# The Illusion of the Phoneme

Andrew J. Lotto
Loyola University Chicago
Lori L. Holt
Carnegie Mellon University

## 0. Caveat

A caveat is warranted here. While our title is provocative, our ambitions are much more prosaic. Obviously the debate on the ontological status of the phoneme has a long and complicated history. We offer neither a summary of this debate nor a last word on the question. We seek only to question the role of the phoneme in the perception of speech and, in doing so, we hope to demonstrate that the empirical evidence for the causal role of the phoneme in perception is limited.

## 1. The Ontological Status of the Phoneme

Classifying speech in terms of discrete abstract labels has been an extremely successful descriptive system for linguistics. While there are well-known concerns about the "shoe-horning" of different languages' phonetic systems into universal phoneme sets, there is little doubt that this descriptive symbol system has been instrumental in the progress of the speech sciences. Phonemes offer a communication system for empirical results. In addition, standardized descriptions of different linguistic sound systems have made the immense variation in language manageable. But has this desire for simplicity and generality blinded us to some of the realities of speech as a communication system?

### 1.1 The "Categorization Error"

The early Gestalt Psychologists (e.g. Köhler, 1930; Koffka, 1935) warned against making the "experience error" when theorizing about human perception. This error arises when "we mistake the result of organization for the cause of organization" (Koffka, 1935). For the Gestalt theorists, structure and organization in perception did not necessarily imply that this structure and organization must be present in the input. We believe that this cautionary note applies beyond the scope of "illusory contours" and perceptual grouping rules. Structure and organization in behavior need not imply that this structure and organization is present in mental representation. It is easy enough in many areas of study to mistake efficient descriptors of the structure of a system's behavior as causal entities responsible for that structure. For example, when subjects are asked to rate members of a category for their typicality (e.g. 'birds' or 'furniture' or 'members of /i/') there is a clear structure in their responses. In particular, there is usually a prominence in the ratings across a subset of the members. This prominence in responses is often described as a *prototype*. Similar prominences in

response structure can be observed for a variety of tasks and measures (e.g. reaction time for identification). The prototype is a very useful descriptor of the response structure for these types of tasks. After all, a prominence in responding is a very salient attribute and it may have functional significance..

Beyond being an important descriptive entity, prototypes in the response structure are often taken as evidence for *mental* prototypes internal to the subject which are responsible for the structure of the output (e.g. Posner & Keele, 1970; Rosch, 1975; 1978). This theoretical approach appears rather compelling: the structure of mental representations is mirrored in the structure of responses. However, we now have a variety of models of categorization that lead to response structures that include a prominence or "prototype". Exemplar models and connectionist net models deliver similar response structure with no explicit representation of a *mental* or internal "prototype" (see, e.g., Reed, 1972; Nelson, 1974; Brooks, 1978; Medin & Schaffer, 1978; Knapp & Anderson, 1984; Kluender, Lotto, Holt, & Bloedel, 1998). It is clear, then, that the presence of a "prototype" in *behavioral data* does not necessitate a "prototype" in the *mental representation* of the stimuli. To make this sort of presumption about mental representation structure from response structure is to commit something akin to the "experience error".[i] Whereas, the Gestalt psychologists reprimanded the presumption that subjective organization is determined by organization in the stimulus, we are suggesting that it is wrong to presume that structure in response follows in a straightforward manner from similar structure in mental representations. The Gestalt version concerned the "experience" of subjective perception. We are discussing the "experience" of objective data. This is stretching the metaphor a bit and, thus, we prefer to refer to this as the *categorization error*. That is, sometimes we mistake the description of response structures (i.e. we *categorize* the structure for efficiency) as causes of response structures (i.e. we presume the *category* to exist as a mental entity).

Another example of a "categorization error" would result from confusing "rule-described" behavior and "rule-following" behavior (Heil, 1983; Ben-Zeev, 1987). The fact that behavior can be described efficiently by rules does *not* entail that the behavior is the result of the following of explicit rules. This distinction was made clear by Wittgenstein's (1953) "skeptical paradox". A contemporary example is the production of the past tense in English. The morphological changes of regular verbs from present to past tense can be described quite efficiently by a rule. This is certainly an example of rule-*described* behavior. However, it would be an error to presume that this is equivalent to saying that the behavior is a result of explicit rule *following*. Connectionist networks can map the morphological change for regular verbs and even novel exemplars without the representation of explicit rules (Rumelhart & McClelland, 1986). These networks even appear to model the trajectory of acquisition of correct production of the past tense by children, again without reference to explicit rules (though, see Pinker & Prince, 1988 for a critical review of these results).[ii]

Are we committing the "categorization error" when we suggest that the phoneme is the fundamental representation or unit of speech? It is an efficient unit for describing speech behavior, but does that give us license to infer that it is a causal entity? And worse, could it be that because we give a prominent role to the phoneme that we start to see more regularity in speech perception and production than is actually there?

## 1.2 Phoneme as the Fundamental Unit of Speech

So, is the phoneme merely an efficient descriptor of language behavior or is it a fundamental functional unit of speech perception (and production)? It appears that across distinct theoretical divisions in speech research there is a pre-theoretical presumption that the phoneme is indeed the fundamental representation of speech perception.[iii] For example, here is a quote from some of the most respected researchers in the field:

> "How is it that on hearing the sounds of speech a listener perceives phonemes? Since the question is reasonable only if we assume that phonemes are perceived...we accept it and go on to ask how such perception might occur." (Liberman, Cooper, Studdert-Kennedy, Harris, & Shankweiler, 1966, as quoted in Walsh, 1989).

In the decades since this quote there have been some concerns raised about the accepted role of phonemes (e.g., Studdert-Kennedy, 1976, 1980). The phoneme representation has also had several explicit defenders (e.g., Nearey, 1990). However, speech perception researchers (the current authors included) too often simply proceed as if the issue has been resolved and conduct experiments to uncover how "a listener perceives phonemes". Most of the disparate theories of speech perception (excepting some forms of Direct Realism) simply presume that the end product of perception is the assignment of a discrete phonemic symbol. This presumption is codified in typical experimental paradigms that demonstrate effects of acoustic parameter manipulation on the forced-choice phonemic labeling of speech sounds. In order to avoid the categorization error, empirical evidence concerning the role of the phoneme in perception has to be brought to bear on the issue. To accomplish this we need to set down some defining characteristics of the phoneme that will lead to testable predictions about perceptual data.

## 1.3 Definition of Phoneme

It should be made clear that the "phoneme" that we are discussing here is not the orhographic unit for phonemic transcriptions of linguists, but a purported internal mental representation that may be presumed to be a functional unit for the speech perceiver-producer. This is the "mentalistic" notion of the phoneme as described by Jones (1967). It has its roots in phoneme theory back to its origins with de Courtenay in the 1870s, who talked of "psychophonetics". The notion of the

phoneme as a psychological (as opposed) to linguistic entity is also explicit in the work of Sapir (1925). Thus, this viewpoint is distinguishable from Trubetzkoy's functional view (1958) or Jones's (1967) "physicalist" notion.

Clearly, the concept of this mentalistic phoneme has gone through many revisions since de Courtenay. We see three elements that are typical (though by no means universal) of mentalistic notions of the phoneme. Phonemes are:

*Discrete*: Some continuous variation in speech sounds is quantized when a phonemic label is assigned.
*Abstract (Symbolic)*: The mental representation itself is a symbolic label similar in conception (though not necessarily similar in structure) to the phonemic transcription labels.
*Language-specific*: A phoneme is meaningful only in relation to a particular language. It is a meaningful functional unit for the idiolect of the particular speaker.[iv]

These properties are in direct contrast to the acoustic signal which carries the phonemic message from speaker to listener. The speech signal is:

*Continuous*: Discrete markers for phoneme boundaries have been notoriously difficult to find. The variation in the acoustic waveform is, for practical purposes, continuous.
*Physical*: The waveform is not an abstract symbol, but a lawful product of the movement and shape of articulators, the medium of sound travel and any intervening objects or sources.[v]
*Not Linguistically Marked*: Obviously, the acoustic waveform is not explicitly marked as characteristic of a particular language. In fact, it isn't even explicitly marked as language.

Because of this mismatch between the continuous signal and the discrete symbol, speech perception research has been confronted with several contumacious problems. The "problems" of signal segmentation, perceptual compensation for coarticulation, lack of invariance, and speaker normalization are all, to some extent, created by the desire to map a continuously varying signal on phoneme quanta. Could these long-standing problems be simply a consequence of our pre-theoretical assumptions?[vi]

To justify the large amount of work going into solving these aforementioned difficulties, it seems incumbent upon us to empirically validate the existence of phonemes or derive their necessity from first principles. We understand that there have been efforts to provide empirical evidence about the ontological status of the phoneme. Much of this previous work is based on patterns of responses in production and perception data (e.g. speech errors, Fromkin, 1971). These data can be quite compelling, but still we are left with the possibility of a categorization error.

We present below some recent empirical data that we feel are relevant to the debate. The questions we pose are these: What role does the phoneme play in speech perception? What are the fingerprints left behind by this purported fundamental functional unit of speech perception?[vii] Can we display the "causal efficacy" of phonemic identity in a non-circular manner?

## 2. Empirical Evidence

### 2.1 Causal Efficacy
How do we decide on the existence of a proposed entity? One important notion may be what is termed "causal efficacy" (e.g., Gasking, 1955). This is the ability of an entity to cause measurable effects by its presence or absence. It is the hallmark of the standard empirical approach that we measure the effects caused by an entity in order to discover something about the entity (e.g. if it exists).

An example is the search for the existence of neutrinos. These nearly massless chargeless particles were hypothesized to exist by Enrico Fermi. Unfortunately, the particles have very weak interactions with other particles. Thus, it was hard to detect the effects of the neutrino. That is, it was difficult to demonstrate the causal efficacy of neutrinos. The existence of the neutrino was finally established because antineutrinos[viii] initiate a particular reaction. Their presence or absence determines the reaction.

Can we empirically establish the causal efficacy of phonemes? One may suggest that we have evidence for phonemes in the fact that human adults can label speech sounds with appropriate phonemic labels. Unfortunately, besides making us vulnerable to a categorization error, this demonstration is a bit circular. We define phonemes by the perceptual behavior of adult humans. We cannot use that same data as proof that they exist. Instead we need to show that the hypothesized presence of a particular phoneme has demonstrable effects on behavior that do not occur in its absence.

### 2.2 Context Effects
One place where we may look to see the causal efficacy of the phoneme is in context effects. That is, can the change of phonemic identity (conceptualized as the activation of a particular mental representation) affect the response made to a nearby speech sound?

We do see such effects. For example, Mann (1980) reported that an ambiguous consonant-vowel (CV) syllable is labeled as 'ga' following /al/ and as 'da' following /ar/. This appears to be a causal effect of a phoneme. The presence of /l/ appears to cause a change in the response to the subsequent CV. It may be proposed that the auditory signal corresponding to the initial VC activated the representation for the phoneme /l/. This activated representation in turn affected the label assigned to the CV. Proposing this from the given data would be an example of the categorization error, however, since we would be presuming that the /l/ phoneme was present in the mind because a phonetic structure was

present in the description of the signal. In order to make this explanation tenable, we need some independent evidence that it was the presence of the mental phoneme /l/, *per se*, that caused the response shift and not the structure of the input signal itself.

Mann (1986) presented these same VC CV disyllables to native Japanese speakers. These listeners could not distinguish between English /l/ and /r/ and, therefore, it is unlikely that they would have the mental representations for these phonemes. If we are to ascribe causal efficacy to phonemes based on this context effect then the effect must disappear in the absence of these particular phonemes in Japanese listeners[ix]. Contrary to this prediction, Japanese listeners shifted their responses to the CV in the same manner as English speakers. Even the size of the effect (in terms of identification boundary shift) did not differ between groups. This demonstration not only fails to provide supportive evidence for the causal efficacy of phonemes; it also provides some disquieting information about the possible role of the phoneme as a fundamental functional unit. In Japanese, there is no phonemic distinction (in the descriptive sense) between English [r] and [l] and both are subsumed under the Japanese /r/. If there *are* mental phoneme representations then the Japanese listeners in Mann's (1986) study most likely represented [al] and [ar] as /ar/ and /ar/. The fact that these equivalent labels led to widely disparate effects leads one to question what fundamental role phonemes play in speech perception. It appears that the context effects are in no way affected by phonemic identity.

Additional evidence concerning the impotence of phonemic labels in this context effect was presented by Lotto, Kluender, and Holt (1997). We trained Japanese quail (*Coturnix coturnix japonica*) to peck to a key when presented with the syllable [ga] and to refrain from pecking when presented with the syllable [da].[x] In a sense, they learned to label the syllables by their pecking responses. After training, they were presented ambiguous CVs preceded by either [al] or [ar]. They demonstrated a shift in "labeling" in a similar manner to human English and Japanese listeners. That is, they pecked more (a "ga" response) to CVs in the presence of [al] than they did to CVs in the presence of [ar]. Since birds are unlikely owners of mental phoneme representations for English /l/ and /r/, this is another example of the context effect occurring in the absence of phoneme representations. Taken together with Mann's (1986) data for Japanese listeners, these data undermine the use of context effects to demonstrate the causal efficacy of phonemes.

Even if we could validate the existence of mental phoneme representations, it would be difficult to claim that phonemes are the fundamental functional units of speech perception. As the studies described above demonstrate, sub-phonemic shifts in acoustic attributes of a stimulus can cause dramatic context effects on perception of subsequent speech. This suggests that there is a unit more fundamental to the processes of speech perception than the phoneme. A natural, and historically popular, candidate for this unit is the

distinctive feature. However, distinctive features will not suffice to explain the parallels in human and quail behavior.

It appears that the causal level for context effects in speech perception may be more general and not language-specific. In a recent study, we looked at the effects of non-speech context on the labeling of CVs. A series of syllables varying from [ga] to [da] were synthesized by manipulating the onset frequency of the third formant. These syllables were preceded by either synthesized versions of [al] or [ar] or non-speech complexes that matched some of the acoustic attributes of [al] and [ar]. In particular, the non-speech stimuli were two flat sine waves at the offset frequencies of the third and second formants for [al] or [ar]. These sine complexes were matched in duration and energy to [al] and [ar]. A 50-msec silent gap separated the CVs and the context sounds. The listeners' task was to label the CVs as "da" or "ga" using keys on a response box. Figure 1 presents the identification data for the speech and non-speech contexts as mean percent 'ga' responses.

**Figure 1.** Context effects for speech and non-speech precursors.
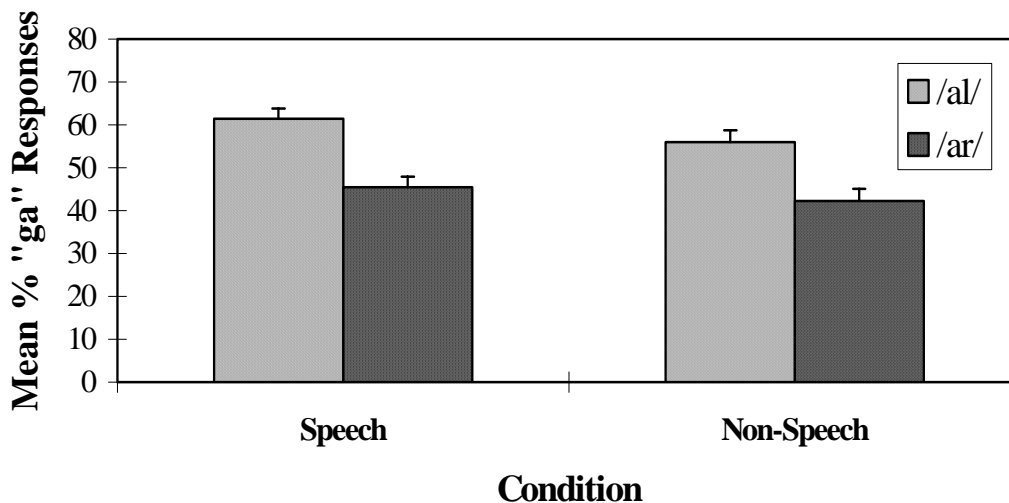


It is clear from the graph that the non-speech context also elicited a shift in the labeling behavior for the CV. In fact, the identifications in the non-speech context were statistically indistinguishable from the speech context. The size of the shift in responses did not statistically differ with context condition ($F_{9,117} < 1$). These data suggest that general auditory attributes of the speech signals are causing the context effects and not any linguistic-specific representation. Over the last couple of years, we have demonstrated a variety of similar non-speech context effects on speech labeling (Holt, Lotto, & Kluender, 1996, 1998; Lotto & Kluender, 1998). We believe that these data require one to consider auditory information as fundamental to speech perception when discussing functional

units. The importance of general auditory attributes in any discussion of speech perception (or production) is also expressed convincingly by Ohala and Perkell (both from this session).

## 2.3 Boundaries of Causal Efficacy

Another way of talking about an "object" of speech perception is to define the boundaries of causal efficacy. That is, we can try to empirically determine what parts of the speech signal act together as a cohesive unit in inducing behaviorial changes. This is basically a problem of functional segmentation. Do these empirically determined boundaries correspond to phonemes?

An experiment reported by Lotto, Kluender, and Green (1996) was designed specifically to determine the boundaries of casual efficacy. The experiment was based on previous research demonstrating that the identification of a syllable-initial consonant can be shifted on the basis of the duration of the subsequent vowel (Miller & Liberman, 1979). This effect has been considered indicative of perceptual normalization for changes in speaking rate. So, for example, the identification boundary between /bi/ and /pi/ will shift depending on the duration of [i], with a longer vowel leading to more /bi/ responses (Green & Miller, 1985).

Of course, it would be a categorization error to simply assume that the object whose duration is important for this effect corresponds to the phoneme /i/ or to any specific linguistic entity. Simply because it is useful to label the stimuli as /bi/ or /pi/ does not mean that the underlying perceptual representation will respect these labels. Lotto *et al.* (1996) tried to determine if it was the duration of phonemes that mattered in this "rate-normalization" effect or whether it was the duration of some other perceptual unit. The experiment included four series containing synthesized CV syllables, where the consonant varied from [b] to [p] through the manipulation of voice-onset time (or more correctly, first formant cutback, Kluender, 1991). One series (the *Long* series) included CVs with a long [i] vowel. The total syllable duration was 350 msec. A second series' (the *Short* series) CVs contained short [i] vowels with a total syllable duration of 110 msec. The final two series contained CVs that were 350 msec in total duration (the duration of the *Long* CVs), but contained a sudden disjunction in the vowel after 110 msec (the duration of *Short* CVs). One disjunction (*ID-Shift*) was slight in spectral change but crossed the identification boundary causing the phonemic identity of the vowel to shift from /i/ to /I/ (as determined by a separate vowel identification task with the same listeners). The second disjunction (*Spectral-Shift*) was large in terms of spectral change, but did not result in a shift in phonemic identity. The vowel shifted from a relatively neutral exemplar of /i/ to a more extreme (in the F1xF2xF3 frequency space) exemplar of /i/.

The predictions are straightforward (even if the stimuli are not). If phonemes define the boundaries of a perceptual unit for the calculation of duration in the rate-normalization effect, then the *ID-Shift* series should be identified similarly to the *Short* series because the effective duration of the

following vowel will terminate at the phonemic shift disjunction. On the other hand, the *Spectral–Shift* series identification function should more closely resemble the *Long* series function because the spectral disjunction can be ignored with no accompanying shift in mental phonemic label. That is, the disjunction in the *ID-Shift* CVs should define two perceptual functional units (phonemes) and the *Spectral-Shift* disjunction should be integrated into one functional unit.

The results listed as identification boundaries[xi] in terms of msec of voice onset time are listed below in Table I.

Table I. Probit boundary values and significance test results from Experiment 2 of Lotto, Kluender and Green (1996).

| Long | Short | ID-Shift | Spectral-Shift |
|---|---|---|---|
| 39.28 msec | 34.63 msec | 38.99 msec | 33.27 msec |
| **Difference:** | **Difference:** | **Difference:** | **Difference:** |
| Short | Long | Short | Long |
| Spectral-Shift | ID-Shift | Spectral-Shift | ID-Shift |
| **No Difference**: | **No Difference** | **No Difference:** | **No Difference:** |
| ID-Shift | Spectral-Shift | Long | Short |

*Note*: Statistical differences were determined using Tukey least significant difference ($\alpha$=.05).

The data were in direct contrast to the predictions arising from a hypothesis of the phoneme as the fundamental functional unit. The identification of the *ID-Shift* series was statistically indistinguishable from the identification of the *Long* series. On the other hand the identification function for the *Spectral-Shift* series was statistically equivalent to the function for the *Short* series. This pattern of results suggests that, for the "rate-normalization" effect, spectral discontinuities define the extent of the duration effect and that phoneme boundary play little demonstrable role. Acoustic information that could be described as separate phonemes (*ID-Shift*) was effectively treated as a coherent perceptual unit.

In addition to these data, Rochelle Newman (this session) presents compelling data from experiments using the rate-normalization paradigm that also demonstrate that spectral changes appear to be of paramount importance in determining perceptual boundaries and that phonemes, *per se*, do little to predict the identification shifts.

## 2.4 Intra-phonemic Variation is Meaningful
The results from the previous two sections demonstrate a surprising dearth of response predictability associated with phonemic identity. At the least, these experiments weaken the claim that the phoneme is "the fundamental functional unit of speech perception". However, one may argue that the phoneme is not a functional unit of the *process* of speech perception, but that phonemes are the *output* of speech perception and are fundamental functional units of language for

lexical access and "higher" linguistic processes. This appears to be the view espoused by Liberman *et al*. (1966) quoted above. They assume that "phonemes are perceived". This also appears to be the underlying assumption of research using forced-choice identification methodology in which the response of concern is the phonemic labeling of a speech sound. This methodological assumption has assured the hegemony of the phoneme in speech perception research. Likewise, there are many computational models of speech perception that output a single phonemic label when provided with a speech signal. For example, Anderson, Silverstein, Ritz and Jones (1977) present the "Brain-State-in-a-Box Model" of speech categorization which gets rid of any differences of responses within a phoneme and increases differences between phonemes.

The "Brain-State-in-a-Box Model" is an extreme example of one of the main difficulties with the phoneme-as-output approach. Mapping the continuous speech signal to a discrete symbolic unit necessitates some loss of information. There is non-random variation in the speech signal that does not differentiate phonemes. Recent studies have made evident that this variation has meaningful consequences on response behavior -- even for behavior related to "higher" linguistic functions. For example, different acoustic stimuli that are identified as the same phoneme can easily be rated by listeners in terms of "goodness" as exemplars of the phoneme (e.g. Miller & Volaitis, 1989; Volaitis & Miller, 1992). This demonstrates the ability of listeners to respond on the basis of intra-phonemic variation which models of speech perception that output phoneme labels like the one mentioned above can not readily do. In addition, exemplars from a single phoneme differ in their effectiveness as adapters or as initiators of identification shifts (e.g. Samuel, 1982; Miller, Connine, Schermer, & Kluender, 1983; Lotto, Kluender, & Holt, 1998). Also, there have been many recent demonstrations of speaker effects on phonemic identification and lexical access (e.g. Schacter & Church, 1992; Nygaard, Sommers, & Pisoni, 1994). Goldinger (1996) shows that the view that non-phonemic variation in the auditory signal is discarded for the more efficient abstract phonemic representation is almost certainly false.

## 3. Conclusions and Future Directions

The data presented in the last section are but a small proportion of recent work showing the causal efficacy of variations within a single phoneme. However, suggesting that variation within a phoneme is meaningful is not the same as suggesting that mental phonemes do not exist. One may argue that it is possible that the phoneme is but one of several representations of the speech signal; all present in parallel with each other. Yes! But this view of phonemes as but one of many possible streams of information in the system forces some changes in our viewpoint.

### 3.1 The Phoneme is Not Fundamental

The question of this session was to describe the "fundamental", "basic" or "primitive" units of speech processing. If we view mental phonemes as but one of several representations of speech signals then it is hard to think of them as "fundamental", especially since, as we argue below, some of the other representations are sub-phonemic.

### 3.2 Basic Functional Units are Task-Specific

The purpose of multiple streams of information must be to allow flexibility with changes in task. We would like to suggest that future research must analyze the experimental tasks to determine what type of information is necessary to perform the task. In a task such as phonemic transcription, some phoneme representation may be functionally essential.

However, let us not presume that all language tasks require phoneme representations, just because people can do phonemic identifications. McMurray and Spivey (this session) presented data on the time course for recognition of a syllable as /ba/ or /pa/. This time course is too slow to account for the speed of word recognition. That is, if recognition of the phonemes in auditory speech was necessary for word recognition then reaction times for word recognition would be much slower. Also note that in the connectionist network model presented by McMurray and Spivey that accounts for their data, there was an all-or-none final decision on the identity of the syllable. However, if there were "word nodes" connected to the output of the phoneme nodes, then there would not need to be full phoneme recognition to recognize the word. The flow of information could determine the word spoken without all-or-none decisions at the phonemic level. Again, the units that are output or are "basic" are determined by the particular task.

As we continue to study speech through forced-choice identification and discrimination paradigms, we should be cognizant of the fact that the representational level we are studying in these tasks may be one of several parallel representations that, perhaps, are not essential for the next "level" of perceptual or linguistic tasks.

### 3.3 Auditory Level is Important for Explaining Language Behavior

As stated earlier, we believe that the evidence supports the notion that some general auditory representation of the speech signal is significant for explaining language behaviors and the development of language as a communication system. This view was also presented by Perkell and Ohala (both from this session). These general auditory representations of the effectively continuous[xii] auditory signal may be present in parallel with phonemic representations along with a number of other possible linguistic entities.

## 4. Summary

So, what is the basic unit of speech perception? That depends on what your definition of *is* is. Our (probably overstated) conclusions are that the answer is dependent on the particular task for the subject. Our nervous system probably retains much of the information in the speech signal until a response is determined. We suggest that hypothesized representations for language be determined through empirical work and by the application of rational first principles. In this way we can avoid the categorization error and determine exactly which "problems" of speech perception need to be solved.

## Notes

[i] This isn't to propose that there are no mental prototypes or that those who propose prototype theories are committing a grave error. In fact, there are many strong reasons to suppose that some kind of prototype theory of categorization is true. We are merely asserting that prototypes in response structure are not *prima facie* evidence for the existence of mental prototypes.

[ii] Another example that is relevant but not concerned with behavioral data is the adaptiveness that biological species show in relation to their ecological niche. The wonderful orderliness and harmony of animal and environment belies the randomness and capriciousness of the process of natural selection. While the current equilibrium of organism and environment may appear to be a clear picture of "God's plan" in nature, it does not necessarily follow that this seeming order resulted from processes exemplified by such order.

[iii] We focus here specifically on the question of speech perception. We believe that many of the same arguments apply to speech production, but this is a separate question.

[iv] This property isn't as obviously true if one presumes that there is a universal set of innately prescribed phonemes. In this case, there would exist mental phoneme representations that were independent of an infant's future "native" language. Given the difficulty that linguists have in defining phonemes in relation to a single speaker's "language", it may be difficult to give precise definitions for these kinds of innate representations.

[v] There are some who argue that mental representations gain meaning only to the extent that they are lawfully determined by physical sources (e.g. Dretske, 1981; 1986). The distinction made here remains true to the degree that the phoneme is not a representation of the physical properties of the acoustic waveform.

[vi] It should be noted that to the extent that distinctive feature systems are discrete, symbolic linguistic systems, they are subject to the same problems discussed here for the phoneme.

[vii] Perhaps indicative of a Freudian slip, this sentence was originally written as "What are the fingerprints left behind by this purported fundamental *fictional* unit of speech perception?"

[viii] The causal efficacy of antineutrinos is presumed to demonstrate the existence of neutrinos, given the surrounding theory of particle physics.

[ix] This is a rather simplistic notion of determining causality and the philosophy of science has much more rigorous standards than those presented here. However, we think that these arguments appeal to our common sense notions of causality.

[x] Some birds were trained to peck to [da] and refrain from pecking to [ga]. These birds pecked more in the presence of [ar].

[xi] Probit analysis is a class of nonlinear models of estimation for binary decision variables. The numbers represent the 50% identification boundary as determined from a linear regression taken through $z$-transformed scores.

[xii] Of course, the auditory representation can't be purely continuous. The signal is transmitted discretely even at the level of neural transduction. However, at the level of perception we are suggesting that the representations can be considered functionally continuous.

# References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. 1977. Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model. Psychological Review, 84: 413-451.

Ben-Zeev, A. 1987. A Critique of the Inferential Paradigm in Perception. Journal for the Theory of Social Behavior, 17: 243-263.

Brooks, L. 1978. Non-analytical Concept Formation and Memory for Instances. In E. Rosch & B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, NJ: Erlbaum.

Dretske, F. 1981. Knowledge and the Flow of Information. Cambridge, MA: MIT Press.

_____ 1986. Misrepresentation. In R. J. Bogdan (Ed.), Belief: Form, Content and Function. Oxford: Oxford University Press.

Fromkin, V. A. 1971. The Nonanomalous Nature of Anomalous Utterances. Language, 47: 27-52.

Gasking, D. 1955. Causation and Recipes. Mind, 64: 479-487.

Goldinger, S. 1996. Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22: 1166-1183.

Green, K. P., & Miller, J. L. 1985. On the Role of Visual Rate Information in Phonetic Perception. Perception & Psychophysics, 38, 269-276.

Heil, J. 1983. Perception and Cognition. Berkeley: University of California Press.

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.

Holt, L. L., Lotto, A. J., & Kluender, K. R. 1998. Spectral Contrast in Perception of VCV Syllables. Journal of the Acoustical Society of America, 104, 1759.

Jones, D. J. 1967. The Phoneme. Cambridge: Cambridge University Press.

Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. 1998. Role of Experience for Language-Specific Functional Mappings of Vowel Sounds. Journal of the Acoustical Society of America, 104: 3568-3582.

Knapp, A. G., & Anderson, J. A. 1984. Theory of Categorization Based on Distributed Memory Storage. Journal of Experimental Psychology, 10: 616-637.

Koffka, K. 1935. Principles of Gestalt Psychology. New York: Harcourt Brace.

Köhler, W. 1930. Gestalt Psychology. London: G. Bell & Sons.

Liberman, A. M., Cooper, F. S., Studdert-Kennedy, M., Harris, K. S., & Shankweiler, D. P. 1966. Some Observations on the Efficacy of Speech Sounds. Paper presented at the XVIII International Congress of Psychology, Moscow, USSR.

Lotto, A. J., Kluender, K. R., & Holt, L. L. 1998. The Perceptual Magnet Effect Depolarized. Journal of the Acoustical Society of America, 103: 3648-3655.

Lotto, A. J., & Kluender, K. R. 1998. General Contrast Effects in Speech Perception: Effect of Preceding Liquid on Stop Consonant Identification. Perception & Psychophysics, 60: 602-619.

Lotto, A. J., Kluender, K. R., & Green, K. R. 1996. Spectral Discontinuities and the Vowel Length Effect. Perception & Psychophysics, 58: 1005-1014.

Lotto, A. J., Kluender, K. R., & Holt, L. L. 1997. Perceptual Compensation for Coarticulation by Japanese Quail (Coturnix coturnix japonica). Journal of the Acoustical Society of America, 102: 1134-1140.

Mann, V. A. 1980. Influence of Preceding Liquid on Stop-Consonant Perception. Perception & Psychophysics, 28: 407-412.

_____ 1986. Distinguishing Universal and Language-Dependent Levels of Speech Perception: Evidence from Japanese Listeners' Perception of English "l" and "r". Cognition, 24: 169-196.

McMurray, B., & Spivey, M. 1999. Categorical Perception of Consonants: The Interaction of Learning and Processing. Presentation at the 35th Regional Meeting of the Chicago Linguistic Society, April 23rd, University of Chicago.

Medin, D. L., & Schaffer, M. M. 1978. A Context Theory of Classification Learning. Psychological Review, 85: 207-238.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. 1983. A Possible Auditory Basis for Internal Structure of Phonetic Categories. Journal of the Acoustical Society of America, 73: 2124-2133.

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.

Miller, J. L., & Volaitis, L. E. 1989. Effect of Speaking Rate on the Perceptual Structure of a Phonetic Category. Perception & Psychophysics, 46: 505-512.

Nelson, K. 1974. Concept, Word, and Sentence: Interrelations in Acquisition and Development. Psychological Review, 81: 248-267.

Newman, R. 1999. The Use of Rate Normalization to Investigate the Basic Units of Speech. Presentation at the 35th Regional Meeting of the Chicago Linguistic Society, April 23rd, University of Chicago.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. 1994. Speech Perception as a Talker-Contingent Process. Psychological Science, 5: 42-46.

Ohala, J. J. 1999. Acoustic-Auditory Aspects of Speech. Presentation at the 35th Regional Meeting of the Chicago Linguistic Society, April 23rd, University of Chicago.

Perkell, J. S. 1999. A Theory of Speech Motor Control with Supporting Data from Speakers with Normal Hearing and with Profound Hearing Loss. Presentation at the 35th Regional Meeting of the Chicago Linguistic Society, April 23rd, University of Chicago.

Pinker, S., & Prince, A. 1988. On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition. In S. Pinker & J. Mehler (Eds.), Connections and Symbols. Cambridge, MA: Bradford Books/ MIT Press.

Nearey, T. M. 1990. The Segment as a Unit of Speech Perception. Journal of Phonetics, 18: 347-373.

Posner, M. I., & Keele, S. W. 1970. On the Genesis of Abstract Ideas. Journal of Experimental Psychology, 83: 28-38.

Reed, S. K. 1972. Pattern Recognition and Categorization. Cognitive Psychology, 3: 383-407.

Rosch, E. H. 1975. Cognitive Representations of Semantic Categories. Journal of Experimental Psychology, 3: 193-233.

_____ 1978. Principles of Categorization. In E. Rosch & B. Lloyd (Eds.), Cognition and Categorization. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. 1986. On Learning the Past Tenses of English Verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models. Cambridge, MA: Bradford Books/ MIT Press.

Samuel, A. G. 1982. Phonetic Prototypes. Perception & Psychophysics, 31: 307-314.

Sapir, E. 1925. Sound Patterns in Language. Language, 1: 37-51.

Schacter, D., & Church, B. 1992. Auditory Priming: Implicit and Explicit Memory for Words and Voices. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18: 915-930.

Studdert-Kennedy, M. 1976. Speech Perception. In N. J. Lass (Ed.), Contemporary Issues in Experimental Phonetics. New York.

_____ 1980. Speech Perception. Language and Speech, 23: 45-65.

Trubetzkoy, N. S. 1958. Grundzüge der Phonologie (3rd Ed.). Göttingen: Vandenhoek & Ruprecht.

Volaitis, L. E., & Miller, J. L. 1992. Phonetic Prototypes: Influence of Place of Articulation and Speaking Rate on the Internal Structure of Voicing Categories. Journal of the Acoustical Society of America, 92: 723-735.

Walsh, M. A. 1989. On the Nature of Functional Linguistic Units: The Influence of Written Language on Spoken Language Forms. Unpublished dissertation, University of Texas at Austin.

Wittgenstein, L. 1953. Philosophical Investigations. New York: MacMilliam.