out two conditions for obtaining contrast that my research failed to meet, they are making up the conditions. When they claim, in addition, that because the phase 2 information in my ramp events had such a large effect on listeners' judgments, and the phase 1 information had a correspondingly small effect, there was "little room for a

second-order effect of durational contrast" on judgments of phase 1, they are making up that claim as well. Intuition might suggest, alternatively, that when the source of an effect (here, contrast) is perceptually salient and when the object of its influence (the phase 1 acoustic signal) is not, the effect might thereby be enhanced, not reduced. Indeed, I used sandpaper on the track ramps to reduce the intensity of phase 1 of the acoustic signals and enhance any effect of phase 2.

Reports of contrast effects of other sorts are quite widespread, and one might extrapolate from their generality a hypothesis that auditory durational contrast might be found under conditions analogous to those under which other ostensible contrast effects are observed. But certainly this literature does not establish the existence of a principle nor does it justify the assertion that any such effects are due to a "general sensory factor."

Diehl and Walsh do not cite the general literature on contrast at all. However, Kluender *et al.* (1988), who invoke durational contrast to explain why long vowels are associated with short voiced-stop closures in many languages, refer to Warren's (1985) review and theoretical work, and they cite some of the research cited by Warren himself. However, Warren's paper is a surprising one for these researchers to cite, because his view of contrast is not compatible with their own. He does not see these effects as sensory distortions of stimulus input. He sees them as products of an adaptive "criterion shift rule" that reflects the way that past experience with tokens of a type affect perceptual judgments of new tokens. In Warren's conceptualization, judgments of a token are made relative to a standard that changes with experience; sensibly, immediately past experience with a token influences the standard for its type disproportionately. Accordingly, a heavy weight hefted on one trial of a weight-judgment test moves the standard considerably in its own direction; a lighter weight hefted next is judged very light, because it is far from the new standard. In turn, the light weight now exerts a strong effect on the standard, and the heavy one, now two trials back, a considerably attenuated one. (See also Bjork and Bjork, in press, and Freyd and Johnson, 1987, for other evidence of "memory averaging" that has this characteristic weighting.) My reading of his perspective on criterion shifts suggests to me that Warren would not predict shifts under the conditions of the research of Diehl and Walsh or Kluender *et al.* because the criterion shift applies only to objects or events to which the past experience reflected in the standard is judged relevant—that is, to tokens of a type. [For example, according to Johnson (1944), in the weight-judging research, hefting objects such as a book or chair that are believed by subjects not to be among the experimental stimuli has no effect on the standard.] Were there durational contrast effects, they would be observed as trial-to-trial effects of vowel durations on judged vowel durations, not as within-trial effects of vowel durations on perceived consonant durations. The effects that Miller and Liberman observed are unlikely to be due to criterion shifts of the sort described by Warren.

Further, a direct test of the hypothesis of Kluender *et al.* that vowel durations exert a contrastive effect on judgments of following consonant closure did not reveal one; rather,

assimilation effects occurred. (Fowler, in press a). In the same study, a direct test for contrast in the analogous nonsense signals used by Kluender *et al.* did not reveal contrast either. In respect to this last test, I was unable to replicate findings by Kluender *et al.*; Kluender (1990) has told me that he too has failed to replicate his earlier findings.

The upshot is this: The literature does not reveal a principle of auditory durational contrast; the literature on criterion shifts does not appear to predict that variation in vowel durations should cause criterion shifts in judgments of consonant durations; compatibly, direct tests for contrast in the speech and nonspeech signals to which Kluender *et al.* applied a contrast hypothesis do not, in fact, reveal contrast. For these reasons alone, Miller and Liberman's account of their findings is not threatened by a contrast hypothesis.

A. Drawing inferences from response patterns to auditory mechanisms

Diehl and Walsh (1989) had found similar response patterns to speech signals modeled after those of Miller and Liberman and to acoustically similar nonsense signals. They ascribed the similar patterns to similar effects of auditory contrast in the similar acoustic signals. I disagreed in Fowler (1990a). One argument I used was that listeners' response patterns reflect their efforts to recover real-world causes of the acoustic signals they intercept. In their commentary, Diehl *et al.* (1991) ask why nonsense signals characterized by listeners as laser sounds and computer noises should give rise to rate normalization and thereby to response patterns like those to the speech syllables of Miller and Liberman. Given that lasers and computers are physically unlike the vocal tract, the reason cannot be that similar events are recovered in perception of the speech and nonspeech signals. In particular, they write that "Fowler provides no rationale for assuming that a rate normalization scheme tailored to...particular facts about stops and glides would be the natural and expected scheme in the case of the above kinds [laser sounds, etc.] of implied source" (p. xxx).

In fact, I thought that I had laid out in considerable detail the conditions (see my footnotes 4 and 5) that a physical event would have to satisfy to undergo rate normalization in the way that /ba/s and /wa/s do. However, a major point that experiments 1 and 2 of Fowler (1990a) make is that the critical response pattern can be obtained for reasons that have nothing to do with either contrast or rate normalization. All that is required is a biphasic physical event in which the acoustic signal caused by phase 2 of the event provides the right kind of information about phase 1. Descriptions such as "laser sounds" and "computer noises" are not very informative. I do not know what about the listeners' percepts led to the response patterns (except that I know now that it was not durational contrast), but neither do Diehl and Walsh. Our level of ignorance would be reduced considerably were the signals not nonsense signals.

II. PERCEPTION

Obviously, Diehl *et al.* look at perception as a considerably different kind of achievement than I do. Here are rea-

sons why I prefer my theoretical perspective and consider theirs unworkable, unwieldy, and unrealistic.

A. Appeals to covert processes and mechanisms

They interpret the results of my experiments 1 and 2 as evidence only that listeners can use their everyday knowledge of events and their physical properties to guide categorization and inference making. They then quote a passage from Diehl and Walsh (1989) in which the authors write as if they assume that an account of Miller and Liberman's findings can only take one of two forms. It can appeal, as Diehl *et al.* do to explain the findings in Fowler (1990a), to the listener's "tacit knowledge," now of speech and rate variation, or it can appeal to general auditory processes. Of the two, they suggest, the appeal to tacit and speech-specific knowledge should be last resort. Appeals to general auditory mechanisms are preferred because they "are simpler, more general, and more likely to be derivable from known principles" (Diehl and Walsh, 1989, p. 2162). (In this case, the "known principle," of course, is auditory durational contrast.)

I agree that appeals to tacit knowledge are to be avoided, but I extend that avoidance to any covert process or mechanism until appeals to public information—here, that in the acoustic signal—have been made. Diehl *et al.* do not explain why such appeals are ruled out for the stimuli of experiments 1 and 2 of Fowler (1990a) or for those of Miller and Liberman (1979).

Notice, too, that, although Diehl and Walsh do not discuss information in the signal as a third source of explanation for listeners' response patterns, that is presumably an oversight, because appeal to that information must be made in any case. Auditory mechanisms and tacit knowledge cannot determine what was said without help from the signal. In addition, acoustic speech signals must be the source of tacit, speech-specific knowledge that listeners may have. It follows, indeed, that use of tacit knowledge cannot be *necessary* for interpretation of perceptual objects, because perception is the source of that knowledge. Further, since tacit knowledge is obtained via perception, it can be no better than the information provided to the perceptual system. (I am assuming that Diehl *et al.* consider the tacit knowledge they mention to be acquired developmentally, since they invoke it to explain performance by subjects in my ramp experiments; if they consider it to be innate, the same problem arises. In order for natural selection to yield tacit knowledge, if it does, there must be systematicities in the environment to shape the selection. I assume also that Diehl *et al.* consider the tacit speech-specific knowledge to be obtainable via the auditory system, since blind persons learn spoken languages.)

A related comment derives from Diehl *et al.*'s acknowledgment that it would be "something of a miracle" were durational contrast to accomplish the equivalent of rate normalization exactly. Where it does not, "genuine rate normalization would presumably come into play." (p. 2906) But how do perceivers know how much rate normalization is enough in a given instance? The only way to know is to recover information for rate from the signal. So the account that Diehl *et al.* propose is that: (1) the information in the

signal specifies rate; (2) auditory contrast overgenerates normalization, undergenerates it, or occasionally, miraculously, gets it right; (3) the perceiver makes an approximate adjustment where necessary. My account is simpler [see (1) above].

A last objection to appeals to explanation in terms of auditory mechanisms is specific to the appeals by Diehl and colleagues; it is that they fall one "because" short of completeness. Our auditory system has evolved, continually attuning itself to its function of acquainting perceivers with relevant properties of the world. If there were a general auditory mechanism responsible for findings that Diehl *et al.* ascribe to contrast or that Warren refers to as criterion shifts, its function would be unlikely to be one of distorting stimulus input. The "because" missing from Diehl *et al.*'s account is the one that explains why contrast is so pervasive in perceptual judgment. Warren's account offers such a "because," but it does not apply to the stimuli of Diehl and Walsh or of Kluender *et al.*

B. Recovery of the environment from stimulation at the sense organs

Diehl *et al.* address my two claims that, in perception universally, perceivers recover physical events, and that their doing so is essential to survival. Whereas they describe the claims exactly right, the objection they raise is irrelevant to them. The objection they raise is that not all different objects or events in the environment structure stimulation differentially. (In their example, there are some objects that look to us as if they are the same color when in fact they absorb and reflect different wavelengths of light.) However, a claim that we perceive the world is not a claim that we perceive everything in the world; nor, correspondingly, is it a claim that every different thing is perceived differently. For something to be perceived at all, it must structure a medium such as light or air that, in turn, can impart its structure to a perceptual system. For two things to be perceived as different, they must structure the medium differently and the differences must be detectable by the relevant perceptual system. Those conditions will hold, other direct realists and I claim, nearly always (mirages excepted) for relevant properties of objects and events with which we interact in the world. Even if objects toward which we must behave differentially do not distinguish themselves by their color, they will distinguish themselves by their shape, size, behavior, etc.

Diehl *et al.* point out that it would not matter for our survival if we heard the acoustic speech signal rather than its vocal-tract source. That may be the case. However, elsewhere two of the authors (Diehl and Kluender, 1989) have argued that speech perception is not special in involving processes distinct from those of the general auditory system. I almost agree. But there is a consequence of this view. If it is crucial sometimes that perceivers know properties of the world (and if it is never crucial, as it never is, that they perceive reflected light, patterned air, or patterned deformations of the skin and joints), then perceptual systems will recover world properties when it is crucial and when it is not crucial. That is, evolution will have "designed" them to use

stimulus structure at the sense organs as information for its cause. Accordingly, they cannot sometimes recover objects and sometimes not depending on whether object recovery is crucial or not. (For one thing, they cannot determine crucialness without knowing what the object is; once they know that, they are done perceiving.)

Diehl *et al.* suggest alternatively that in cases where there is no adaptive advantage to perceiving physical events directly, "what counts in these cases is whether the listener is able to judge what kind of object or event is associated with a given sound." For example, we need not perceive a lion's vocal tract via its roar; to use the roar to our advantage, we need only associate the perceived acoustic signal with the lion. I infer that Diehl *et al.* consider it never crucial to recover the world from the acoustic signal. In their view, that is, we always perceive acoustic perceptual objects. For two reasons, I consider this idea highly implausible.

First, it makes the auditory system quite different in nature from the visual and haptic systems, which cannot yield light-ray or skin-deformation perceptual objects. Consider the authors' own example. How can a perceiver associate an acoustic signal of a lion's roar with a lion? The lion is in the world and the perceived signal is in the perceiver's head. If the visual system were to work in the way that the auditory system putatively works, there could be no association of the roar signal with the lion; the roar signal would have to be associated with reflected light patterns caused by lions. This would be very bad for the perceiver, who needs to know that there is a lion nearby. So there must be a perceptual system, here the visual system (and Diehl *et al.* grant this in their reply), that uses structure in stimulation, not as something to be perceived, but as information for world objects and events. The lion, perceived, thanks to the visual system, may now be associated with the acoustic signal caused by its roar. Now consider blind persons. They must acquire knowledge of the world perceptually too if they are to guide their actions in it. Lacking a working visual system that would yield the world, they must get information about the world of events in some other way. The only viable options are the auditory system and the haptic system. To maintain the view of Diehl *et al.* that objects of auditory perception are acoustic, we must pick the haptic option. (Are there independent grounds?) So now there is a second perceptual system that uses structure in stimulation at the sense organs, not as an object of perception in itself, but as information for the world. Is it reasonable to suppose that the auditory system works differently from these systems? In visual and haptic perception, stimulus structure that has been caused by an event is used as information for the event. In acoustic perception, Diehl *et al.* aver, however, stimulus structure in the air that has been caused by an event is heard in itself. Why? And why does this allegedly acoustic-signal perceiving system localize sound, not where the acoustic signal is (in the ear), but where the acoustic-signal causing event is in the world? Give it up, Diehl *et al.*

I suspect that Diehl *et al.* consider that it involves an extra perceptual step to go from the signal to the world so that perceivers would not bother taking the step if they did not need to. However, in fact, the difference has to do, not

with processing steps, but with what it is that structure in stimulation counts as perceptually: As a thing or as information for its causal source.

A second reason why I consider it implausible that the auditory system, uniquely, perceives stimulation at the sense organ is the literature on speech perception. Consider these findings.

(1) Listeners to speech recover a parsing of the acoustic speech signal that is surprising were they aiming to recover an acoustic perceptual object (see Fowler, in press b, for a review of the literature). In particular they recover parsings of integral dimensions of the acoustic signal such as fundamental frequency. For example, effects on fundamental frequency of vowel height variation and declination (e. g., Silverman, 1987) are not heard as part of the pitch melody of an utterance even though their effects are perceptually significant in ways to be considered next. Complementarily, acoustically distinct consequences of the same linguistic gesture of the vocal tract can be perceptually equivalent. So, for example, either a higher fundamental frequency (Reinholt Peterson, 1986; see also Kingston, in press) or a lower F_1 , both consequences of a gesture of the vocal tract for achieving a high vowel, can make a vowel sound linguistically higher. In short, listeners appear to use the constellations of acoustic consequences of a linguistically significant gesture of the vocal tract as information for their causal source—the gesture itself.

(2) In the "McGurk effect" (McGurk and MacDonal, 1976), optical information for a spoken syllable (in the form of videotaped presentation of someone talking) can affect what a listener experiences hearing when the tape is dubbed. This effect is strong phenomenally, and it is naturally explained in a theory in which physical causes of optical and acoustic stimulation are perceived. Diehl and Kluender (1989) suggest that such effects are due to learned associations (compare the association that Diehl *et al.* invoke between the roar and the visible lion). However, that account does not explain the data. McGurk effects are absent when the video display is replaced by printed syllables (Fowler and Dekle, in press), despite the demonstrated existence of associations in memory between spellings and sounds (e.g., Seidenberg and Tanenhaus, 1979). In the same study, crossmodal effects are present when the video display is replaced by the manual feel of a face producing a syllable, despite the absence of relevant crossmodal associations in memory. Crossmodal influences on the sound of a syllable apparently occur when the perceiver is tricked into perceiving one physical event whose character is signaled crossmodally.

C. Premature allegations of a lack of specificity in acoustic speech signals

There is no guarantee as far as I know that structure in a medium that is caused by an event will be specific to the event. However, it is not wholly surprising to find that different events generally do structure media in ways distinctive to themselves, because it is the different physical properties of the events that cause the structure in the light, air, or on a

body, surface. This is how structure in the media can inform. Even so, if two different events structure light, air, or the body in sufficiently similar ways, they will not be distinguished perceptually. This is how we experience mirages. Outside the laboratory, however, mirages are not commonplace. I ascribe that fact to the general validity of the "doctrine of necessary specificity" (Turvey and Shaw, 1979)—the idea that, for our survival in the world, information at the sense organs must generally specify its causal source.

I am wholly unconvinced by Diehl *et al.*'s discussion of the lack of specificity in speech signals. Possibly, / α / is an example. That is, either the several ways of producing / α / do not yield distinct acoustic patterns (although I doubt this) or the differences they do cause are not reliably distinguished by perceivers. That certainly helps to explain the extreme unpopularity of / α / both crosslinguistically and among dialects of English. Unless Maddieson (1984) uses a symbol I do not recognize, not one of the 317 languages he surveys has an / α /. (English is not one of the sampled languages.) Further, many dialects of English lack / α / (including dialects of British English, Australian English, New York City English, and others). However, as for the general claim that the acoustic speech signal does not specify its source, I suggest that the reader wait and see. One can find in the literature on visual perception similar confident claims of specificational failures of the optic array [my favorite is Molyneux's premise (Pastore, 1971) that distance cannot be perceived based on optical information] that were subsequently shown to be false. Research is just beginning on ecological acoustics (e.g., Rosenblum *et al.*, 1987; Schiff and Oldak, 1989); as it progresses, we will begin to have a better handle than we seem to have now on the kinds of macroscopic structure that audible events at the human scale cause in air.

One source of progress in investigation of informative optical information was the realization that considerably more information is available in optical information that unfolds over time than is available in a temporal snapshot (e.g., Gibson, 1966). The same realization will help the search for specifying information in speech. Diehl *et al.* cite findings by Ladefoged *et al.* (1978) that more than one static configuration of the vocal tract can give rise to the same acoustic signal. However, that must largely cease to be true if the signal is produced in running speech. In running speech, there are the physical constraints that any vocal-tract "configuration" must be achievable smoothly from what came before and that it must permit smooth achievement of what comes later; these constraints must mostly rule out all but the actual "configuration."

A lesson that recent research on haptic information for physical objects has made is that finding information for an object requires knowing the relevant object properties that cause patterning in stimulation at the sense organs (Solomon and Turvey, 1988; Solomon *et al.*, 1989). It is difficult to know what information to look for without having an explicit idea of what it is informing about. The critical information in the speech signal will be the structure in the air that is caused by the linguistically significant actions of the vocal tract. It will not do to look for information for jaw

motion, for example, even though, of course, the jaw does move and structure the signal. Jaw movement is mostly the vector sum of more than one phonetic gesture of the vocal tract. Listeners sort out the gestures, as I have already indicated, via the constellation of acoustic effects they cause, and so we must discover what the gestures are and how *they* structure the signal.

D. Mirages and caged animals

I found this brief discussion rather muddling. Literally, I am asked whether a quail or chinchilla could "distinguish" synthetic from natural speech. I answer "sure," because the signals are systematically different in whatever ways causes the one signal to sound mechanical and the other not.

Perhaps something more is being asked, however. Possibly, Diehl *et al.* are asking whether a chinchilla or quail could actually perceive a synthesizer producing the acoustic signals it produces. If that is the question, my answer is "no," but I argue that humans do not perceive the synthesizer either. It is true that our verbal response is appropriately "synthesizer" or "natural speech;" however, it does not follow that we *perceive* the synthesizer from the signal. Consider this analogy. I look out on the road on a hot day and see an apparent puddle. I know that it is not a puddle, because it has not rained for 6 weeks, and so, when someone asks me what is out there, I say that there is heat rising off the road. But the fact is that my *knowledge that* it has not rained has no effect on what the thing *looks* like; it looks wet. It does so because the heat structures the light in very much the same way as water does, thereby creating a mirage; we *see* water rather than heat rising, because water is considerably more ecologically, and hence perceptually, salient (Coss and Moore, 1990) than is rising heat. (This, by the way, is my answer to the begged question to which Diehl *et al.* allude; see also Fowler, 1990b. It is, further, testimony to the irrelevance of knowing that, whether tacit or explicit, to perception.) By the same token, I do not believe that humans perceive the synthesizer doing whatever it does when it mimics acoustic speech signals. Even though we *know that* it is a synthesizer, we *hear* a vocal-tract-like physical system producing linguistically significant gestures. We do so because the synthesizer structures the air in much the same way as human vocal-tract gestures for speech do, thereby creating a mirage; we hear speech when either speaker or synthesizer productions are signaled because of the "profound biological significance of speech" (Whalen and Liberman, 1987, p. 171) and the corresponding relative insignificance of synthesizers mimicking speech signals. Moreover, I expect that, given synthesized human speech, quail perceive it more-or-less as we do, albeit stripped of its linguistic significance. Given synthesized quail vocal productions, we hear it more-or-less as the quail do, stripped of its communicative significance. In each case, we wrongly recover a physical system similar to the natural vocal source rather than the synthetic one, because of the distinct relative ecological, and hence perceptual, salience in evolution of the behaviors of other animals, in comparison with the natural-acoustic-signal-mimicking properties of recent human inventions.

ACKNOWLEDGMENTS

Preparation of the manuscript was supported by NICHD Grant HD-01994 to Haskins Laboratories.

- Bjork, R., and Bjork, E. (in press). "A new theory of disuse and an old theory of stimulus fluctuation," *Festschrift for William K. Estes*, edited by A. Healy, S. Kosslyn, and R. Shiffrin (Lawrence Erlbaum Associates, Hillsdale, NJ).
- Coss, R., and Moore, M. (1990). "All that glistens: Water connotations in surface finishes," *Ecol. Psychol.* **2**, 367-380.
- Diehl, R., and Kluender, K. (1989). "On the objects of speech perception," *Ecol. Psychol.* **1**, 121-144.
- Diehl, R., and Walsh, M. (1989). "An auditory basis for the stimulus-length effect in the perception of stops and glides," *J. Acoust. Soc. Am.* **85**, 2154-2164.
- Diehl, R., Walsh, M., and Kluender, K. (1991). "On the interpretability of speech/nonspeech comparisons," *J. Acoust. Soc. Am.* **89**, 2905-2909.
- Fowler, C. A. (1990a). "Sound-producing sources as objects of speech perception: Rate normalization and nonspeech perception," *J. Acoust. Soc. Am.* **88**, 1236-1249.
- Fowler, C. A. (1990b). "Calling a mirage a mirage: Direct perception of speech produced without a tongue," *J. Phon.* **18**, 529-541.
- Fowler, C. A. (in press a). "Vowel duration and closure duration in voiced and unvoiced stops: There is no contrast effect here," *J. Phonet.*
- Fowler, C. A. (in press b). "Listener-talker attunements in speech," in *Human Development and Communication Sciences*, edited by T. Tighe, B. Moore, and J. Santrock (Lawrence Erlbaum Associates, Hillsdale, NJ).
- Fowler, C. A., and Dekle, D. J. (in press). "Listening with eye and hand: Crossmodal contributions to speech perception," *J. Exp. Psychol.: Human Percept. Perform.*
- Freyd, J., and Johnson, J. Q. (1987). "Probing the time course of representational momentum," *J. Exp. Psychol.: Learn. Mem. Cognit.* **13**, 259-268.
- Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems* (Houghton-Mifflin, Boston).
- Johnson, D. (1944). "Generalization of a scale of values by the averaging of practice effects," *J. Exp. Psychol.* **34**, 425-436.
- Kingston, J. (in press). "Integrating articulations in the perception of vowel height," *Phonetica*.
- Kluender, K. R., Diehl, R. L., and Wright, B. A. (1988). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," *J. Phon.* **16**, 153-169.
- Kluender, K. R. (1990). Personal communication.
- Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L. (1978). "Generating vocal-tract shapes from formant frequencies," *J. Acoust. Soc. Am.* **64**, 1027-1035.
- Lieberman, A., and Mattingly, I. (1985). "The motor theory revised," *Cognition* **21**, 1-36.
- Maddieson, I. (1984). *Patterns of Sounds* (Cambridge U.P., Cambridge, England).
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature* **264**, 746-748.
- Miller, J., and Liberman, A. (1979). "Some effects of later-occurring information on the perception of stop-consonant and semivowel," *Percept. Psychophys.* **25**, 457-465.
- Pastore, N. (1971). *Selective History of Theories of Visual Perception, 1650-1950* (Oxford U.P., New York).
- Reinholt Peterson, N. (1986). "Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations," *Phonetica* **43**, 31-42.
- Rosenblum, L., Carello, C., and Pastore, R. E. (1987). "Relative effectiveness of three stimulus variables for locating a moving sound source," *Perception* **16**, 175-186.
- Schiff, W., and Oldak, R. (1989). "Accuracy of judging time to arrival: Effects of modality, trajectory and gender," *J. Exp. Psychol.: Human Percept. Perform.* **16**, 303-316.
- Seidenberg, M., and Tanenhaus, M. (1979). "Orthographic effects on rhyme monitoring," *J. Exp. Psychol.: Human Learn. Mem.* **5**, 546-554.
- Silverman, K. (1987). "The structure and processing of fundamental frequency contours," Ph.D. dissertation, Cambridge University, Cambridge, England.
- Solomon, H. Y., and Turvey, M. T. (1988). "Haptically perceiving the distances reachable with hand-held objects," *J. Exp. Psychol.: Human Percept. Perform.* **14**, 404-427.
- Solomon, H. Y., Turvey, M. T., and Burton, G. (1989). "Perceiving extents of rods by wielding: Haptic diagonalization and decomposition of the inertia tensor," *J. Exp. Psychol.: Human Percept. Perform.* **15**, 58-68.
- Turvey, M. T., and Shaw, R. (1979). "The primacy of perceiving: An ecological reformulation of perception for understanding memory," in *Perspectives on memory research: Essays in Honor of Uppsala University's 500th Anniversary*, edited by L. G. Nilsson (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 167-222.
- Warren, R. M. (1985). "Criterion shift rule and perceptual homeostasis," *Psychol. Rev.* **92**, 574-584.
- Whalen, D., and Liberman, A. (1987). "Speech perception takes precedence over nonspeech perception," *Science* **237**, 169-171.