863

Journal of Phonetics (1993) **21**, 147–155

Discovering phonetic function

Michael Studdert-Kennedy

Haskins Laboratories, New Haven, CT 06511, U.S.A.

The paper considers two topics on which one or other of the target papers disagrees, explicitly or implicitly, with one or other of its companion papers, and tries to resolve the differences. The topics are: (i) "categorical-like" perception of phonetic contrasts by infants, as evidence of "innate" phonetic capacity; (ii) units of perception, and the relations between perception and production.

1. Introduction

The four target papers form a fair sample of the styles of work now going on in studies of speech development over the first one to two years of life. The papers are diverse in focus, scope, data base and theoretical inclination, so that common themes are not easy to isolate. I have therefore chosen two topics that seem to me important, on which one or other of the papers, implicitly or explicitly, disagrees with its peers, and have tried to assess their several positions.

2. Phonetic function: innate or emergent?

The first point of difference concerns the nature of the infant's "initial state", its perceptual capacities at or soon after birth. Is the child born with a capacity to discriminate the phonetic properties of a spoken syllable, or are its early capacities purely auditory? Are syllables, segments, features given at birth, or do they emerge gradually over the first year of life by perceptual differentiation from larger prosodic structures? The conflict is most clearly drawn between JUSCZYK who assumes the latter, and WERKER & POLKA who assume the former. SUOMI and VIHMAN do not explicitly address the "initial state", but clearly see phonological structure as emergent rather than innate, and this is my view also. The question is of interest because it bears on the interpretation of the important work of Werker herself and her colleagues, reviewed by WERKER & POLKA. Do these findings reflect a "reorganization", between six to eight and 10–12 months of age, of infants' phonetic abilities, as WERKER & POLKA propose (p. 87)? Or do they rather reflect a shift in perceptual process from non-linguistic (auditory) to linguistic (phonetic) function, as I assume JUSCZYK would argue?

Our answers to these questions will depend on how we interpret infants' "categorical-like" perception of native and non-native phonetic contrasts during the first six months of life. WERKER & POLKA apparently concur with Eimas (1975) who concluded from a review of the evidence that categorical perception reflects the operation of innate mechanisms, specialized for processing speech and that "...

0095-4470/93/01147 + 09 $08.00/0

these early categories serve as the basis for future phonetic categories" (p. 342). WERKER & POLKA speculate indeed that among the various phonetic classes "... it might be the case that stop consonants are perceived in the most categorical-like manner, and are the most innately specified, whereas vowels are perceived in the least categorical manner and are most directed by experiential influences ..." (p. 97).

This statement reflects a misunderstanding of the conceptual status of categorical perception and an apparent ignorance of related studies of dichotic listening and short-term memory in which analogous stop consonant–vowel differences have been observed. A fair conclusion from over a dozen studies in these diverse paradigms is that the differences between stop consonants and vowels are due not to differences in phonological status or function, but to a difference in acoustic energy (the product of intensity and time)—a difference which WERKER & POLKA themselves remark ("... vowels ... have more prominent and long-lasting acoustic cues", p. 89). The higher energy of vowels than of voiced stop consonants makes them more discriminable (less easily confused). If we reduce their energy, or select a stimulus set from closer neighbors in acoustic space, vowels tend to induce consonant-like perceptual functions. By the same token, although we cannot easily increase the energy of stop consonants, we can sensitize listeners to their differences by an appropriate change of paradigm (e.g., Pisoni & Tash, 1974, cited by JUSCZYK), and we can select a consonant stimulus set more widely spaced acoustically than voiced stops. When we do either of these things, consonants tend to induce vowel-like perceptual functions. Let us briefly review the evidence for these claims. (For a fuller review than is possible here, see Studdert-Kennedy, 1975.)

Categorical perception, first reported for synthetic speech by Liberman, Harris, Hoffman & Griffith (1957) was formally defined by Studdert-Kennedy, Liberman, Harris & Cooper (1970) as a mode in which subjects can discriminate only between stimuli that they identify differently, giving rise to characteristic differences in ABX discrimination functions for synthetic continua of consonants and vowels: for consonants, high peaks between categories and troughs within categories, for vowels, relatively flat functions with low peaks between categories. Studdert-Kennedy *et al.* (1970) attributed these differences between consonants and vowels to supposed differences in their degree of "encoding" in the speech sound stream. Fujisaki & Kawashima (1970) proposed instead that the differences were simply due to the vowels' greater acoustic energy and the resulting longer duration of their traces in short term memory. Fujisaki & Kawashima themselves presented little data. It was left to Pisoni (1973) to manipulate vowel duration, systematically testing and largely confirming Fujisaki & Kawashima's account. WERKER & POLKA indeed cite Pisoni (1973), and several other relevant studies, demonstrating that shortened vowels are more categorically perceived than longer ones, but because even the shortened vowels (Pisoni's were 40 ms in duration) were somewhat more continuously perceived than stop consonants, WERKER & POLKA assume that a real consonant-vowel difference remains. My own view is that short enough vowels—a single glottal pulse can often be correctly identified—would be as "categorically" perceived as stop consonants. Such an experiment has not been done, so far as I know. But results from two other experimental paradigms encourage this prediction.

First are the results of dichotic studies, specifically designed to test for possible lateral differences in the hemispheric processing of consonants and vowels

(Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970). These studies of synthetic CV and naturally spoken CVC syllables demonstrated significant right ear advantages for initial consonants and non-significant right ear advantages for vowels, again inviting the interpretation of a possible difference in the "encoded" status of the two classes (e.g., Crowder, 1971). However, performance was significantly higher (in fact, close to ceiling) for vowels than for consonants on both ears, hinting that vowels were less subject to dichotic competition than consonants, so that a larger right ear advantage might be induced by a reduction in their effective acoustic energy. Godfrey (1974) confirmed this hypothesis, demonstrating a significant right ear advantage for vowels either reduced in duration or presented in noise. He also demonstrated a significantly greater right ear advantage for a set of four lax vowels, close neighbors in acoustic space, than for a set of four more widely distributed tense vowels.

Studies of immediate ordered recall of spoken word lists tell a similar story. A typical finding for lists of eight or nine random digits is that the last item is significantly better recalled than those immediately before it (recency effect). Crowder & Morton (1969) attributed the effect to an echoic "precategorical acoustic store" (PAS), and they showed that a redundant spoken suffix at the end of the list interfered with the store, significantly reducing the recency advantage (suffix effect). Extending the paradigm to lists of stop-vowel syllables that varied in either stop consonant (/b, d, g/) or vowel, Crowder (1971) found the standard recency and suffix effects for vowel lists, but neither effect for stop consonant lists. He concluded that vowels were represented in PAS, but stop consonants were not. Liberman, Mattingly & Turvey (1972) accepted Crowder's interpretation, arguing further that phonetic coding of the stops "strips away all auditory information", terminating auditory display. Darwin & Baddeley (1974) scotched this conclusion, however, by demonstrating: (i) a recency effect for /ga/, when the other members of the set in the lists to be recalled were acoustically distinct, /ʃa, ma/; (ii) a significantly reduced recency effect for vowels that were both very brief (30 ms steady-state in a 60 ms CV syllable) and close neighbors on an F_1–F_2 plot.

In short, three independent experimental paradigms initially invited the hypothesis that differences in perception of, or memory for, consonants and vowels reflected specialized phonetic or phonological processes. In each case, tests of the hypothesis revealed that the phonological status of the stimulus materials was irrelevant, and that the observed phenomenon could be comfortably drawn within the purview of standard auditory psychophysics. We should not infer from this that speech perception is a mere matter of auditory psychophysics. Nor should we conclude that consonants and vowels do not differ in phonological function, for they surely do, as WERKER & POLKA themselves suggest. Indeed, it is precisely from their differences in acoustic energy (and so perceptual salience) that their functional differences arise (Studdert-Kennedy, 1975, pp. 119, 120). What we *should* conclude is that the infant's "categorical-like" perceptual functions in its early months are evidence not of an "innate" specialization for phonetic perception, but of its mammalian auditory heritage.

Several consequences follow. First, we can be confident that the phenomena discovered by Werker and her colleagues do not reflect "reorganization" of the infant's phonetic capacities, because, on the available evidence, the infant at birth has no phonetic capacities. Second, by relinquishing the assumption that infants are born with specialized "language universal" or "language general" phonetic capacities, we are relieved of the puzzle as to how capacities, never characteristic of any individual speaker, could have evolved. Third, we resolve the troubling discrepancy between the supposedly unlimited capacities of the infant at birth and the decidedly limited capacities of older children (cf. Studdert-Kennedy, 1986). Finally, we can bring the discovery of apparent shifts in the perception of an indefinitely large number of non-native phonetic contrasts into line with the work of JUSCZYK and his colleagues: the capacity to perceive the segmental structure of a syllable is not the start, but the end of a long process of gradual differentiation from larger prosodic structures through clauses to words and segments (cf. Studdert-Kennedy, 1991*a*).

In this light we can perhaps understand WERKER & POLKA's discovery as a reflection of the same process that JUSCZYK and his colleagues have revealed by demonstrating the development of a preference for words having the phonotactic structure of the ambient language. The infant who distinguishes a non-native contrast at six months, but not at 12, has neither lost nor reorganized a phonetic capacity. Rather it has become attuned to, attentive to, interested in, and able to recognize at least some of the phonetic categories of its native language (cf. Best, in press). That this process should begin earlier with vowels than with consonants is (as WERKER & POLKA suggest) very much what the salience and early communicative function of vocalic melody would lead us to expect.

3. Relations between perception and production

A second area of difference among the four papers concerns the proposed units of perception, and the role, explicitly stated or implicitly assumed, that these units play in speech production. The papers differ not only in their substantive proposals, but in the extent to which they address the issue. JUSCZYK has little to say, and WERKER & POLKA, despite their call for attention to the child's changing functional requirements, have even less. Yet a central function of perception in the infant is surely to guide production: by learning to listen the child learns to speak. SUOMI clearly acknowledges this fact in his proposed "auditory-to-motor conversion", as does VIHMAN in her concept of an "articulatory filter" on perception.

In the following paragraphs I will first briefly sketch certain constraints that studies of audio-visual speech perception evidently place on accounts of how speech is perceived. I will then consider how well the accounts of WERKER & POLKA, JUSCZYK, SUOMI and VIHMAN seem to meet these constraints.

3.1. *Audio-visual speech perception*

The importance of lipreading for the hard of hearing and its utility for normal listeners in noisy environments have long been known. But the theoretical implications of the skill have only begun to emerge from experiments in several independent paradigms over the past 10–15 years. The burden of these experiments is that the two sources, acoustic and optic, specify the same physical events, and that observers integrate auditory and visual information "... before phonetic or lexical categorization takes place; the two streams are analogue at their conflux ..." (Summerfield, 1987, p. 16). The latter statement implies that observers integrate auditory and visual information in a continuous, common metric, and it raises the question of the nature of this metric.

I have no space to review here the extensive evidence for precategorical audio-visual integration (for which see Summerfield's lucidly argued chapter). I simply draw attention to two remarkable examples. First is evidence that the rate at which observers can correctly track passages of connected discourse by watching a talker's lips more than doubles if they are able to hear, in addition, a synchronized frequency pulse train picked up from the talker's larynx (Rosen, Fourcin & Moore, 1981). The pulse train itself carries no segmental information, only the talker's pitch, conveying intonation, stress and voice timing. All the segmental information comes from the lips. Yet experienced observers "... report a surprisingly complete degree of integration. Subjectively the pulse train ceases to sound like a buzz; it acquires vowel color and other acoustical attributes" (Summerfield, 1987, p. 16, fn. 3). Similar impressions of actually hearing what observers know themselves to have seen are reported for the well-known "McGurk effect" (McGurk & MacDonald, 1976).

A second example comes from recent work on the recency and suffix effects, in verbal recall, described above. Originally, these effects were taken to be peculiar to audition, because they did not occur for graphically presented word lists. New work has revealed, however, that the effects are cross-modal and peculiar to speech. This work has shown that: (i) lists seen (lipread) but not heard give rise to a recency effect (Campbell & Dodd, 1980); (ii) a spoken suffix, heard but not seen, reduces a lipread recency effect (Campbell & Dodd, 1980) and *vice versa* (Spoehr & Corin, 1978); (iii) an ambiguous acoustic suffix that can be heard as either a spoken syllable or a nonspeech sound reduces an auditory recency effect, if listeners take it as speech, but not if they take it as a musical sound (Ayres, Jonides, Reitman, Egan & Howard, 1979), or as an animal sound (Surprenant, Neath & Crowder, 1991; Crowder & Surprenant, in press). Evidently, speech seen, but not heard, and speech heard, but not seen, share a common representation in immediate memory. Perhaps we should not be surprised by this since the optic signal that we read from the lips is shaped by the same physical events as the acoustic signal that we hear: the shared representation is structurally isomorphic with the talker's articulations.

Importantly, for our understanding of the development of perceptual function, infants at three to four months are already sensitive to structural correspondences between the acoustic and optic properties of some speech sounds: they recognize an optic-acoustic match for the vowels /a/ and /u/ and they imitate optic-acoustic pairs, if they are matched, but not if they are mismatched (Legerstee, 1990). Similar, or related, results have been reported for the vowels /a/ and /i/ in four- to five-month-old infants (Kuhl & Meltzoff, 1982) and for various CV disyllables in six- to seven-month-old infants (MacKain, Studdert-Kennedy, Spieker & Stern, 1983). Evidently, infants may begin to discover certain auditory-articulatory properties of speech even before they begin to babble (cf. VIHMAN's paper).

In short, as Summerfield (1987) has remarked: "... any comprehensive account of how speech is perceived should encompass audiovisual speech perception. The ability to see as well as hear has to be integral to the design, not merely a retro-fitted afterthought" (p. 47). We can say much the same about a comprehensive account of how infants perceive speech and how they learn to talk.

We should not expect these four papers to address the issue directly, since it was not part of their charge. But if sensitivity to articulatory structure is indeed crucial to infant perceptual function, we may reasonably ask how well the papers conform to this constraint.

3.2. *Werker & Polka*

WERKER & POLKA assume purely linguistic units, and so can hardly address the perceptuomotor link. Nonetheless, the work of Werker and her colleagues, reviewed by WERKER & POLKA, meshes neatly with evidence for changes in perceptual function during the second half of the first year, discussed by JUSCZYK and, particularly, by VIHMAN. With the onset of babbling, and with the discovery of correspondences between the sounds it hears and the sounds it makes, the infant begins to focus attention on the phonetic (articulatory) properties of native sounds. The change cannot be picked up directly in the infant's responses to these sounds because the change is not in discriminability, but in the grounds for discrimination (phonetic rather than auditory). Werker and her colleagues hit upon a subtle index of this change, namely, a loss of interest in speech sounds (though not in non-speech sounds) for which the child can find no corresponding pattern in its own productions. What I am proposing, then, is that the apparent shift in the infant's response to any one of the infinity of non-native contrasts to which it has not been exposed is, in fact, a side effect of a shift in perceptual function with respect to the infant's native sounds—a neat instance, perhaps, of the operation of VIHMAN's "articulatory filter" on perception (see below).

3.3. *Jusczyk*

JUSCZYK's approach, like WERKER & POLKA's, stems from studies of synthetic speech perception in which properties of the signal are manipulated without regard to whether they are or are not articulatorily possible. Not surprisingly, he evinces no interest in production. To be sure, unlike WERKER & POLKA, JUSCZYK sees the child's "initial state" as pre-linguistic. And, very properly in my view, he sees the hierarchy of linguistic units (prosodic patterns, words, segments) as gradually emerging from "global" auditory sketches of the utterances a child hears. Yet the origins of the "global" sketch are unclear. because JUSCZYK's model has no principles of perceptual organization. He posits (i) a set of "innate" spectral analyzers (reminiscent of a channel vocoder) that divide the acoustic signal into discrete features. and (ii) a "pattern extraction process" that integrates the features of a speech signal into the standard "cues" of terminal analog synthesis. derived by manipulation of the spectrogram: "...stop closures, nasal resonances, frication, glide-like transitions, vocalic elements and the like" (p. 14). Presumably, some further "pattern extraction process" integrates these cues into larger units, but JUSCZYK says nothing about it. He does not justify the assumption of channels of analysis, or explain how a more or less continuous signal is divided into temporally discrete features. Nor does he describe the principles by which patterns are formed and cohere. "Pattern extraction" seems to have been conceived *ad hoc* as a process that will get the model, by fiat, from one level of description to the next. Certainly, it is not always invoked where it might seem to be necessary: for example, JUSCZYK evidently regards syllables as "basic units of processing" (p. 6), for which no pattern extraction is needed.

In this I believe he is correct. For the syllable is precisely the point at which the child gains access to phonetic structure, and to the essential isomorphism of articulatory and auditory form: rhythmic oscillation of the mandible in babbling and early speech gives rise to rhythmic trajectories of amplitude and frequency through acoustic space. Learning to talk is, in the first instance, a matter of discovering how variation in the articulatory structure of the syllable corresponds to variation in its auditory form, and *vice versa.* If JUSCZYK were to shed the largely programmatic "front end" of his model—essentially a pious nod to the history of cue-based speech synthesis—and start with the syllable, as an acoustic and articulatory integer, he might give us a child who can learn to talk.

3.4. *Suomi*

SUOMI fully engages the question of the perceptuomotor link. He adopts the gesture of Browman & Goldstein's (e.g., 1990) Articulatory Phonology as the motor primitive from which words are built, a solution to the puzzle of how we are to align perception and production for which I have argued elsewhere (Studdert-Kennedy, 1991*a, b*). He assumes (without committing himself as to its developmental origin) that a child has the capacity to "... directly imitate the detailed motor orchestrations that he has extracted from adult words" (p. 33). I take this statement to imply that, in SUOMI's view, the acoustic/auditory form of a word fully specifies (corresponds to, is isomorphic with) its articulatory form.

What puzzles me then is the distinction that SUOMI wants to draw between the input lexicon (a collection of "... continuous, phonetically unsegmented spectral prototypes" (p. 47)) and the output lexicon (a collection of word motor plans, each a "... blueprint of the gestures required to approximate the associated prototype" (p. 35)). Obviously, input and output are neurologically distinct, whether in locus or process or both, but they are formally identical. We have no reason to suppose that the acoustic specifications of a word are not identical whether we are shadowing the word (intending to repeat it immediately), holding it briefly in immediate memory, or storing it in the lexicon. If word motor plans are indeed implicit in the continuous spectral structure of every word (as SUOMI seems to assume in his auditory-to-motor conversion). nothing is to be gained by positing one lexicon. accessed by LAFS when we are simply listening, and a parallel lexicon of word motor plans when we hear a new word or are planning to speak.

Moreover, LAFS is not easily reconciled with the facts of lipreading. Summerfield (1987) has considered the problem of "making LAFS lipread". He proposed, for example, the possibility of specifying both LAFS sequences of spectral states and their corresponding two-dimensional (2-D) images of the talker's lips and teeth as three-dimensional (3-D) vocal tract configurations and compiling the three lists into a single lexicon (cf. Klatt, 1979, p. 306). Recognition would proceed by tracing the path through the network that yielded the best matches between 3-D vocal tract configurations, on the one hand, and both spectral states and 2-D facial configurations, on the other.

Such an undertaking might be of interest for engineers concerned with automatic speech recognition, but has little insight to offer the student of speech perception. The main objection to LAFS, whether purely auditory or auditory and visual, is that it operates by brute force compilation rather than by exploiting the lawful correspondences between speech sounds and articulation. We have already seen that infants at three to four months are sensitive to auditory-visual matches in spoken vowels, and that when matches (but not when mismatches) are present, infants attempt to imitate them. Presumably, they have not acquired this capacity by compilation.

3.5. *Vihman*

VIHMAN's paper contributes to the ethology of early speech: its data are observational rather than experimental, and therefore are primarily concerned with production. All the more striking then is the fact that she is drawn to posit close perception-production relations. Of particular interest here is VIHMAN's construct of an "articulatory filter".

As we have seen, infants already at three to four months can recognize certain auditory-articulatory correspondences in adult speech and may even, in the right experimental situation, be impelled to imitate them. But only with babbling and, later, with first words does the infant come into possession of an array of gestural scores, so that it can begin to recognize correspondences between these scores and the auditory patterns of its own and adult speech. From this capacity there then arises an "articulatory filter", operationally defined as the child's habit of selecting from the available adult repertoire some words to say, some to avoid, and some to make a stab at (the IN, OUT and ATTEMPTED of Schwartz, 1988).

VIHMAN's "filter", conceived to account for the facts of phonological selectivity, also bears on two other matters discussed above. First is the puzzling question of WERKER & POLKA's infants who cease to discriminate an infinity of sounds to which they have never been exposed. Perhaps we have here a parallel to the child's apparent "avoidance" of certain sounds in its early words. Avoidance is a difficult concept, because it seems to require that the child have a representation of what it chooses to avoid, so that avoidance should no longer be necessary. VIHMAN proposes a positive solution: the child chooses to attempt only those adult words for which it has a matching gestural score. The sounds it "avoids" are simply those selected because they do not yet have a match. Perhaps WERKER & POLKA's children, at 12 months and on the cusp of their first words, attend only to those speech contrasts for which they have at least a sketch of a matching score.

The second matter on which VIHMAN's "filter" bears is the hypothesis put forward in discussion of SUOMI's paper, namely, that the articulatory representation of a word is implicit in its auditory representation (and *vice versa*). If this is so, as the phenomenon of phonological selectivity in a child's early words seems to require, what functional gain would accrue from the emergence of a parallel representation, suitable for LAFS, from which articulatory specification had been excluded? So far as I can see there would be none, and this would invite a welcome simplification of SUOMI's otherwise elegant model.

Preparation of this paper was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

References

Ayres, T. J., Jonides, J., Reitman, J. S., Egan, J. C. & Howard, D. A. (1979) Differing suffix effects for the same physical suffix, *Journal of Experimental Psychology: Human Learning and Memory*, 5, 315–321.

Best, C. (in press) Emergence of language-specific constraints in perception of non-native speech. In *Developmental neurocognition: speech and face processing in the first year of life* (B. de Boysson-Bardies & S. de Schonen, editors). The Netherlands: Kluwer.

Browman, C. P. & Goldstein, L. (1990) Gestural specification using dynamically-defined articulatory structures, *Journal of Phonetics*, **18**, 299–320.

Campbell, R. & Dodd, B. (1980) Hearing by eye. *Quarterly Journal of Experimental Psychology*, **32**, 85–99.

Crowder, R. G. & Morton, J. (1969) Precategorical acoustic storage (PAS), *Perception and Psychophysics*, **5**, 365–373.

Crowder, R. G. (1971) The sound of vowels and consonants in immediate memory, *Journal of Verbal Learning and Verbal Behavior*, **10**, 587–596.

Crowder, R. G. & Surprenant, A. M. (in press) On the linguistic module in auditory memory. In *Language and literacy: comparative approaches* (de Gelder, B. & Morais, J.). Cambridge, MA: M.I.T. Press.

Darwin, C. J. & Baddeley, A. D. (1974) Acoustic memory and the perception of speech. *Cognitive Psychology*, **6**, 41–60.

Eimas, P. D. (1975) Auditory and phonetic coding of the cues for speech: Discrimination of the /r-l/ distinction by young infants. *Perception and Psychophysics*, **18**, 341–347.

Fujisaki, H. & Kawashima, T. (1970) Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, University of Tokyo*, **29**, 207–214.

Godfrey, J. J. (1974) Perceptual difficulty and the right ear advantage for vowels. *Brain and Language*, **1**, 323–335.

Klatt, D. H. (1979) Speech perception: a model of acoustic-phonetic analysis and lexical access, *Journal of Phonetics*, **7**, 279–312.

Kuhl, P. K. & Meltzoff, A. N. (1982) The bimodal perception of speech in infancy, *Science*, **216**, 1138–1141.

Legerstee, M. (1990) Infants use multimodal information to imitate speech sounds, *Infant Behavior and Development*, **13**, 343–354.

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries, *Journal of Experimental Psychology*, **54**, 358–368.

Liberman, A. M., Mattingly, I. G. & Turvey, M. T. (1972) Language codes and memory codes. In *Coding processed in human memory* (A. W. Melton & E. Martin, editors), pp. 307–334. Washington: V. H. Winston.

Mackain, K. S., Studdert-Kennedy, M., Spieker, S. & Stern, D. (1983) Infant intermodal speech perception is a left-hemisphere function, *Science*, **219**, 746–748.

McGurk, H. & MacDonald, J. W. (1976) Hearing lips and seeing voices, *Nature*, **264**, 746–748.

Pisoni, D. B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, **13**, 253–260.

Pisoni, D. B. & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories, *Perception and Psychophysics*, **15**, 285–290.

Rosen, S. M., Fourcin, A. J. & Moore, B. C. J. (1981) Voice pitch as an aid to lipreading, *Nature*, **291**, 150–152.

Schwartz, R. G. (1988) Phonological factors in early lexical acquisition. In *The emergent lexicon* (M. D. Smith & J. L. Locke, editors), pp. 185–222. New York: Academic Press.

Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears, *Quarterly Journal of Experimental Psychology*, **19**, 59–63.

Spoehr, K. T. & Corin, W. J. (1978) The stimulus suffix effect as a memory coding phenomenon, *Memory and Cognition*, **6**, 583–589.

Studdert-Kennedy, M. (1975) From continuous signal to discrete message: Syllable to phoneme. In *The role of speech in language* (J. F. Kavanagh & J. E. Cutting, editors) pp. 113–125. Cambridge, MA: MIT, Press.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In *Invariance and variability in speech processes* (J. S. Perkell & D. H. Klatt, editors), pp. 58–76. Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1991*a*) Language development from an evolutionary perspective. In *Biological and behavioral determinants of language development* (N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. Studdert-Kennedy, editors), pp. 5–28. Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1991*b*) Learning to speak: a note on the units of speech production and speech perception, *Perilus*, **XIV**, 173–183.

Studdert-Kennedy, M. & Shankweiler, D. P. (1970) Hemispheric specialization for speech perception, *Journal of the Acoustical Society of America*, **48**, 579–594.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S. & Cooper, F. S. (1970) Motor theory of speech perception, *Psychological Review*, **77**, 234–249.

Summerfield, A. Q. (1987) Some preliminaries to a comprehensive account of audio-visual speech perception. In *Hearing by eye: the psychology of lip-reading* (B. Dodd & R. Campbell, editors), pp. 3–51. Hillsdale, NJ: Lawrence Erlbaum Associates.

Surprenant, A. M., Neath, I. & Crowder, R. G. (1991) A context-dependent stimulus suffix effect. Paper presented at Psychonomic Society Meetings, San Francisco, November.