surements can be conducted in concert with perceptual experiments (Nakayama 1994). Where this is not possible, physiological hypothesis are only of value to the extent that they can constrain or substantially add to perceptual theories; for example, by showing that some putative invariant is specifically tuned to facts of auditory neural processing. The "orderly output constraint" might have served this purpose if in fact the locus equations could serve a similar role in the perception of stop consonants as interaural time difference arrays do in the barn owl. However, they cannot carry this burden on their own since "other information, such as the release burst, shape of the onset spectra, and voice onset time will also contribute to stop place identification during normal speech perception" (sect. 6.1, para. 4). In the absence of a detailed model of the interaction of these various cues, speculations as to a perceptual role for locus equations is difficult to evaluate.

Let me illustrate with an example from my own work of what I take to be the advantage of Gibsonian approach to speech perception. I have for some time been looking at the question of invariance as it relates to the perception of quantity in Icelandic, a language that distinguishes long and short vowels and consonants in stressed syllables (Pind 1986; 1995). Of particular interest are those kinds of syllables where a long vowel is followed by a short consonant or vice versa. Consider typical production data as shown in Figure 1. It can readily be seen that speaking rate affects the overall durations of vowels and consonants. Indeed, a close examination of the figure would reveal that a phonemically short vowel, spoken slowly, can easily become longer than a phonemically long vowel spoken at a fast rate. Because listeners are usually not troubled by changing speaking rates, it may be surmised that some invariant can be found for the speech cue of duration. Indeed, looking at the figure, it can readily be seen that there is no overlap in the data as plotted here on a two-dimensional scatterplot, showing simultaneously vowel and consonant duration. This suggests that a ratio of vowel to consonant duration could serve as the higher-order invariant. This is borne out by perceptual studies that show (Pind 1995) that the listener more or less bisects the vowel-consonant (VC)-plane as shown in Figure 1, hearing syllables of type V-C if the vowel is longer than the consonant and vice versa.

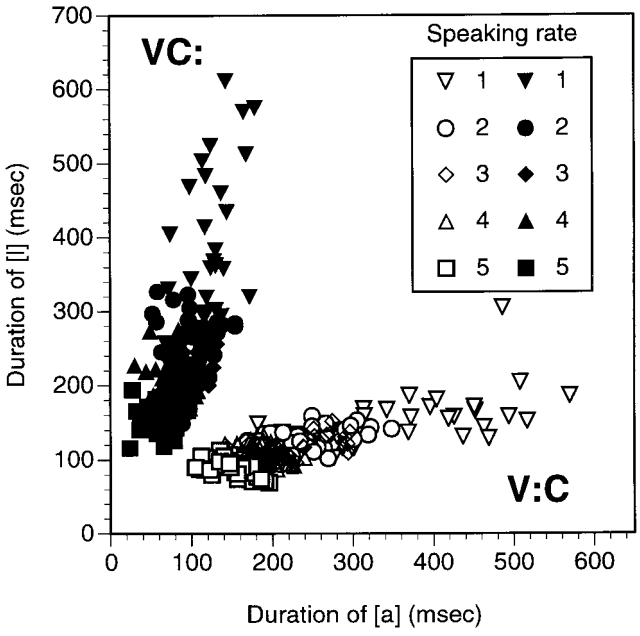

Figure 1 (Pind). Measurements of the durations of the vowel [a] followed by [l] in two-syllabic Icelandic words, spoken by four speakers at five different speaking rates from very slow (1) to very fast (5). The words either have a long vowel followed by a short consonant (type V:C -- open symbols) or vice versa. The distributions of these durations suggest an invariant for quantity expressed in terms of the ratio of vowel to consonant durations (from Pind 1995).

The interesting thing about this relational cue is that it is self-normalizing with respect to speaking rate. Changes in speaking rate will affect the durations of vowels and consonants, and the overall durations of the syllables. The relational speech cue needs no rate adjustments; it will stay invariant in the face of quite large transformations of rate.

Although it has been claimed that the case for invariants in speech is often overstated (Lindblom 1986), I would argue that the notion of invariants provides a convenient reference from which to pursue the study of speech perception. As an exhortation to experimental studies it is still without equal.

On the ontogeny of combination-sensitive neurons in speech perception

Athanassios Protopapas[a] and Paula Tallal[b]

[a]Scientific Learning Corporation, Berkeley, CA 94704; [b]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, NJ 07102.

protopap@scilearn.com www.scientificlearning.com; tallal@axon.rutgers.edu

Abstract: The arguments for the orderly output constraint concern phylogenetic matters and do not address the ontogeny of combination-specific neurons and the corresponding processing mechanisms. Locus equations are too variable to be strongly predetermined and too inconsistent to be easily learned. Findings on the development of speech perception and underlying auditory processing must be taken into account in the formulation of neural encoding theories.

The issue of acoustic invariance in phonetic perception has long baffled speech scientists. Reliable derivation of place of articulation from acoustic information remains essentially an unsolved problem, for both automatic speech recognition and human perceptual modeling. Sussman et al. propose that locus equations constitute a consistent cue and speculate on the possibilities for the emergence of the observed regularity and its perceptual significance. Despite several remaining questions, the idea that combination-responsive neurons constitute a cross-species mechanism for solving species-specific problems touches on many important issues. We would like to comment on the interplay between genetic and environmental constraints in the ontogeny of speech perception as it might apply to locus-equation specific, combination-sensitive neurons.

Several lines of evidence support the notion that humans are born with the capacity to discriminate between phonetic contrasts despite cross-linguistic differences that influence subsequent phonetic development (see Jusczyk, 1997, for discussion and review of findings). Neural mechanisms are likely to exist for the detection of formant frequencies, perhaps as an evolution of species-specific call detectors (Rauschecker et al. 1995) or for the estimation of body size (Fitch 1997). Neurons sensitive to spectral energy transitions of specific slopes such as those found in the ferret cortex (Shamma et al. 1993) may in turn constitute formant transition detectors. Whatever the specifics turn out to be, there is certainly a strongly innate component to basic auditory processing that underlies the infant's earliest phonetic perception.

On the other hand, support for a learning-based notion of relatively low-level perceptual functions comes from findings on the phonetic development of language-learning impaired (LLI) children showing that (1) there exist individuals with severe impairments in phonetic perception and in nonspeech auditory processing (Tallal & Piercy 1973; 1974), and (2) the observed deficits in these individuals can be substantially ameliorated through specialized training in auditory processing of speech and nonspeech stimuli (Merzenich et al. 1996; Tallal et al. 1996). There is now mounting evidence to suggest that the perceptual

deficits in LLI children are not speech-specific but stem from a generalized impairment in auditory processing (Wright et al. 1997; see Bishop, 1992, and Farmer & Klein, 1995, for review). This impairment has been found to be present within the first 6 months of life in children genetically at risk for LLI and to predict subsequent language delay (Benasich & Tallal 1996). The relatively rapid improvement that can be brought about by specialized auditory training indicates that basic auditory perception underlying speech perception is subject to powerful learning effects, as language-specific phonetic perception must also be.

Analogies from nonhuman species can be powerful when operating on similarly predetermined processing mechanisms, either genetically "hardwired" or strongly biased in terms of physiological and environmental constraints. The speech perception literature, in particular, has gained substantially from cross-species research. The analogies from nonhuman species offered by Sussman et al., however, differ from locus equations and speech perception in some important respects. Specifically, the overlap between locus-equation combination cues for different places of articulation stands in contrast to the unambiguous mapping from combination cues for both the isovelocity categories in the mustached bat and the iso-interaural time difference (ITD) categories in the barn owl. Consequently, what is relatively straightforward for the bat to learn may be very difficult if at all possible in the case of speech perception.

Furthermore, velocity and ITDs are well-defined physical properties that do not vary between individuals, groups, or time frames. In the cases of the nonhuman species used to illustrate the orderly output constraint principle, the corresponding combination-specific neural responses to a great extent may be genetically encoded, as a result of adaptation on an evolutionary time scale. Human listeners, however, must learn (or at least fine-tune) during development the specific places of articulation and their combinations with manner of articulation of their language. In contrast to the nonhuman analogies of Sussman et al., a hardwired processing mechanism for locus equation cues in human speech perception seems unwarranted.

In summary, it is doubtful that locus equations for speech perception are on par with isovelocity or iso-ITD cues, regardless of the relative degree of environmental (signal-bound) and genetic (physiology-bound) constraints. It remains possible, however, that a neural mechanism of cue combination exists that forms higher-order features from perceptual inputs. Advances in neural network simulations have shown many ways in which such learning is possible and, indeed, functional (if still speculative with respect to human perceptual learning). It remains to be specified, however, where in the speech/auditory processing system such combination-sensitive neurons are to be found, to what extent their connectivity (and function) is dependent on the acoustic environment, and how language-specific properties are fine-tuned throughout development.

Listening to speech in the dark

Robert E. Remez

Department of Psychology, Barnard College, New York, NY 10027-6598.
remez@paradise.barnard.columbia.edu
www.columbia.edu/ barnard/psych/fac-rer.html

Abstract: This commentary questions the proposed resemblance between the auditory mechanisms of localization and those of the sensory registration of speech sounds. Comparative evidence, which would show that the neurophysiology of localization is adequate to the task of categorizing consonants, does not exist. In addition, Sussman et al. do not offer sensory or perceptual evidence to confirm the presence in humans of processes promoting phoneme categorization that are analogous to the neurophysiology of localization. Furthermore, the computational simulation of the linear model of second formant variation is not a plausible sensory mechanism for perceiving speech sounds.

Osteoarthritis is universal in humans by age 70. It is also observed in elderly fish, amphibia, reptiles (including dinosaurs), birds, bears, whales, and dolphins. The universality of this form of articular disorder has been taken to reflect the action of a paleozoic mechanism of joint repair rather than a specific disease afflicting humans. A satisfactory account of the biology of osteoarthritis would describe the cellular functions by which the tissues are established, and the mechanical, biochemical, and enzymatic forces that promote hypertrophy. To accomplish this descriptive and explanatory goal, animal models are exploited, and only the species that exhibit the ailment are suitable to model it. Despite wide distribution of degenerative joint disease among vertebrates, it is nonetheless possible to make an unlucky choice of animal model. Bats do not manifest it at all, nor do sloths, though both are bony and are similar in evolutionary history and physiology to animals that, like the rest of us, exhibit structural changes in aged joints.

When contemplating the biology of language, far rarer among species than joint disease, there can be little hope of exploiting an animal model. There is simply no veterinary instance of language. Without an animal model of language, Sussman et al. propose instead to use the mustached bat as a partial model. In doing so, they went out on a limb already well populated by those of us who have asserted analogies between aspects of language and all sorts of ways that animals think or act. The present case is distinguished by a reliance on assertions of rough similarity, on claims that are cautious albeit hopeful, and on indirect empirical tests. Despite its ambition and its well-informed rendition of the neurophysiology of localization, the target article is not convincing about language, leaving even this modest and partial correspondence of human and animal nature merely arguable and conjectured.

The target article does succeed in a goal it set for itself: to propose an analogy between the auditory functions that promote phonetic perception and the neurophysiological vignettes of bats and owls. Indeed, the exposition is a profusion of analogies: (1) Localization by bats is analogous to localization by owls, both using combination-sensitive neurons (sect. 1, para. 2). (2) Auditory localization is analogous to phonetic categorization (sect. 1.2), both requiring the recognition of acoustic elements in combination and permutation. (3) An owl or bat recognizing an auditory pattern is analogous to a human listener recognizing an auditory pattern (sect. 1.3.1). (4) The auditory systems that support these functions are analogous, perhaps necessarily so, if not homologous (sect. 1.3.2). (5) The auditory maps representing interaural phase differences as iso-velocity contours are analogous to maps that represent frequency transitions in formant-centers as iso-stop-place territories, regions within the space unique to phonetic features of place of articulation (sect. 7; Figs. 2 and 16). (6) Localization in bat and owl exploits low-variance linearities in an impinging signal correlated with direction; by analogy, so would an auditory mechanism responsible for pattern recognition in speech (sect. 6.2). (7) The coevolution of auditory and motor components of speech is analogous to the coevolution of the visual sensitivity of bees and the production of pigment by flowers (sect. 6.2). Throughout the exposition, analogies pile up with no defense of the aptness of any of them, a circumstance in which an allegation of unelaborated similarity between localization and categorization of phonetic segments fits. This format allows Sussman et al. to endorse an answer that appeals to them – linearity and low-variance sensory maps – before defining the compliant question. We should find nothing unusual about this. It is a customary pretheoretical way to appraise the psychological applicability of findings in sensory physiology, and is the only way available to us for devising a physiologically justified account of the causes of phonetic perceptual impressions (cf. Rock 1970). When we discover a specific mechanism, we consider the likelihood that its operating characteristic is global, rather than local. Does the strategy work here?

The enterprise fares poorly in implementing a computational analog of this neural mapping mechanism that proves adequate to