Animal models of speech perception phenomena

Andrew J. Lotto, Keith R. Kluender, and Lori L. Holt
University of Wisconsin-Madison

1. Introduction

Science, by tradition and foundational philosophy, is conservative. General theories that are broad in application and austere in constructs and variables are preferred to proliferation of specific explanations for individual phenomena. When general, simple accounts no longer suffice as explanations of the available data, novel constructs and theories become necessary. Fifty years ago, speech perception was considered to be explainable in terms of general psychoacoustics and learning theories. However, after the development of the first speech synthesizers, several phenomena were discovered which led many researchers to suggest that general processes would no longer suffice to explicate speech perception. Thus, processes specific to phonetic perception were proposed (for a history of these events see Liberman, 1996). It is our contention that these proposals were premature and that general perceptual and learning theories have not been exhausted as explanatory mechanisms for phenomena of speech perception. We intend to demonstrate in this paper that many of the results which have been argued to be indicative of species-specific processes specialized for phonetic perception are replicable with non-human animals and are, in turn, potentially explainable by several general theories and findings which were proffered in the first half of the twentieth century.

Proponents of a "speech-is-special" view have offered many examples of phenomena that are considered to be products of processes exclusive to speech perception. One way to falsify these claims is to show analogous results for non-speech acoustic stimuli. There have been many demonstrations of sine-wave tones, musical chords, noise bursts, slamming doors, etc. which elicit labeling behavior which had previously been described as specific to speech sounds (e.g. Blechner, 1977; Pisoni, 1977; Pisoni, Carrell, & Gans, 1983; Diehl & Walsh, 1989; Fowler & Rosenblum, 1990; Holt, Lotto, & Kluender, 1996; Lotto & Kluender, in press). These studies have made it clear that the proposition of speech-specific effects may be problematic. However, there are some concerns with experiments utilizing non-speech analogs. One is that it is difficult to describe which aspects of a speech sound must be retained in a non-speech sound in order to make a theoretically effective analog. Another concern is that non-speech sounds may be processed by sufficiently broad application of putative speech-specific processes (Kuhl, 1978; 1986a;b). Although this complicates the definition of "speech-specific", a reasonably good argument can be made for this possibility. Speech is probably the most ecologically important periodic (and quasi-periodic) signal for humans. If complex processes have evolved to perceive these signals, then one may expect that over-generalized application to non-speech sounds could be most adaptive. One would rather analyze a non-speech sound (with small ecological importance) in a non-veridical manner than not properly process an important speech signal (e.g. "Duck!").

Fortunately, there is another way to obtain evidence that falsifies claims of speech-specificity. No matter the details of the particular theory, demonstrations of putative "speech-specific" behavior by non-human animals is *prima facie* evidence against the assertions of "speech-is-special" theories. Animals are unlikely recipients of innate processes that are dedicated to sounds created by a human vocal tract. Analogous human and animal behaviors call for explanations with rather general application. We hope to demonstrate that, in fact, many "speech-specific" behaviors have been replicated with animal models and that, therefore, proposal of novel processes specific to the perception of phonetic strings are prodigal.

According to our reading, there are three classes of phenomena that have led to the proposal of and maintenance of theories of specialized speech processing: 1) the apparent lack of invariant acoustic attributes which map directly onto the perceived phonemic identity of speech sounds; 2) the symmetry between speech production and speech perception, e.g. the apparent perceptual compensation for acoustic effects of coarticulation; 3) the ability of human infants to discriminate and categorize speech sounds in a linguistically-relevant manner. These phenomena are often explained with reference to a proposed species-specific module which is dedicated to the perception of phonetic signals. We will present data from two avian species which address each of the speech-perception phenomena listed above and suggest some general theories which may account for both the human and animal results.

2. Lack of Invariance

Perhaps the phenomena most responsible for the development of "Motor Theory", the paradigmatic "speech-is-special" theory", is the lack of invariance between acoustics and linguistic categories (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Acoustic information specifying a particular phonetic segment varies dramatically with a change in the identity of surrounding segments. For example, the second-formant (F2) transition of a naturally-produced syllable-initial /d/ varies substantially in shape depending upon the identity of the following vowel. Preceding /i/, F2 increases in frequency. Preceding /u/, F2 decreases in frequency. Despite this difference, human listeners will label both as beginning with /d/ (Liberman, Delattre, Cooper, & Gerstman, 1954). While there has been some limited success in describing invariant acoustic attributes of speech sounds (e.g. Stevens & Blumstein, 1981; Sussman, McCaffrey, & Matthews, 1991), it is clear that for most phonetic categories variability of attributes is the norm. How is it that human listeners can so reliably label phonetic segments in the face of such variability?

“Motor Theory” proposes that a specialized module performs the inverse transform from acoustics to motor commands (which can be described in terms of phonemes) through reference to the mechanisms which caused the variability, i.e. vocal-tract gestures (Liberman & Mattingly, 1985).

Presumably, such a module would be specific to humans, since other species have little need to disentangle the variability arising specifically from human vocal tracts. Given this, is it possible for animals to “properly” label phonetic signals which contain the variability inherent to speech sounds? Kluender, Diehl, and Killeen (1987) answered this question using an avian species, the Japanese Quail (*Coturnix coturnix japonica*). Three female quail were trained to discriminate naturally-produced /dVs/ from /bVs/ and /gVs/ with the vowels /i/, /u/, /✍/, and /a/. The birds were reinforced for pecking a lighted key during repeated presentation of [dis], [dus], [d✍s], or [das] and were required to refrain from pecking to [bis], [bus], [b✍s], [bas], [gis], [gus], [g✍s], or [gas]. The birds eventually were able to discriminate the /d/ stimuli from non-/d/ stimuli. One may suggest that this is evidence that non-human animals can label stimuli in a phonetically-appropriate manner despite the lack of invariance between the acoustic signals and their responses. However, it is possible that the birds simply learned to peck to the four “positive” (/d/) stimuli individually and that they had not abstracted information about /d/ regardless of context. In order to determine whether the birds had learned a phonetic category one has to test the animals with novel (i.e. previously unpresented) syllables.

After the birds achieved asymptotic performance on the training stimuli, Kluender *et al*. presented two of the birds with stimuli containing eight novel vowels, /ɪ/, /ᴜ/, /✍/, /✍/, /e^y/, /o^w/, /o^y/, and /✍/. The difference in peck rates for /dVs/ stimuli versus /bVs/ and /gVs/ stimuli would indicate whether the birds had generalized from the training stimuli. Figure 1 displays a histogram of these data collapsed across the two birds’ responses.

It is clear from these data that the quail were able to respond differentially to the novel /dVs/ and the novel /bVs/ and /gVs/. This suggests that the birds were responding on the basis of the presence of /d/. What allowed the birds to make this response? Measurements of the acoustic stimuli used in the experiment did not reveal any obvious invariant attributes (e.g., spectral templates discussed by Stevens & Blumstein, 1981) that could have cued the bird to a proper response. The shape of the F2 transition varied from syllable to syllable in a manner similar to that reported in Liberman *et al*. (1954). Despite this variability the birds the /dVs/ stimuli as functionally equivalent.

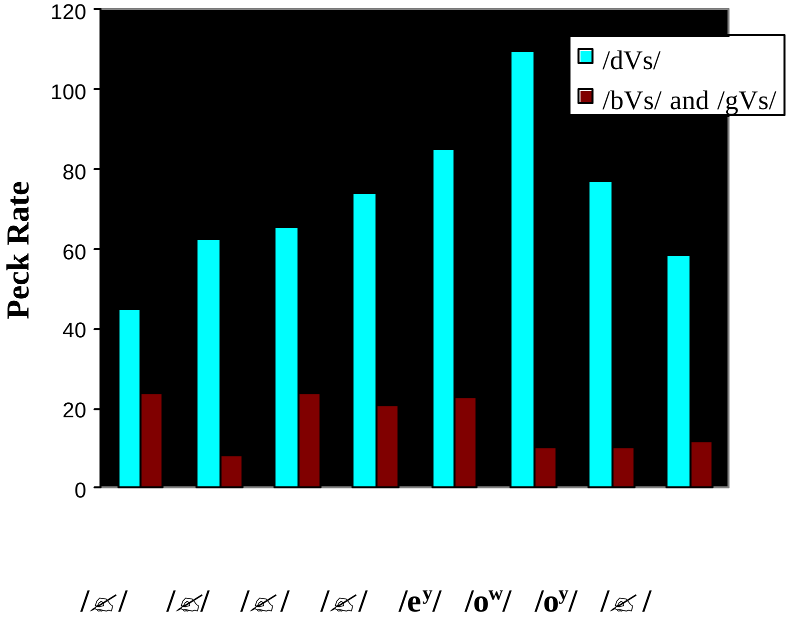

Figure 1. Peck Rates from Kluender *et al*. (1987) for novel stimuli. Collapsed across Bird 716 and Bird 730.

What does it mean “to respond to the presence of /d/” for an animal that presumably has no lexicon and has never been taught about the orthographic reference ‘d’ and when there are no invariant physical attributes which signal the presence of /d/? The birds must have extracted some information from the training stimulus on which they could base a response. We would suggest that birds (and humans) are not using any single reliable attribute to categorize the sound, but instead are relying on a combination of imperfectly-valid attributes which are related in some degree to the phonetic category. For example, F2-onset frequency, F2-transition shape, burst characteristics, third-formant-onset frequency, and steady-state formant frequency values are all informative for syllable-initial stop identification, but none of them alone can perfectly discriminate exemplars from different phonetic categories. Taken together, however, these various sources of information may be synthesized and respectable categorization ability can be achieved. This multiple-attribute approach is not specific to speech and is certainly not neoteric. Brunswick (1950; 1955) proposed a similar account for perceptual constancies called the “Lens Model”. According to Brunswick an object or event in the environment (distal stimulus) gives rise to an array of proximal attributes. These attributes tend not to be perfectly indicative of their source but instead are probabilistically related to the identity of the distal stimulus. These probabilities are referred to as the *ecological validity* of attributes (Brunswick, 1955; Postman & Tolman, 1959). These multiple proximal attributes are weighted by the perceiver and contribute to the final perception. No single attribute is responsible for the perception. Brunswick used the model to explain phenomena such as size constancy in vision, but the "Lens Model" applies equally well to speech phenomena. The gestures of a vocal tract (distal stimulus) give rise to an array of acoustic attributes each of which is probabilistically related to the intended phonetic string. By relying on all of these probabalistic cues (properly weighted) the perceiver, whether a bird or a human, can perceive the intended string. This description will remind some of connectionist modeling efforts in which inputs are differentially weighted and combined into the final output.

One prediction of a "Lens Model" approach is that the absence of one attribute can be offset by the presence of another, because of the summation of information from all of the attributes. This perceptual tradeoff has been demonstrated for many attributes associated with various phonetic categories. Repp (1982) provides a good review of these "phonetic trading relations" and suggests that these results are supportive of a phonetic mode. Repp argues that it *is* highly unlikely that these trading relations will arise from the general operating characteristics of the auditory system. This may be true, but it is highly likely that such trading relations will arise from general learning processes which weight multiple attributes differentially as in Brunswick's Lens Model. One would predict "phonetic trading relations" from general perceptual principles.

Of course, some of the phonetic trading relations still may be due to the general auditory system and, again, animal models can elucidate which phonetic attributes are related in this way. For example, one familiar trading relation is between first-formant (F1) frequency and voice-onset time (VOT) in identifying voiced versus voiceless stops in English. When F1-onset frequency is lower, a longer VOT is required for humans to label consonants as voiceless (Lisker, 1975; Summerfield & Haggard, 1977; Kluender, 1991). Kluender and Lotto (1994) trained Japanese Quail to peck to good exemplars of voiced (short VOT) or voiceless (long VOT) /Ca/ stimuli. They then varied F1-frequency for ambiguous novel stimuli (moderate VOT). Just as with human listeners, birds "labeled" more low-frequency F1 stimuli as voiced than for high-frequency F1 stimuli. This result was obtained despite the fact that the birds had no chance to learn the covariations between F1 frequency and VOT with phonetic categories. In this case, the trading relation must have arisen from some general aspect of the auditory system which is shared between humans and quail. Similarly, Kuhl and Miller (1978) showed that the shift in VOT boundary with place of articulation evidenced by human listeners is displayed in the responses of chinchilla. It is possible that all of the "phonetic trading relations" can be accounted for by a combination of general auditory processing and general multiple-attribute learning processes. Until an aspect of phonetic categorization is described which can't be accounted by these twin explanatory mechanisms, scientific virtues of theoretical asceticism compel us to reject speech-specific proposals.

3. Perceptual Compensation for Coarticulation

The second class of phenomena which purportedly support speech-specific theories is the apparent symmetry between perception and production. Articulatory dynamics and constraints shape the resultant waveform and, in many cases, perceptual processes of the listener act in ways that appear to respect these constraints and dynamic properties. This symmetry has been a motivating factor behind "speech-is-special" theories. For example, in the revised Motor Theory of Liberman and Mattingly (1985), symmetry is said to arise because of the shared currency of speech perception and production: gestural representations.

One case of symmetry between speech perception and production that has received a considerable amount of empirical attention is the effect of a preceding liquid on stop-consonant perception. Mann (1980) presented listeners with members of a series of synthesized consonant-vowel (CV) syllables varying perceptually from /da/ to /ga/ preceded by natural utterances of either /al/ or /ar/. Subjects identified the CVs as /ga/ more often following /al/ than following /ar/. This shift in response is complementary to the coarticulatory influences on CV production following /al/ and /ar/. Due to the assimilative nature of coarticulation, the place of vocal-tract occlusion of a stop consonant is more anterior following /al/ than following /ar/. Because the alveolar stop [d] is produced at an anterior place in the oral cavity, and the velar stop [g] is produced with a posterior occlusion, /al/ productions result in subsequent CVs being more [da]-like articulatorally and acoustically, whereas CVs following /ar/ are more [ga]-like. Thus, subjects' perceptual responses seem to compensate for acoustic effects of coarticulation; more /ga/ identifications result from CVs following /al/.

The perceptual effect is quite compelling and subsequent research has demonstrated labeling shifts for 4-month old infants (Fowler, Best, & McRoberts, 1990) and for Japanese listeners who could not discriminate /l/ and /r/ (Mann, 1986). This impressive evidence led Mann (1980) to claim that human listeners must be making "tacit reference to the dynamics of speech production" in order to compensate for the effects of coarticulation. Many people see these sorts of perceptual phenomena to be the most coercive argument in favor of specialized processes for speech perception. Certainly non-human animals have no mechanisms that could accommodate the acoustic effects of human vocal-tract coarticulation.

Lotto, Kluender, and Holt (in press) conducted an experiment to discover whether animals show such speech context effects. Two Japanese quail were trained to peck a key when presented good exemplars of the syllable /da/ and to refrain from pecking to good /ga/ stimuli. Two other quail were trained with /ga/ as the positive syllable and /da/ as the negative syllable. After reaching a preset criterion of 10:1 ratio of pecks to positive stimuli versus pecks to negative stimuli, the birds were presented with novel ambiguous CVs preceded by either /al/ or /ar/. The task was essentially the same as that used by Mann (1980) with human subjects. The birds 'labeled' the CVs as /da/ or /ga/. There was no chance for the birds to learn the covariation between the acoustics of the CV and the preceding context. Would there be any effect of the preceding liquid context as there was for human listeners? Figure 2 displays the peck rates for the novel stimuli in the /al/ and /ar/ contexts.

In fact, all four birds displayed a significant shift in peck rates across the change in preceding liquid. The two /da/-positive birds pecked substantially more to CVs preceded by /ar/ and the /ga/-positive birds pecked more to CVs preceded by /al/. This is the same shift that Mann (1980) showed for human listeners. Are Japanese quail perceptually compensating for the acoustic effects of coarticulation arising from human vocal tracts? Lotto *et al*. (in press) suggest that the similarities in human and animal responses to these speech stimuli are probably due to rather general properties of perceptual systems. In particular, these context effects may be described as examples of perceptual contrast.

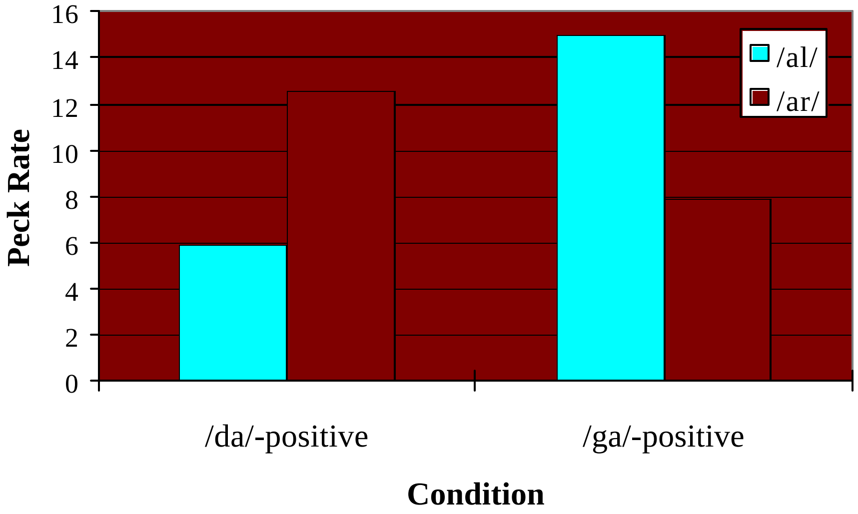

Figure 2. Peck rates for novel CVs from Lotto *et al*. (in press). Rates are collapsed across two birds in each condition.

If one considers the F3 frequencies for the CVs and the /al/ and /ar/ stimuli, one sees that the behavior of the animals (and humans) can be described as contrastive. Quail trained to peck to CVs with *low* F3-onset frequencies (/ga/-positive) pecked more to intermediate values of F3-onset frequency (novel ambiguous stimuli) when CVs were preceded by a syllable with a *high*-frequency F3-offset (/al/). Quails trained to peck to CVs with *high* F3-onset frequencies (/da/-positive) pecked more to intermediate values of F3-onset frequency when CVs were preceded by a syllable with a *low*-frequency F3-offset (/ar/). The bird data suggest that this contrastive behavior is a product of the auditory system and is not a pattern of behavior evolved in humans to deal specifically with coarticulation. One could refer to this general auditory behavior as *frequency contrast*. The effective frequency of a formant (F3 in this case) is shifted in a direction away from a preceding prominence of spectral energy.

Again, frequency contrast is not a new concept. Cathcart and Dawson (1928-1929) demonstrated shifts in responses to tones of differing frequencies which can be described as contrastive. Christman (1954) also described non-speech examples of frequency contrast. It is possible that this general perceptual contrast is responsible for the "perceptual compensation for coarticulation" evidenced by humans and Japanese quail. Surely such an explanation should be falsified before more extravagant proposals are made.

In a similar vein, perceptual contrast may account for other context effects in speech perception. For example, human identification boundaries for members of a stop/glide series varying in transition duration shift toward longer transitions (i.e. more *short*-transition responses) when syllable duration is increased (Miller & Liberman, 1979). It has been suggested that this is the result of a mechanism specialized to accommodate changes in speaking rate and its effect on temporal attributes of speech sounds. However, it may be that the listener is perceiving the transition duration as *shorter* in contrast to the *longer* overall syllable duration. This general perceptual account is bolstered by data from Stevens, Kuhl, and Padden (1988) that demonstrate a similar shift in responses by macaques to a stop/glide series.

4. Infant Speech Categorization

The final phenomenon that has been argued to be indicative of specialized processing is the ability of pre-lingual infants to categorize speech sounds in a linguistically-relevant manner. Kuhl (1983) demonstrated that six-month-old infants can differentiate two vowels even across changes in speaker gender. This is rather remarkable considering the substantial differences in fundamental frequency and formant frequencies between the sexes. However, this does not to appear to be a special human ability. Burdick and Miller (1975) showed that chinchillas can differentiate vowels despite changes in talker and overall amplitude.

Of more recent concern is the ability of infants to respond to sounds from the same phonetic category. Kuhl, Williams, Lacerda, Stevens, and Lindblom (1992) utilized a head-turn paradigm to test the degree to which infants treat instances of a vowel distribution equivalently. Six-month olds from Swedish- and English-speaking households exhibited quite different responses to Swedish- and English-typical vowel distributions. The results were easily related to the phonetic categories of the infants' native languages. That is, by six months of age, infants respond to speech sounds in a way relevant to their language experience. Some may view these results as remarkable because these infants probably have little to no lexicon and thus no idea of minimal contrasts.

Some researchers have placed themselves on the *cordon sanitaire* between the “speech-is-special” and “speech-is-general” factions by proposing that while auditory processes involved in speech perception may be general, human infants may have learning mechanisms specialized for the categorization of speech sounds. For example, Jusczyk (1993) has suggested that the development of speech perception capacities is due to “innately guided learning processes”. (Also see Kuhl, 1993 for another example of proposed specialized learning processes).

Is language-relevant categorization after at least six-months of ambient exposure to speech sounds remarkable enough to warrant discussion of innate specialized mechanisms? Are general learning mechanisms unable to accomplish such a feat? While there may be no definitive answer to these questions at this time, there are some suggestive data. Lotto, Kluender, and Holt (1995) report on a vowel-learning experiment involving the avian species European or Common starling (*Sturnus vulgaris*). In this experiment, birds were trained to discriminate members of the English vowel categories /i/ and /✍/ or two Swedish vowel categories. Since the data from the different “languages” were similar, we will concentrate on birds trained to peck to /i/. The task for these birds was to peck to exemplars of a distribution varying widely in F1 and F2 characteristics centered on a typical version of a male-produced /i/ (Peterson & Barney, 1952). These birds also learned to refrain from pecking to members of a distribution of vowels surrounding a typical /✍/ production. The distributions were quite similar to those used by Kuhl *et al*. (1992). After less than 100 hours of exposure, the starlings were presented with novel instances from each of the vowel categories. The birds showed facile generalization to these novel stimuli. That is, they categorized these previously unheard vowels in a linguistically-relevant manner.

Interestingly, the birds did not treat all novel instances as equivalent. They pecked much more to the vowels with the highest F2 frequencies, i.e. vowels that were maximally different from the /✍/ exemplars. This response gradient can be predicted from general principles of learning which have been described long ago (e.g. Spence, 1936; 1937; 1952; 1960). One of the facts that these early learning theorists wished to explain was that a positive response to one stimulus (S+) was affected by the nature of a second stimulus (S-) which discouraged responding. A classic experiment in this regard (Hansen, 1959) demonstrated that the peak of the discrimination function for responses by pigeons that were trained to respond to a visual stimulus at one wavelength (S+) would shift to a longer wavelength when S- was a shorter wavelength. Basically, this “peak shift effect” consisted of the response pattern S+ (excitatory) being skewed away from S- (inhibitory).

What is truly informative about the “peak shift” in peck rate responses to /i/ stimuli isn’t the historical precedent, but the fact that this pattern of responses has been evidenced by humans listening to similar vowel stimuli. Lively (1993) presented humans with an /i/ distribution and asked them to make judgments as to the "goodness" of the stimuli as exemplars of the vowel /i/. Humans judged the stimuli with the highest F2 frequencies to be the best stimuli despite the fact that these are not typical of fluent production. Lotto *et al*. (1995) also had humans rate the "goodness" of the stimuli presented to the starlings. The ratings agreed with those from Lively (1993) and with the peck-rate values of the starlings: the "best" /i/ exemplars were those with the highest F2 frequencies. In fact, the agreement between the starling responses and the human judgments was quite remarkable. The correlation coefficient calculated between the positive vowel responses of the birds and the /i/-goodness of the humans was r=.671. This is an astounding degree of agreement given the difference in task and experience of the participants.

Is it simply serendipity that human adults and starlings have such similar patterns of responses to the same stimuli? Or are these results readily explainable in terms of general principles of learning and categorization? In order to provide some evidence to this debate, Lotto *et al*. (1995) trained a simple linear associator with the same vowel distributions presented to the humans and birds. This associator can be conceptualized as a simple neural network which operates by principles similar to Hebbian learning (Hebb, 1949). The pattern of responses from this network for the novel test stimuli were very similar to the pattern elicited from the birds (r=.678). These results serve as an existence proof for the ability of simple general learning mechanisms to account for the categorization ability of birds, human adults and human infants.

5. General Discussion

The debate over the generality of speech-perception processes is fundamental to future progress on a comprehensive theory of speech communication. The set of permissible explanatory constructs will depend on the resolution of this dialectic. We believe that until we exhaust the explanatory possibilities of classic theories of perception and learning, there should be no proposals of specialized mechanisms.

As an addendum to this discussion, It should be noted that we believe that learning is an essential part of speech perception. In the past, opponents of "speech-is-special" theories have been considered slaves to the operating characteristics of the auditory system. Whether this was a fair characterization or not, we do fully acknowledge that most phonetic categories and many other aspects of speech perception do not arise because of quirks of the auditory system, but are due to processes of learning. We are proponents of what we call a General Auditory and Learning Approach (GALA) to speech communication. The basic premise of this approach is that the development of the human communication system precipitated upon the general auditory and learning capabilities of the genus *Homo* (which are apparently shared by many other animals). Many aspects of our modern speech system take advantage of these general capabilities. Because of the general perceptual processes described by the "Lens Model", the communication system utilizes multiple stochastic attributes; because of general perceptual contrast, a production system which is assimilatory (coarticulation) is effective; because Hebbian learning is a part of our biology, phonetic categories have their characteristic structure. This may appear backwards from typical speech theorizing (i.e. the production system determines the attributes of the perceptual system), but we believe it adheres better to the scientific principles of generality and parsimony than do theories which proffer specialized processes for the perception of speech signals.

6. Author Note

This work was supported by NIDCD Grant DC-00719 and NSF Young Investigator Award DBS-9258482 to the second author.

References

Blechner, M. J. 1977. Musical skill and the categorical perception of harmonic mode. *Haskins Laboratories: Status Report on Speech Research*, SR- 51/52: 139-174.

Brunswick, E. 1950. The conceptual framework of psychology. In O. Neurath (Ed.) *International Encyclopedia of Unified Science*, 1(10).

Brunswick, E. 1955. Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62: 193-217.

Burdick, C. K., & Miller, J. D. 1975. Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustic Society of America*, 58: 415-427.

Cathcart, E. P., & Dawson, S. 1928-1929. Persistence (2). *British Journal of Psychology*, 19: 343-356.

Christman, R. J. 1954. Shifts in pitch as a function of prolonged stimulation with pure tones. *American Journal of Psychology*, 67: 484-491.

Diehl, R. L., & Walsh, M. A. 1989. An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, 85: 2154-2164.

Fowler, C. A., Best, C. T., & McRoberts, G. W. 1990. Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48: 559-570.

Fowler, C. A., & Rosenblum, L. D. 1990. Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance*, 16: 742-754.

Hansen, H. M. 1959. Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 58: 321-372.

Hebb, D. O. 1949. *The Organization of Behavior*. New York: Wiley.

Holt, L. L., Lotto, A. J., & Kluender, K. R. 1996. Perceptual compensation for vowel undershoot may be explained by general perceptual principles. *Journal of the Acoustical Society of America*, 99: 2589.

Jusczyk, P. W. 1993. Sometimes it pays to look back before you leap ahead. In D. de Boysson-Bardies *et al*. (Eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. The Hague: Kluwer Academic Publishers. pp. 227-235.

Kluender, K. R. 1991. Effects of first formant onset properties on voicing judgments result from processes not specific to humans. *Journal of the Acoustical Society of America*, 90: 83-96.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. 1987. Japanese Quail can learn phonetic categories. *Science*, 237: 1195-1197.

Kluender, K. R., & Lotto, A. J. 1994. Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans. *Journal of the Acoustical Society of America*, 95: 1044-1052.

Kuhl, P. K. 1978. Predispositions for the perception of speech-sound categories: A species-specific phenomenon. In F. D. Minifie & L. L. Lloyd (Eds.) *Communicative and Cognitive Abilities - Early Behavioral Assessment*. Baltimore: University Park Press. pp. 229-255.

Kuhl, P. K. 1983. Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, 6: 263-285.

Kuhl, P. K. 1986a. Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology*, 45: 233-265.Kuhl, P.

Kuhl, P. K. 1986b. The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.) *Categorical Perception*. Cambridge, MA: Cambridge University Press. pp. 355-386.

Kuhl, P. K. 1993. Innate predispositions and the effects of experience *in* speech perception: The Native Language Magnet Theory. In D. de Boysson- Bardies *et al*. (Eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. The Hague: Kluwer Academic Publishers. pp. 259-274.

Kuhl, P. K., & Miller, J. D. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 63: 905-917.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants six-months of age. *Science*, 255: 606-608.

Liberman, A. M. 1996. *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review*, 74: 431-461.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. 1954. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 68(8, Whole No. 379).

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21: 1-36.

Lisker, L. 1975. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 57: 1547-1551.

Lively, S. E. 1993. An examination of the perceptual magnet effect. *Journal of the Acoustical Society of America*, 93: 2423.

Lotto, A. J., & Kluender, K. R. in press. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*.

Lotto, A. J., Kluender, K. R., & Holt, L. L. 1995. Animal and computational models of development of graded vowel categories. *Journal of the Acoustical Society of America*, 98: 2965.

Lotto, A. J., Kluender, K R., & Holt, L. L. in press. Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*.

Miller, J. L., & Liberman, A. M. 1979. Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25: 457-465.

Mann, V. A. 1980. Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28: 407-412.

Mann, V. A. 1986. Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners perception of English “l” and “r”. *Cognition*, 24: 169-196.

Peterson, G. E., & Barney. H. L. 1952. Control methods used in the study of the vowels. *Journal of the Acoustical Society of America*, 24: 175-184.

Pisoni, D. B. 1977. Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61: 1352-1361.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. 1983. Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception & Psychophysics*, 34: 314-322.

Postman, L., & Tolman, E. C. 1959. Brunswick’s Probabilistic Functionalism. In S. Koch (Ed.) *Psychology: A Study of a Science. Volume 1. Sensory, Perceptual, and Physiological Formulations*. New York: McGraw Hill.

Repp, B. H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92: 81-110.

Spence, K. W. 1936. The nature of discrimination learning in animals. *Psychological Review*, 43: 427-449.

Spence, K. W. 1937. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 44: 430-444.

Spence, K. W. 1952. The nature of the response in discrimination learning. *Psychological Review*, 59: 89-93.

Spence, K. W. 1960. *Behavior Theory and Learning*. Englewood Cliffs, NJ: Prentice-Hall.

Stevens, E. B., Kuhl, P. K., & Padden, D. M. 1988. Macaques show context effects in speech perception. Paper presented at *116th Meeting of the Acoustical Society of America*, Honolulu, Hawaii.

Stevens, K. N., & Blumstein, S. E. 1981. The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.) *Perspectives in the Study of Speech*. Hillsdale, NJ: Erlbaum.

Summerfield, Q., & Haggard, M. 1977. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 62: 435-448.

Sussman, H., McCaffrey, H., & Matthews, S. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, 90: 1309-1325.