

ON THE DISSOCIATION OF AUDITORY AND PHONETIC PERCEPTION

MICHAEL STUDDERT-KENNEDY

Queens College and Graduate Center, City University of New York and Haskins Laboratories, 270 Crown Street, New Haven, CT. 06510, U.S.A.

My charge for this meeting is to review some facts about speech perception that might seem to be explicable, or not explicable, by reference to a peripheral representation of the speech signal. Since many later speakers will surely report what they think can be explained, I shall confine my remarks to what I think cannot.

Let me state, at the outset, that my view is, in the current jargon, uncompromisingly "bottom up." Whatever "top down" processes may contribute to speech perception (and, in my view, it is very little) I shall not consider it here. In other words, my initial and unequivocal assumption is that all information necessary to support the perception of speech is present in the acoustical signal and in its peripheral auditory representation. Of course, much remains to be learned about the acoustic structure of speech and about its auditory transformations. But, beyond that, lies the deeper question of what the acoustic structure means. What does it refer to? Certainly, it conveys information, but information about what? We lack insight into the nature of the perceptual object, and therefore of the mechanism by which raw audition becomes speech perception.

There is nothing biologically unsound here. On the contrary, all animals live in the same terrestrial world and are exposed, or could be exposed, to the same patterns of radiant or mechanical energy, but they differ in their capacities to use those patterns for information about the world. Indeed, as von Uexküll (28) suggested many years ago, the interlocking of animal and environment, that is, the peculiar fit of each species to its niche, is a definition of that species. We might do well in our study of speech to take a leaf from the ethologist's notebook.

Consider, for example, the barn owl and the extraordinary precision with which it can localize sources of sound (4,8). All the information necessary for localization is present in the peripheral sensory arrays, properly viewed.

But the physiologist contemplating those arrays could never discover their import, if he were not guided by the ethologist's knowledge of how the owl lives. Moreover, even when the physiologist has guessed that the changing auditory patterns provide a map, he still cannot understand how they do so in isolation, because the map only emerges from a central, midbrain comparison of input from the two ears (4). For speech perception a double input is not necessary, but we do need a central mechanism to read the phonetic import of the auditory array.

In what follows, I will do four things:

- cite evidence that auditory and phonetic perception are distinct processes that can be dissociated experimentally;
- infer from certain experimental results something about how phonetic perception uses auditory information;
- illustrate how phonetic perception may be engaged, or disengaged, ontogenetically and experimentally;
- speculate briefly on the biological function of phonetic perception.

Briefly, my broad hypotheses are: that phonetic perception is a specialized mode of auditory localization, in which we use sound to inform ourselves about the positions and changes in position of a speaker's articulatory apparatus; that the phonetic percept is an abstract buffer intermediate between the auditory array and the articulatory mechanisms necessary to reproduce that array; and that one function of phonetic perception is to facilitate language learning.

THE DISSOCIATION OF AUDITORY AND PHONETIC PERCEPTION

Twenty years ago, Fant (6) drew a distinction between the speech signal and its linguistic message. He illustrated compellingly the discrepancy between the sound segments of the acoustic event and the phonemic segments of its percept. He emphasized that the two levels are not contradictory, but "different aspects" of the same event. I propose to follow Fant's lead, extending his description of the signal upstream into the auditory system and his description of the linguistic message downstream into phonetics. In other words, I shall distinguish between two qualitatively different aspects of a single percept: auditory and phonetic. I shall argue that the two perceptual modes are different, active, "attentional" modes of scanning the signal for information.

Duplex perception. The distinction can be demonstrated experimentally in an effect, discovered by Rand (20), recently elaborated and dubbed "duplex perception" by Liberman and his colleagues (12,15): two different percepts, one auditory, the other phonetic, arise simultaneously when the acoustic constituents of a synthetic syllable are separated and presented dichotically. Figure 1 displays a 9-step continuum of patterns sufficient to induce the effect, taken from Mann, Madden, Russell and Liberman (15). If the base (bottom left) is presented alone, it is usually heard as [da]; if one of the isolated transitions (bottom right) is presented alone, it is heard as a non-speech "chirp." If the two patterns are presented dichotically in appropriate temporal alignment, the listener hears a fused syllable (either [da] or [ga], depending on which transition is presented) and, at the same time, a non-speech chirp perceptually identical to the chirp heard in isolation. If, now, the patterns are presented for discrimination in pairs of stimuli, separated by three steps along the continuum, with instructions to attend on one series of trials to the speech percepts and on the other series of trials to the non-speech chirps, the results for a typical subject are those of Figure 2: for the non-speech judgments, a more-or-less continuous descending function, for the speech judgments, a discrimination function peaked at the phoneme boundary in the fashion typical of categorical perception (11). We thus have clear behavioral evidence for perception of a single acoustic event in two different perceptual modes at the same time.

Audio-visual adaptation. A second, perhaps even more compelling demonstration of the dissociation derives from an effect discovered by McGurk and MacDonald (16). They demonstrated that listeners' perceptions of a speech sound presented over a loudspeaker could be changed, if they simultaneously watched a videotape of a speaker producing another sound. For example, presented with audio [ga] and video [ba], a subject will typically report a cluster [bga] or [gba]. On the other hand, presented with the reverse arrangement, audio [ba] and video [ga], a subject will typically report a sort of auditory-articulatory blend [da]. Notice that, in this instance, the phonetic percept corresponds to neither visual nor auditory pattern. The effect is not well understood, but evidently arises from a process by which two continuous sources of information, acoustic and optical, are actively combined at an abstract level at which each has already lost its distinctive sensory quality (for fuller discussion, see Summerfield (27)).

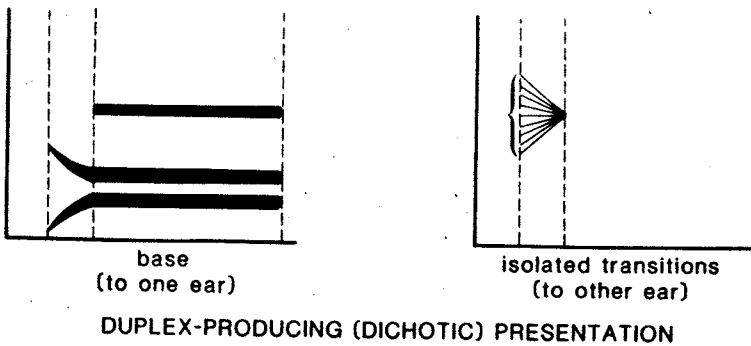
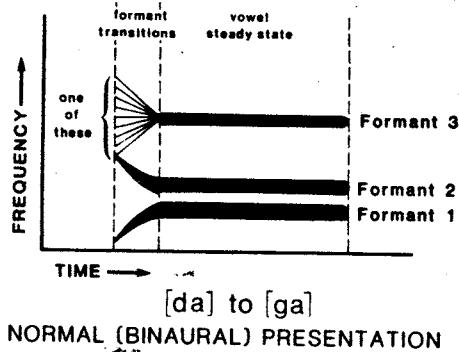


Fig. 1. Schematic representation of the stimulus patterns used to study the integration of formant transitions (from (15)).

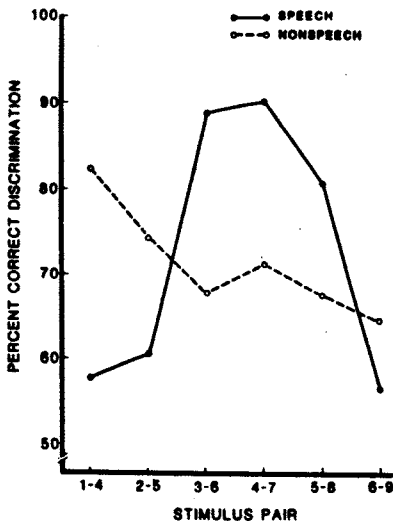


Fig. 2. Discriminability of formant transitions when, on the speech side of the duplex percept, they supported perception of stop consonants and when, on the nonspeech side, they were perceived as chirps (from (15)).

Roberts and Summerfield (24) took advantage of this phenomenon to test the nature of speech adaptation, and in so doing demonstrated dissociation of auditory and phonetic processes. Adaptation, it will be recalled, is a form of perceptual contrast, induced by prolonged exposure to some fixed pattern of energy. The most familiar example comes from color vision. If we stare for several minutes at a circle of red light and then look away at a neutral background, we will see a circle of a somewhat unsaturated, complementary green. A similar effect can be demonstrated in speech. First, a baseline is established by asking subjects to identify items along a synthetic speech continuum, stretching from, say [b_e] to [d_e]. Next, subjects are exposed to repeated presentations (typically, over a hundred presentations at a rate of about two per second) of one or the other endpoints of the continuum. Finally, subjects are again asked to identify items drawn from the continuum. The last two steps are repeated as often as is necessary to obtain a reliable adapted identification function. On this function, subjects typically report fewer instances of the class of items on which they have been "satiated" or "adapted" than they did before they were adapted. In other words, the category boundary is shifted toward the end of the continuum to which the subject has been repeatedly exposed.

Roberts and Summerfield's ingenious twist on this paradigm was to expose subjects to an audio-visual adaptor, audio [b_e], video [g_e], intended to be perceived phonetically as [d_e]. In the event, six of their twelve subjects perceived the adapting pattern as either [d_e] or [ʒ_e], four as [kl_e], one as [fl_e], one as [ma]. Not a single subject perceived the phonetic event corresponding to the auditory event actually presented, namely [b_e]. Yet every single subject displayed a significant shift of his [b_e-d_e] boundary toward [b_e]. In other words, the procedure effectively dissociated auditory and phonetic processes. Subjects' only conscious percepts were phonetic, but their auditory systems demonstrated, by their response to adaptation, that they were simultaneously registering the acoustic signal presented over the loudspeaker.

I take these two experimental procedures of duplex perception and audio-visual adaptation to demonstrate unequivocally the "on-line" dissociation of auditory from phonetic perception. Moreover, following Summerfield (27),

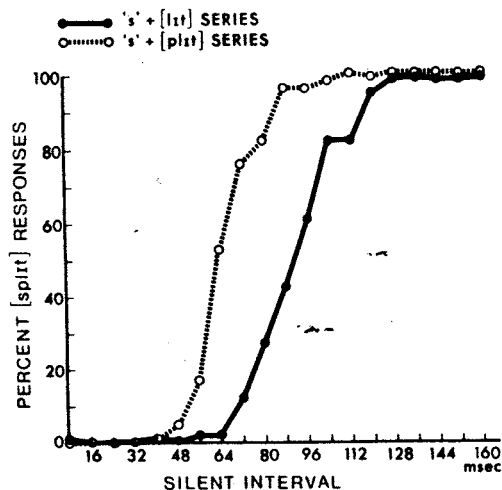


Fig. 3. Effect of silent interval on the perception of [sliit] vs. [spliit] for the two settings of the labial transition cue. (Reprinted from (7) by permission of the Psychonomic Society, Inc.)

I take the results of the audio-visual adaptation study to demonstrate that the support for phonetic perception is information about the common source of acoustic and optical information, namely, articulatory dynamics.

THE USE OF AUDITORY INFORMATION FOR PHONETIC PERCEPTION

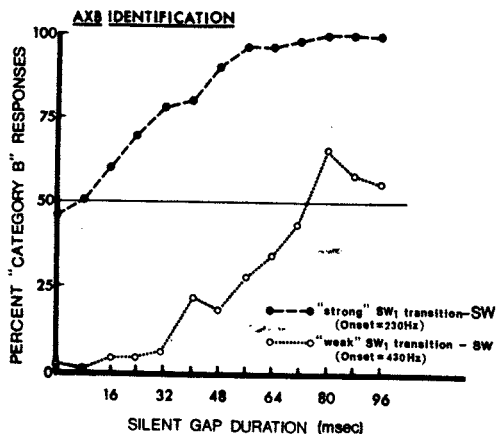
A commonplace in discussions of speech perception is that the information for particular phonetic segments is often widely scattered, both temporally and spectrally, through the signal (13). Again, Fant (6) remarked that a single segment of sound may carry information about several segments of the message and a single segment of the message may draw information from several segments of the sound. For an example of the latter, consider the work of Bailey and Summerfield (1). They showed that perceived place of articulation of a stop consonant, induced artificially by inserting a brief silence between [s] and a following vowel (as in [si] and [su]), depends in English on the duration of the silence, on spectral properties at the offset of [s] and on the relation between these properties and the following vowel. One function of the phonetic mode is apparently to integrate, or to exploit the integration of, such diverse auditory properties into a coherent percept.

Support for this hypothesis comes from two recent experiments. The first was done by Fitch, Halwes, Erickson and Liberman (7), who demonstrated the perceptual equivalence of two distinct cues to a voiceless stop in a fricative-stop-liquid cluster: silence and rapid spectral change. They constructed two synthetic syllables [pIIt] and [lIIt], the first differing from the second only in having initial transitions appropriate to a labial stop. If a brief band-passed noise, sufficient to cue [s], was placed immediately before these syllables, both were heard as [sIIt], but if small interval of silence, long enough to signal a stop closure, was introduced between [s] and the vocalic portion, both were heard as [spIIt]. What is of interest is that the silent interval necessary to induce the stop percept was shorter when the vocalic portion carried labial transitions than when it did not. Figure 3 displays the perceptual effect of systematically manipulating the duration of the silent interval before each of the two syllables: the procedure titrates the initial transition and shows that it is equivalent to roughly 25 msec of silence.

On the face of it, there would seem to be no psychoacoustic grounds for this spectral-temporal equivalence. However, Delgutte (this volume, Figure 4), testing the output of his model of the peripheral auditory system for an analogous equivalence between the rise-time of friction noise and the duration of preceding silence in the fricative-affricate distinction (studied by Repp, Liberman, Eccardt and Pesetsky (22)) finds that these different cues do have similar effects on the response of his model: "...both a decrease in rise time and a longer silence duration increase the amplitude of the peak in discharge rate at the onset of friction."

Nonetheless, it seems that listeners are only able to make use of this auditory information, if they are listening in the speech mode. This is the implication of a study by Best, Morrongiello and Robson (2) using "sine-wave speech" (21). Best and her colleagues constructed a sound from three sine waves modulated to follow the path of the center frequencies of the three formants of a naturally spoken syllable [dei], in two forms: one form had a relatively long initial F_1 transition ("strong" [deI]), one had a relatively short initial F_1 transition ("weak" [deI]).

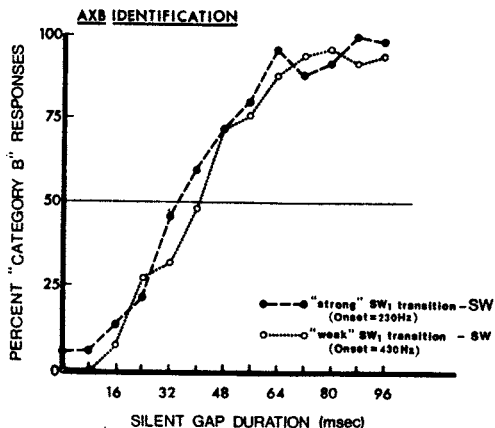
"SPECTRAL" LISTENERS



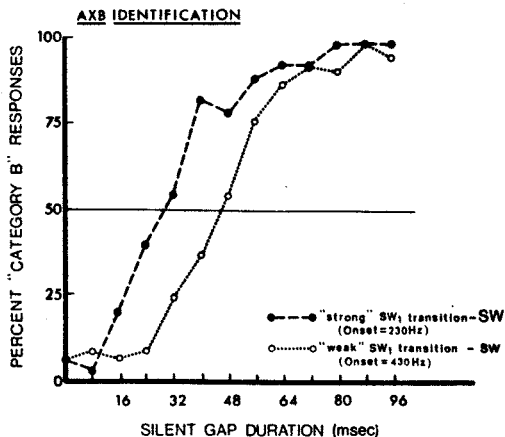
Figs. 4 and 5 are for those who perceived them as non-speech, divided, according to their reports of what the sounds were like, into those who were apparently attending to the transition cue (Figure 4, "Spectral" listeners) and those who were apparently attending to the silence cue (Fig. 5, "temporal" listeners). Fig. 6 is for the subjects who perceived the sounds as speech ("say-stay" listeners)).

Figs. 4, 5 and 6. Effect of silent interval on "identification" of sine-wave analogues of "say-stay" stimuli. (Reprinted from (7) with permission.)

"TEMPORAL" LISTENERS



"SAY-STAY" LISTENERS



Given a perceptual set for speech, some listeners identify these sounds as [deI] and [eI], while others hear them as different non-speech chords. If a suitable patch of noise is placed immediately before these sounds, they can be heard as [sei]; if a sufficient silent interval is introduced between noise and sine waves, a so-called "speech" listener will hear [steI], and he will hear it with a shorter interval before "strong" [deI] than before "weak" [deI].

On this basis, Best et al. constructed two continua, analogous to those of the earlier experiments, varying silent interval in combination with one or other of the [deI] syllables. To obtain identification functions without an explicit request for identification, they used an AXB procedure. In this procedure A and B are endpoints of a synthetic continuum, X a variable item from the continuum, to be judged on each trial as "more like A" or "more like B." Thus, despite the bizarre quality of their stimuli, Best and her colleagues obtained identification functions and assessed the perceptual equivalence of silence and formant transitions in a manner analogous to that of the earlier [sIIt - splIt] studies.

Their fifteen listeners divided themselves neatly into three groups of five. The results are displayed in Figures 4, 5 and 6. One group (the "spectral" listeners of Figure 4), though not entirely insensitive to the temporal variations, tended to hear a pattern with the short F_1 transition as more like the [sei] end of the continuum and a pattern with the long F_1 transition as more like the [steI] end of the continuum. A second group (the "temporal" listeners of Figure 5) disregarded spectral differences and based its judgments on the duration of the silent interval. Only the five "speech" listeners of Figure 6 were sensitive to the integrated auditory pattern and displayed the expected perceptual equivalence of silence duration and spectral change.

Evidently, one aspect of phonetic function is a capacity either to integrate or to make use of an auditorily integrated pattern of, temporally and spectrally distributed acoustic properties, to arrive at a unified phonetic percept.

ONTOGENETIC ENGAGEMENT AND DISENGAGEMENT OF THE PHONETIC MODE

Let us turn now to the development of the phonetic mode. I will describe a small set of interconnecting studies that converge on a view of phonetic perception as a specialized mechanism of attention that normally begins to be differentiated from general auditory perception during the first year of life.

We begin with the well-known observation that native speakers of Japanese typically cannot discriminate English [r] from English [l]. In fact, Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura (19) formally demonstrated this fact in an experimental test. Subsequently Eimas (5) showed that 4-6 month old, prelinguistic English infants discriminated between, but not within, the [r] and [l] categories very much as English-speaking adults do. Although comparable tests have not, so far as I know, been done on Japanese infants, presumably they would behave like the English infants. We might then reasonably infer that learning the sound system of a language involves, among other things, not only learning to tell the difference between sounds that are put to linguistic use, but also learning to disregard differences between sounds that are not.

In fact, Werker and her colleagues have evidence of precisely this (30). Their initial finding was that 7-month old Canadian English infants, tested in a head-turning paradigm, could discriminate between natural, non-English contrasts in Hindi that English-speaking adults could not. Werker (29) followed this up by tracking the decline of discriminative capacity over the second half of the first year of life. She used a conditioned head-turning paradigm to test three groups of infants on two non-English sound contrasts: Hindi voiceless, unaspirated retroflex vs. dental stops and Thompson (Interior Salish, an American Indian language) voiced glottalized velar vs. glottalized uvular stops. On the Hindi contrast, the number of infants successfully discriminating were: 11/12 at 6-8 months, 8/12 at 8-10 months, 2/10 at 10-12 months; for the Thompson contrast the results were essentially the same. (An infant was only classified as having failed to discriminate, if it had successfully discriminated an English contrast both before and after the failure on a

non-English contrast.) Finally Werker (29) reports longitudinal data for six Canadian-English infants on the same two non-English contrasts. All six discriminated both contrasts at 6-8 months, but by 10-12 months none of them could make the discriminations. By contrast, the one Thompson and two Hindi infants so far tested at 10-12 months could all make the called-for discriminations in their own language.

Is infant loss of discriminative capacity a permanent loss, followed by neural atrophy, or can it be reversed? MacKain, Best and Strange (14) addressed this question in a study of native Japanese speakers. They constructed a [rak-lak] series on the Haskins Laboratories ØVE IIIc synthesizer (see Figure 7). They then tested 10 American English speakers and 12 native Japanese speakers, divided into two groups on the basis of their experience with English. On the average, experienced Japanese speakers (n=5) and inexperienced Japanese speakers (n=7), respectively, had been spending either more than half of their time for over two years or about a quarter of their time for less than one year speaking English, and were taking either eight hours or less than one hour of English instruction a week. Figure 8 summarizes the results of an identification test and two types of discrimination test. The native American English speakers display standard categorical perception with consistent identification functions and discrimination peaks at the category boundary. By contrast, the inexperienced Japanese listeners display more-or-less chance performance across the board, while the experienced Japanese identify the syllables very much as the American English do, and display moderate discrimination peaks at the category boundary.

The final experiment in the series is an ongoing study by Catherine Best and myself, intended to assess how far either auditory or phonetic perception of the same acoustic pattern can be induced by instructional set. We constructed a sine-wave three-formant analog continuum, modeled on a [ra-la] series (similar to that of Figure 7), by varying only the sine-wave analog of F₃ from a steady pattern [la] to an upglide [ra]. One group of subjects (the music group) was informed that they would hear computer-made violin-like sounds and was given several examples of the isolated, endpoint F₃ analogs to be identified as "steady" or "upglide." Then, in a training series, the lower two endpoint sine waves

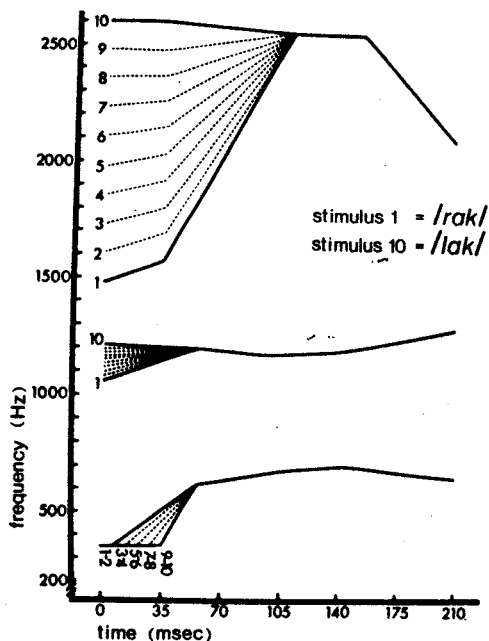


Fig. 7. Schematic spectrogram representations of the 10 stimuli in the synthetic [rak] - [lak] series. (Reprinted from (14) with permission).

SUMMARY OF RESULTS

- Americans (10)
- Japanese-experienced (5)
- Japanese -NOT experienced (7)

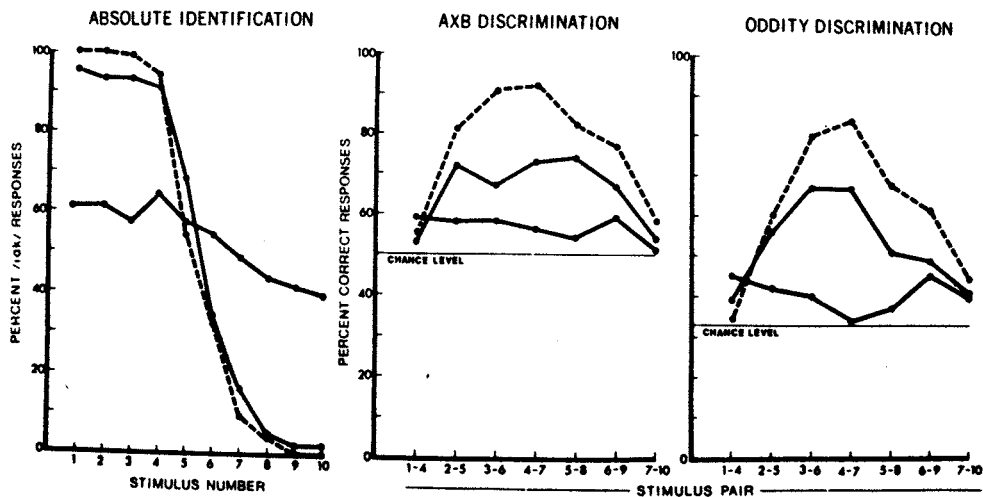


Fig. 8. Comparison of Americans, Experienced Japanese and Not-experienced Japanese results on three tests. (Reprinted from (14) with permission.)

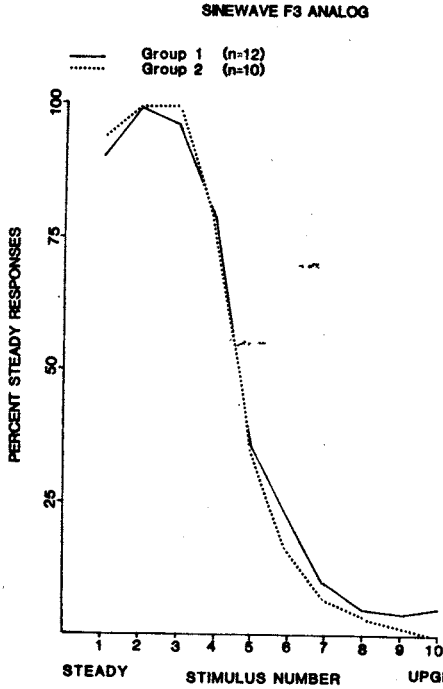


Fig. 9. Identification functions for speech listeners (Group 1) and violin listeners (Group 2), classifying the sinewave F_3 analog as steady or rising.

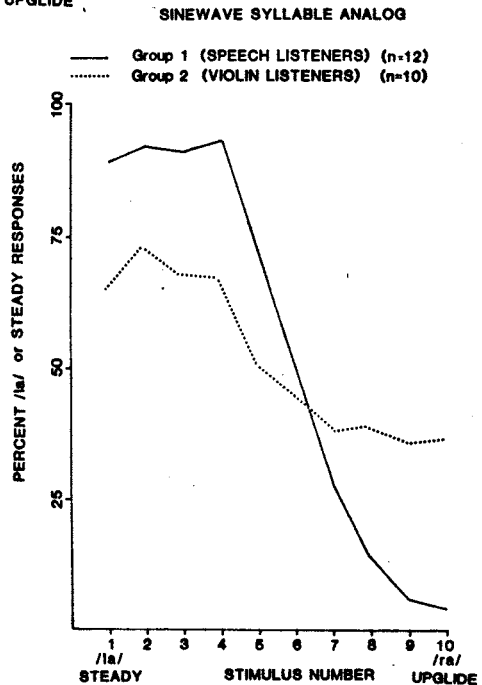


Fig. 10. Identification functions for the two groups of listeners, classifying the whole three-sine-wave patterns as [la] vs. [ra] (speech listeners) or as steady vs. upglide (violin listeners).

were gradually added to the upper endpoint waves until subjects were hearing the full chord. Finally, subjects were tested on the whole series with instructions to listen for the upper element of each musical chord and to identify it as either "steady" or "upglide". By contrast, the speech group was first played full, synthetic ØVE IIIC endpoint patterns to identify as [la] or [ra]. Then, in a training series, the synthetic formants were gradually faded out, while the sine-wave chord was faded in. Finally, subjects were tested on the whole series with instructions to identify each sine-wave syllable as [ra] or [la]. Subsequently, both groups were tested on the series of isolated F_3 analogs and instructed to identify them as "steady" or "upglide".

Figure 9 displays the results of this last test. Both groups classify the isolated F_3 sine-wave analogs quite consistently, with a 50% boundary around stimulus 5, close to the midpoint of the series. Figure 10 displays the results of the full-chord test. The speech listeners are quite consistent, with a 50% boundary shifted away from the midpoint of the series toward stimulus 6. The violin listeners, on the other hand, give a sloping function close to chance (similar, in fact, to the identification function of the inexperienced Japanese listeners in the left panel of Figure 8). Thus, although both groups of listeners were well able to classify the varying F_3 analog in isolation, only the speech listeners were able to use the F_3 sine-wave information to classify the full pattern.

Evidently, the phonetic mode begins to be engaged for some sounds and disengaged for others during the second six months of life. Subsequently, with appropriate acoustic materials and instruction it may be reengaged for certain sounds on which discrimination has been lost, or disengaged for certain other sounds on which discrimination has been retained.

A POSSIBLE BIOLOGICAL FUNCTION OF THE PHONETIC MODE

The distinction between auditory and phonetic perception is not new (10) and by now has substantial support from studies of hemispheric specialization in both normal and pathological populations. Some years ago, Donald Shankweiler and I concluded from the results of a dichotic

study of normals that "...while the general auditory system common to both hemispheres is equipped to extract the auditory parameters of a speech signal, the dominant hemisphere may be specialized for the extraction of linguistic features from those parameters" (26, p. 579).

Subsequently, Levy (9) and Zaidel (31,32) reached a similar conclusion on the basis of studies of split-brain patients, individuals whose two cerebral hemispheres have been surgically separated for control of epilepsy. Levy showed that only the left hemisphere can carry out the phonological analysis needed to recognize written rhymes; Zaidel showed that, while the right hemisphere may have a sizeable auditory and visual lexicon, only the left hemisphere can carry out the auditory-phonetic analysis necessary to identify synthetic nonsense syllables or the phonological analysis necessary to read new words. Moreover, dissociation of auditory discrimination and phonetic classification is commonplace among certain types of aphasic patients, suffering from left hemisphere lesions. For example, Blumstein, Cooper, Zurif and Caramazza (3), and Riedel (23) have described patients who could discriminate between syllables differing in the acoustic correlates of voice onset time or in medial vowel duration, but could not use these acoustic properties to classify the syllables phonetically.

We may gain some insight into the biological function of this dissociation, if we remember that the single most securely established specialization of the left hemisphere in over 90% of the population is for motor control of the right hand and of the speech apparatus. Whatever the perceptual capacities of the right hemisphere, studies of aphasics (17), of split-brain patients (32) and of the effects of sodium amytal injection (18; Borchgrevink, this volume) agree that the right hemisphere is essentially mute.

A plausible hypothesis then is that phonetic processes emerged phylogenetically, within the hominid line, in close association with the motor processes of the left hemisphere. The phonetic percept is then an abstract, sensorimotor link between the sounds of speech and the motor controls necessary to reproduce them. One source of ethological support for this hypothesis comes from the phenomenon of dialect. Every child,

uninstructed and without effort, learns the dialect of his peers. This seems to call for, on the one hand, recognition of subtle details of temporal and spectral structure of the kind we have been discussing and, on the other, the discovery of precise patterns of articulator placement and interarticulator timing necessary to reproduce the perceived dialectal variants. Whatever the ultimate adaptive value of dialect--whether selective breeding or group cohesion--we can hardly doubt its biological status nor the need for the specialized sensorimotor processes that it entails.

Finally, we should not shrink from viewing man's capacities for speech and language as both central and without biological parallel. We would not have come to understand how the barn owl tracks its prey from study of its peripheral auditory input alone. Nor would we expect to advance our knowledge of, say, echolocation in the bat solely by studying the auditory neural responses to bat sonar signals of chin-chillas. We can, of course, learn much of interest about the common properties of different mammalian auditory systems that may guide us in our study of speech. But, in the end, we must recognize that speech is peculiar to man. The purpose of Darwin's great work was not, after all, to demonstrate that all animals are the same, but to explain why they are not.

ACKNOWLEDGMENTS

The arguments developed in this paper rest squarely on the work of others, mostly of my colleagues at Haskins Laboratories. I thank them all, but absolve them from responsibility for the conclusions. I am particularly grateful to Alvin Liberman, Ignatius Mattingly and Bruno Repp for invigorating discussions. Writing of the paper was supported in part by Grant NICHD 01994 to Haskins Laboratories.

REFERENCES

1. Bailey, P.J. and Summerfield, Q. (1980) Information in speech: Observations on the perception of [s] + stop clusters. *J. Experimental Psychology: Human Perception and Performance*, 6, 536.
2. Best, C.T., Morrongoiello, B. and Robson, R. (1981) Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, 29, 191.

3. Blumstein, S.E., Cooper, W.E., Zurif, E.B. and Caramazza, A. (1977) The perception and production of voice onset time in aphasia. *Neuropsychologia*, 15, 371.
4. Bullock, T.H. (1982) Where do we go from here: need for cross-taxa comparisons. Paper read at NRP Associates Symposium on Ethological approaches to behavioral and neural plasticity, Boston, March 14-17.
5. Eimas, P. (1975) Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception and Psychophysics*, 18, 341.
6. Fant, C.G.M. (1962) Descriptive analysis of the acoustic aspects of speech. *Logos*, 5, 3r
7. Fitch, H.L., Halwes, T., Erickson, D.M. and Liberman, A.M. (1980) Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception and Psychophysics*, 27, 343.
8. Konishi, M. (1973) How the owl tracks its prey. *American Scientist*, 61, 414.
9. Levy, J. (1974) Psychobiological implications of bilateral symmetry. In Dimond, S.J. and Beaumont, J.G. (eds.) *Hemispheric function in the human brain*. Elek, London, pp.121-183.
10. Liberman, A.M. (1970) The grammars of speech and language. *Cognitive Psychology*, 1, 301.
11. Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review*, 74, 341.
12. Liberman, A.M., Isenberg, D. and Rakerd, B. (1981) Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception and Psychophysics*, 30, 133.
13. Liberman, A.M. and Studdert-Kennedy, M. Phonetic perception. In Held, R., Leibowitz, H. and Teuber, H.-L. (eds.) *Handbook of Sensory Physiology*. Vol. 8: Perception, Springer Verlag, New York, pp. 143-178.
14. MacKain, K.S., Best, C.T. and Strange, W. (1981) Categorical perception of English [r] and [l] by Japanese bilinguals. *Psycholinguistics*, 2, 369.
15. Mann, V.A., Madden, J., Russell, J.M. and Liberman, A.M. (1981) Further investigation into the influence of preceding liquids on stop consonant perception. *J. Acoustical Society of America*, 69, 591.
16. McGurk, H. and MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, 264, 746.
17. Milner, B. (1974) Hemispheric specialization: scope and limits. In Schmitt, F.O. and Worden, F.G. (eds.) *The Neurosciences: Third study program*. M.I.T. Press, Cambridge, pp. 75-89.

18. Milner, B., Branch, C. and Rasmussen, T. (1964) Observations on cerebral dominance. In DeReuck, V.S. and O'Connor, M. (eds.) Disorders of Language (Ciba Foundation Symposium). J. and A. Churchill, London, pp. 200-214.
19. Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J. J. and Fujimura, O. (1975) An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 19, 331.
20. Rand, T.C. (1974) Dichotic release from masking for speech. *J. Acoustical Society of America*, 55, 678.
21. Remez, R.E., Rubin, P.E., Pisoni, D.B. and Carrell, T.D. (1981) Speech perception without traditional speech cues. *Science*, 212, 947.
22. Repp, B.H., Liberman, A.M., Eccardt, T. and Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *J. Experimental Psychology: Human Perception and Performance*, 4, 621.
23. Riedel, K. (1982) Durational factors in the phonetic perception of aphasics. Unpublished doctoral dissertation, City University of New York.
24. Roberts, M. and Summerfield, Q. (1981) Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics*, 30, 309.
25. Studdert-Kennedy, M. (1974) The Perception of Speech. In Sebeok, T. A. (ed.) *Current Trends in Linguistics*, Vol. 12. Mouton, The Hague, pp. 2349-2485.
26. Studdert-Kennedy, M. and Shankweiler, D.P. (1970) Hemispheric Specialization for Speech Perception. *J. Acoustical Society of America*, 48, 579.
27. Summerfield, Q. (1979) Use of visual information for phonetic perception. *Phonetica*, 36, 314.
28. von Uexküll, J. (1957) A stroll through the worlds of animals and men. Reprinted in Schiller, C.H. (ed.) *Instinctive Behavior*. International Universities Press, New York, pp. 5-80.
29. Werker, J.F. (1982) The development of cross-language speech perception: the effect of age, experience and context on perceptual organization. Unpublished doctoral dissertation, University of British Columbia, Vancouver, B.C.
30. Werker, J.F., Gilbert, J.H.V., Humphrey, K. and Tees, R.C. (1981) Developmental aspects of cross-language speech perception. *Child Development*, 52, 349.
31. Zaidel, E. (1976) Language, dichotic listening and the disconnected hemispheres. In Walter, D.O., Rogers, L. and Finzi-Fried, J.M. (eds.) *Conference on Human Brain Function*. Brain Information Service/BRI Publications Office, UCLA, Los Angeles.
32. Zaidel, E. (1978) Lexical organization in the right hemisphere. In Buser, P.A. and Rougent Buser, A. (eds.) *Cerebral correlates of conscious experience*. Elsevier/North-Holland, Amsterdam, pp. 177-197.