



Published in final edited form as:

*Curr Dir Psychol Sci.* 2010 October 1; 19(5): 308–313. doi:10.1177/0963721410383976.

## New Approaches to the Study of Childhood Language Disorders

**Susan Nittrouer** and  
The Ohio State University

**Bruce Pennington**  
University of Denver

### Abstract

Not long ago, poor language skills did not necessarily interfere with the quality of a person's life. Many occupations did not require sophisticated language or literacy. Interactions with other people could reasonably be restricted to family members and a few social or business contacts. But in the 21st century, advances in technology and burgeoning population centers have made it necessary for children to acquire high levels of proficiency with at least one language, in both spoken and written form. This situation increases the urgency for us to develop better theoretical accounts of the problems underlying disorders of language, including dyslexia. Empirical investigations of language-learning deficits largely focus on phonological representations and often ask to what extent labeling responses are "categorical." This article describes the history of this approach and presents some relevant findings regarding the perceptual organization of speech signals—findings that should prompt us to expand our investigations of language disorders.

---

Although considerable progress has been made in understanding the etiology and risk factors associated with disorders of speech, language, and reading acquisition (Pennington & Bishop, 2009), the exact nature of the problem underlying these related conditions is still puzzling. Much research on children with these disorders has reliably shown that something is awry with their phonological representations, but precisely what that problem is remains elusive. Here we suggest that at least one factor impeding progress toward identifying that problem is that our collective notion of phonological representations has fallen out of line with current directions in perceptual psychology. The goal of this paper is to review trends in this area that bear on how we think about phonological representations in typical and atypical spoken and written language development, to help redirect our research efforts. In turn, that redirection should affect how we intervene with children experiencing difficulty learning language.

The term *phonological representations* refers to a broad set of structures in language, including syllables, onsets, rimes, and phonemes. Of these, the last—the phoneme, or phonetic segment as it is realized in the speech signal—is generally studied in experiments of both normal speech perception and language disorders, including those of reading. Typical language users automatically and efficiently (in a rapid manner that demands little of other attentional or cognitive resources) recover phonetic structure when listening to their native language. This suggests that most of us have a keen awareness of that structure. Probably because of just that fact, most accounts of processing limitations in children with language deficits assume that these elements are readily available for recovery in the physical signal of spoken language. Indeed, they are not.

One way to see that individual phonemes are not necessarily easy to recover from the speech stream is to consider the experience of listening to an unfamiliar language. There is usually more to that experience than just failing to understand the words. The speech may sound fast, consisting of not easily separable phonetic segments. That is likely the perceptual

impression of individuals with poor awareness of phonetic structure when listening to their native language, which includes the majority of children with language deficits. It's little wonder that these children have difficulty maintaining attention to what is being said. Once we recognize that phonetic structure is not transparently available in the physical signal, the problems these children face can be better appreciated. For them, phonetic structure is a slippery object to latch onto, like catching fish in a stream with bare hands.

Quite clearly, difficulty discerning phonetic structure in spoken language could explain phonologically based reading disorders, at least for languages with alphabetic orthographies, such as English. Our alphabetic symbols represent individual phonetic segments. However, phonetic structure is also called into service in linguistic processes other than reading. It is used to code signals being deposited in working memory, extending the length of material that can be stored with the help of phonetic coding over that which can be stored without it. As children progress through the school grades, teachers' instructions become longer and more complex. This situation can be increasingly problematic for children who have difficulty recovering phonetic structure from the speech signal. Phonetic structure also permits efficient storage and retrieval in the lexicon of vocabulary items, especially new ones. In school, children are constantly presented with new vocabulary items, and children who are unable to rapidly represent those vocabulary items with a precise phonetic code are likely hindered terrifically in their academic performance. Phonetic structure is what allows us to generate syntactically ordered strings, or sentences, complete with grammatical markers. Efficient recovery of phonetic structure from the speech signal even aids our understanding of speech in noisy environments: Listeners with normal auditory capacities can encounter difficulty understanding speech in noise if they fail to recover precise phonetic representations from the signal. Given all these functions of phonetic structure, it is easy to see why a child who has difficulty recognizing that structure is at risk for language and learning problems beyond dyslexia. Consequently it is important to develop a cogent account of what it takes to recover phonetic structure from the signal. Unfortunately, we still lack such an account.

In this article, we present a brief history of the predominant approach to studying how listeners recover phonetic structure from the speech signal, the process usually termed *speech perception*. We offer explanations for why that approach has likely failed to explain the process completely, and so why the approach has also failed to account for childhood language disorders. We then give an overview of how the general study of speech perception has been expanded in recent years, and lastly suggest ways in which we might enlist that work as a basis for modifying the study of childhood language disorders.

## A brief history of research into speech perception

As described above, speech perception is typically viewed as the recovery of strings of phonetic segments from the acoustic signal. Historically, investigators have examined (a) what acoustic properties are used to make decisions about specific phonetic segments and (b) what values or settings on those properties specify each phonetic label. These acoustic properties are traditionally termed *cues*, defined as temporally brief (several tens of milliseconds long) bits of the spectral array that, when experimentally manipulated, can be shown to affect phonetic labeling (Repp, 1982). For decades, speech perception experiments have involved manipulating these spectro-temporal bits of the signal in a well-controlled manner and measuring how those manipulations influence phonetic labeling. This line of investigation led to the development of another cornerstone of speech perception research: *categorical perception*.

Categorical perception of speech is examined by creating synthetic syllables in which all acoustic properties are held constant across a series of stimuli at settings providing only ambiguous information about phonetic identity, with one exception. That one property (or in a few cases, several perfectly correlated properties) changes in a linear fashion across the series, spanning a range from a setting that disambiguates labeling in favor of one phonetic category to a setting that disambiguates labeling in favor of another phonetic category. So, for example, the vocalic portion of a fricative-vowel syllable might be constructed to be ambiguous between a syllable that begins with 's' and a syllable that begins with 'sh.' The fricative-noise portion of the syllable would then vary in frequency across the continuum from one that is appropriate for 's' to one that is appropriate for 'sh.' All stimuli are presented to listeners multiple times for labeling in a binary choice format (i.e., the only possible responses are *s*-vowel or *sh*-vowel). Subsequently a labeling function, showing the proportion of one or the other category label assigned to each stimulus, is derived from listeners' responses. The tell-tale mark of categorical perception is that labeling functions are nonlinear, meaning that the same size changes in acoustic properties do not always evoke equivalent changes in labeling. Flat regions of the functions are associated with stable categories, while rapidly changing sections are found near the boundaries between those categories. Figure 1 illustrates with dotted lines the labeling functions that would be obtained from an ideal observer. The solid line in this figure shows ideal discrimination results: Discrimination between adjacent pairs of stimuli is near chance in the regions of stable categories and excellent in the boundary region.

Evidence of categorical perception has intrigued psychologists studying a variety of visual and auditory phenomena for decades (Harnad, 1990). In particular, categorical perception of the kind of highly stylized speech stimuli already described came to be viewed as evidence of a uniquely human capacity to recognize speech (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Where individuals with language deficits are concerned, poor performance on categorical-perception tasks is commonly viewed as evidence of deficient phonological representations (Hazan, Messaoud-Galusi, Rosen, Nouwens, & Shakespeare, 2009). In particular, shallower labeling functions are taken to reflect less distinct phonetic categories, and individuals with language impairments often exhibit shallow labeling functions.

There are problems, however, in positing categorical perception as an explanation of anything other than of how attentive a listener is to the particular cue being manipulated. Over the years, investigators have discovered categorical perception for a wide variety of organisms that we would never credit with being able to understand speech, including chinchillas (Kuhl & Miller, 1978) and quail (Kluender, Diehl, & Killeen, 1987). At the same time, studies with people demonstrate that human listeners fail to show categorical perception for speech stimuli in a language other than their native language (MacKain, Best, & Strange, 1981), supporting the anecdotal impressions most of us have when we listen to a foreign language. This is true in spite of the fact that listeners in such studies demonstrate keen sensitivity to the acoustic property being manipulated when it is presented separately from the speech signal (Miyawaki et al., 1975). These collective results suggest that it may have been imprudent of us all along to be attributing so much significance to outcomes of categorical-perception studies in our quest to understand childhood language disorders. Although categorical perception is an interesting and, for many purposes, important observation, its presence or absence can not reliably index how well language is processed by the individual.

## What we have learned from perceptual psychology

While evidence was mounting that one's ability to categorically label stimuli varying systematically on one acoustic property might not be as critical to explaining human speech perception as originally thought, some investigators were busy studying other kinds of perceptual processes for speech. In 1981, a team of scientists tremendously reduced the amount of spectro-temporal detail (i.e., acoustic cues) in the signal by using sine waves to replicate only the center frequencies of the first three formants (Remez, Rubin, Pisoni, & Carrell, 1981). Formants are the resonant frequencies of the vocal tract, and change continuously as the configuration of the vocal tract changes. Figure 2 shows spectrograms of a natural sentence and the sine wave replica of that sentence. Remez and colleagues reported that adult listeners could comprehend sine wave sentences like this one. This finding meant that there is structure of a kind and at a level not represented in the details of the signal that is nonetheless relevant to speech perception. Although the exact significance of this general result for theories of speech perception was not recognized at the time, it would have to be considered in subsequent theories of speech perception.

Not long after that experiment, other investigators showed that infants are sensitive to acoustic structure in their native language before they even utter their first word. As with the study we just discussed, however, the kind of structure involved was not the signal detail associated with acoustic cues. Instead, these researchers found that young children are sensitive to overall spectral shape in their native language (Boysson-Bardies, Sagart, Halle, & Durand, 1986): Long-term spectra, derived from integrating across several seconds of speech, were computed on the babbling productions of 10-month-olds living in four distinct language communities. Those spectra matched the long-term spectra of meaningful speech produced by adults in each respective language community. Thus, these infants were clearly recovering structure that did not match the definition of acoustic cues—that was more “global” in nature—and making the perception–production link before they were saying words.

In 1989, Best, Studdert-Kennedy, Manuel, and Rubin-Spitz designed a study explicitly to ask if listeners' sensitivity to acoustic cues is sufficient to explain how the signal gets organized perceptually. To measure sensitivity, Best et al. examined how well listeners could discriminate brief spectral glides replicating the initial portion of the third formant in 'ra' and 'la' syllables. Each glide was a single sine wave, and they differed from one another along a ten-step acoustic continuum. When combined with the rest of the spectral components that normally make up these syllables, that initial third formant transition specifies whether a syllable is 'ra' or 'la' and consequently fits the definition of an acoustic cue. Listeners also labeled sine wave replicas of the first three formants of the syllables that incorporated the spectral glides used in the discrimination task. Half the listeners were trained to hear the three-tone stimuli as music; the other half were trained to hear the very same stimuli as speech. Labeling functions differed for the two groups, with only listeners trained to hear the stimuli as speech showing categorical perception. Listeners who received music training showed shallow labeling functions with no clear categories. Because discrimination results for the glides were comparable across groups, the difference in labeling could not be attributed to how sensitive listeners were to that isolated cue. Those results led the authors to suggest that listeners perceptually organize signals in distinctive ways when they are recognized as being speech. There is apparently more to speech perception than can be explained by sensitivity to discrete cues.

In some sense, what Best et al. (1989) described was the process of auditory object formation, except that the object being recovered from the waveform was the moving vocal tract. Accordingly, it could be argued that speech perception needs no further explanation

because we already understand the principles underlying auditory object formation. These principles have been described under the rubric of auditory scene analysis (Bregman, 1990) and include observations such as separate spectral streams cohere if they share a common fundamental frequency. In 1994, however, Remez, Rubin, Berns, Pardo, and Lang designed a series of experiments explicitly to examine whether the perceptual organization of speech signals can be explained by these general auditory principles. Using stimuli that carefully precluded these principles from being invoked, the authors observed that listeners nonetheless organized signals in order to recover linguistic structure. Consequently they ended up discounting general auditory principles as being responsible for the perceptual organization of speech. Again it appeared as if listeners organize complex spectral signals in a distinctive manner if those signals are presumed to have been generated by a moving vocal tract.

Beyond obvious advantages to speech perception, there are other benefits accrued by being able to perceptually fuse signal components in a distinctive way when that signal is recognized as speech. For example, Gordon (2000) measured listeners' thresholds in noise for the accurate labeling of first formants derived from the vowels [I] and [e]. In a separate condition, a stable spectral band in roughly the region of the second and third formants was added. Both because that band was ambiguous in its specification of each vowel and because it remained stable across stimuli, it provided no unique information about vowel identity. Nonetheless, listeners were able to recognize vowels accurately at poorer signal-to-noise ratios when it was present. The ability to fuse signal components in speech perception apparently serves to protect against the deleterious effects of noise masking.

## New Approaches

This brief historical review of speech perception research indicates that we can no longer view the human listener as merely a passive recipient of acoustic cues specifying a string of phonemes. There is more to speech perception. Typical listeners use signal components not exclusively affiliated with individual phonetic segments to recover the linguistically significant object, as long as those components can be organized properly. Results congruent with this suggestion emphasize that the job of our sensory systems is to fuse all signal components reaching us in order to create coherent perceptual objects. Accordingly, phonological representations arise when various levels of signal structure, both detailed and more global, are integrated over time. By this view, phonemes are not recoverable as separate entities; rather, phonetic structure emerges from ongoing perceptual processes. Being able to organize signal components in a certain manner is clearly an important skill, yet little effort has been expended investigating how well children with language deficits are able to perform these sorts of perceptual feats. Numerous studies have measured the sensitivity of children or adults with language impairments to the acoustic cues that underlie phonetic labeling, and they have come up empty handed in terms of unequivocal explanation for those impairments (e.g., Hazan et al., 2009). It is time to consider other perceptual processes as possible culprits in the problems faced by these individuals.

First, we need to ask how sensitive listeners with language impairments are to signal structure other than that of traditional acoustic cues. Each method of processing the speech signal preserves and coincidentally eliminates specific kinds of structure. For example, sine wave replicas of speech preserve the spectral undulations arising from long-term, gradual changes in vocal-tract configuration, but largely eliminate brief spectro-temporal cues. Signal-processing techniques need to be selected in experimental paradigms with these considerations in mind, and results need to be interpreted accordingly. Second, we need to examine the abilities of children with language impairments to appropriately fuse all signal components relevant to phonetic perception. Empirical evidence shows that language



experience affects how listeners perceptually organize different kinds of signal structure, even when care is taken to ensure that top-down linguistic influences (i.e., one's knowledge of and ability to use phonotactic and syntactic constraints) are equivalent across groups (Nittrouer & Lowenstein, 2010). Thus there is evidence that these skills in organizing sensory information vary among people, raising the possibility that individuals with language deficits fail either to recover in typical fashion signal structure beyond the level of detail associated with acoustic cues, and/or fail to perceptually integrate signal structure at all levels, both detailed and more global. In fact, one study has reported preliminary evidence that children with language impairments are unable to appropriately organize signal structure that is more global in nature (Johnson, Pennington, Lowenstein, & Nittrouer, in press). More work using these paradigms that involve global structure is needed.

Extending the focus of research on childhood language disorders to look at how both global and detailed structure in the speech signal is integrated could eventually lead to modifications in how language impairments are treated. Most interventions focus explicitly on the phonetic level of linguistic structure, which at first blush seems appropriate given the pervasive difficulty exhibited by these children with tasks involving phonological awareness. However, if that difficulty actually stems from problems recovering structure other than acoustic cues or from difficulty integrating various kinds of signal structure in speech, then the unit used for treatment purposes needs to be broadened. It is difficult to recover structure such as that arising from long-term modulations in vocal tract configurations from brief sections of speech; longer units are required.

## Summary

It is tempting for cognitive and developmental scientists whose main interest is in the acquisition of language or literacy to relinquish concern about the details of phonological development—to simply assume that such development happens and provides the child with phonological representations like phonemes that can then be used in other language processes. But the intricacies of how listeners recover phonologically significant structure turn out to be very important for understanding both typical and atypical spoken and written language acquisition. Relying too heavily on older theories of speech perception to explain this phenomenon is a mistake. It is time that we incorporate alternative experimental approaches such as those we reviewed here into our study of childhood language disorders, and consider the theoretical implications in our accounts of those disorders.

## Acknowledgments

This work was supported by research grants R01 DC-00633 and R01 DC-006237 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, to Susan Nittrouer and NIH grants R01 HD-049027 and P50 HD-027802 to Bruce Pennington. We thank Carol A. Fowler for comments on an earlier draft of this manuscript and Joanna H. Lowenstein for help with figures.

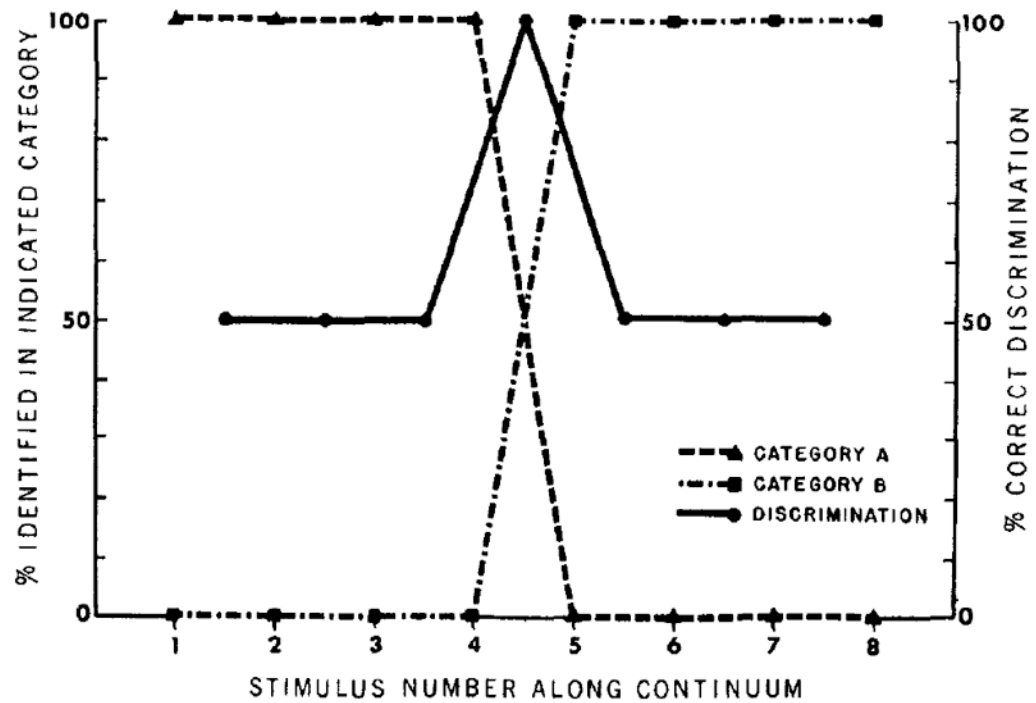
## References

- Best CT, Studdert-Kennedy M, Manuel S, Rubin-Spitz J. Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics*. 1989; 45:237–250. [PubMed: 2710622]
- Boysson-Bardies, B de; Sagart, L.; Halle, P.; Durand, C. Acoustic investigations of cross-linguistic variability in babbling. In: Lindblom, B.; Zetterstrom, R., editors. *Precursors of early speech*. New York, NY: Stockton Press; 1986. p. 113-126.
- Bregman, AS. *Auditory scene analysis*. Cambridge, MA: The MIT Press; 1990.
- Gordon PC. Masking protection in the perception of auditory objects. *Speech Communication*. 2000; 30:197–206.

- Harnad, S. *Categorical perception: The groundwork of cognition*. Cambridge, England: Cambridge University Press; 1990.
- Hazan V, Messaoud-Galusi S, Rosen S, Nouwens S, Shakespeare B. Speech perception abilities of adults with dyslexia: Is there evidence of a true deficit? *Journal of Speech, Language, and Hearing Research*. 2009; 52:1510–1529.
- Johnson EP, Pennington BF, Lowenstein JH, Nittrouer S. Sensitivity to structure in the speech signal by children with speech sound disorder and reading disability. *Journal of Communication Disorders*. in press.
- Kluender KR, Diehl RL, Killeen PR. Japanese quail can learn phonetic categories. *Science*. 1987; 237:1195–1197. [PubMed: 3629235]
- Kuhl PK, Miller JD. Speech perception by the chinchilla: Identification function for synthetic VOT stimuli. *Journal of the Acoustical Society of America*. 1978; 63:905–917. [PubMed: 670558]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- MacKain KS, Best CT, Strange W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*. 1981; 2:369–390.
- Miyawaki K, Strange W, Verbrugge R, Liberman AM, Jenkins JJ, Fujimura O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*. 1975; 18:331–340.
- Nittrouer S, Lowenstein JH. Learning to perceptually organize speech signals in native fashion. *Journal of the Acoustical Society of America*. 2010; 127:1624–1635. [PubMed: 20329861]
- Pennington BF, Bishop DVM. Relations among speech, language, and reading disorders. *Annual Review of Psychology*. 2009; 60:283–306.
- Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM. On the perceptual organization of speech. *Psychological Review*. 1994; 101:129–156. [PubMed: 8121955]
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981; 212:947–949. [PubMed: 7233191]
- Repp BH. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*. 1982; 92:81–110. [PubMed: 7134330]
- Studdert-Kennedy M, Liberman AM, Harris KS, Cooper FS. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*. 1970; 77:234–249. [PubMed: 5454133]

## Suggested Reading

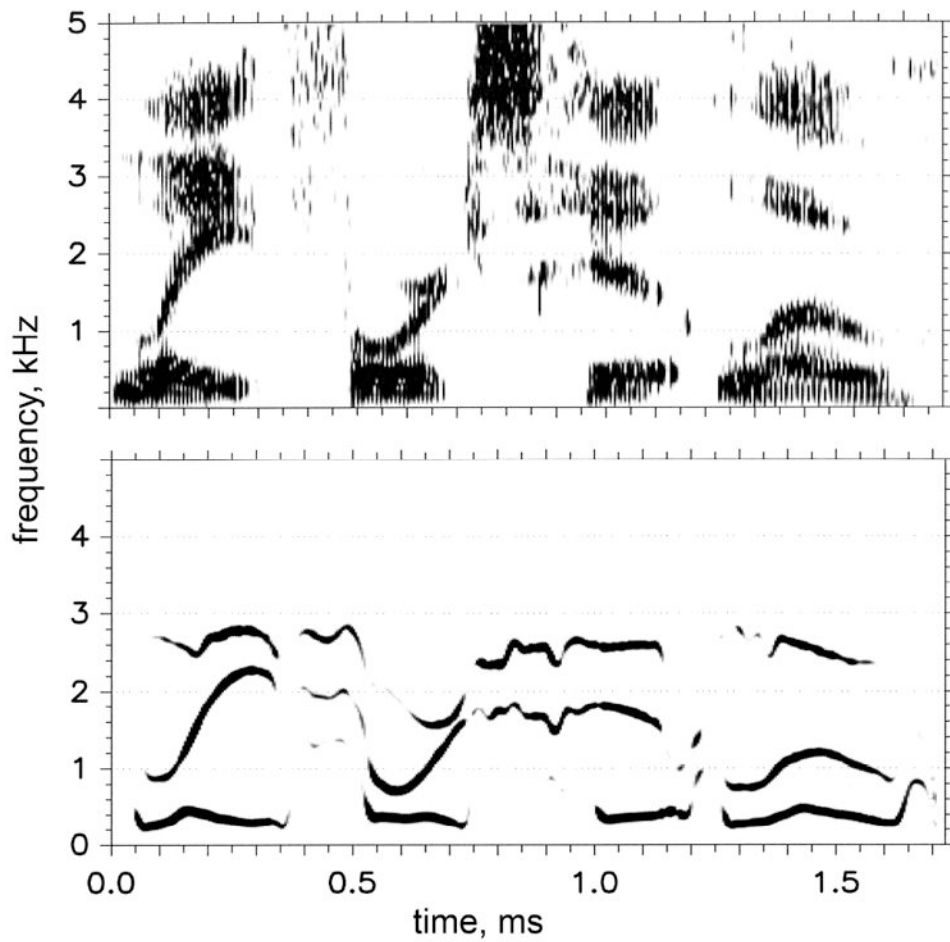
- Galantucci B, Fowler CA, Goldstein L. Perceptuomotor compatibility effects in speech. *Attention Perception & Psychophysics*. 2009; 71:1138–1149.
- Pennington BF, Bishop DVM. 2009 See References.
- Remez, RE. Perceptual organization of speech. In: Pisoni, DB.; Remez, RE., editors. *The handbook of speech perception*. Malden, MA: Blackwell; 2008. p. 28-50.



**FIGURE 1.**

Labeling and discrimination functions from an ideal listener in a categorical-perception experiment. Labeling functions show the proportion of times out of multiple presentations of each stimulus that a specified category label was assigned, assuming a binary choice was provided (i.e., only available labels were Category A or B). The discrimination function shows the proportion of times that adjacent stimuli, presented in series, were correctly recognized as different. From Studdert-Kennedy, Liberman, Harris, and Cooper (1970).





**FIGURE 2.** Spectrogram of a natural production of *Late forks hit low* spoken by a man (top), and as a sine wave replica (bottom).