G+1 6 More Next Blog» Create Blog Sign In

Talking Brains

News and views on the neural organization of language

moderated by Greg Hickok and David Poeppel

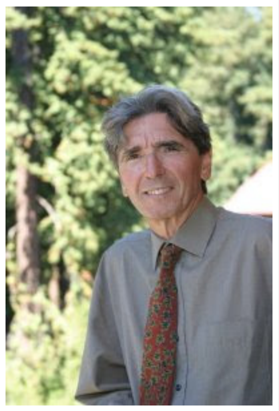
Showing posts sorted by relevance for query **massaro**. Sort by date Show all posts

Tuesday, February 22, 2011

Reflections on the syllable as the perceptual unit in speech perception by Dom Massaro

Given that there have been some interesting debates here on Talking Brains regarding the basic unit of speech perception, I asked Dom Massaro, a prominent and long-time player in this debate, to put together a comment on the topic for publication here. He graciously agreed to do this for us and here it is. Thanks Dom!
-greg

****



Some reminiscences on how I was led to propose the syllable as the perceptual unit in speech perception. I relied mostly on my writings in the literature rather than undocumented memory.
Dom Massaro

During my graduate studies in mathematical and experimental psychology and also during my postdoctoral position, I developed an information-processing approach to the study of behavior (see Massaro & Cowan, 1993, for this brand of information processing). Two important implications arose from this approach: 1) the proximal influences on behavior and 2) the time course of processing are central to a complete description of behavior (as opposed to simple environment-behavior relationships. My early studies involved a delineation of perception and memory processes in the processing of speech and music. The research led to a theory of perception and memory processes that revealed the properties of pre-perceptual and perceptual memory stores and rules for interference of information in these stores and theories of forgetting (Massaro, 1970).

Initiating my career as a faculty member, I looked to apply this information-processing approach to a more substantive domain of behavior. I held a graduate seminar for three years with the purpose of applying the approach to language processing. We learned that previous work in this area had failed to address the issues described above, and our theoretical framework and empirical reviews anticipated much of the research in psycholinguistics since that time in which the focus is on real-time on-line processing (see our book entitled, Understanding Language: An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics, 1975)

My own research interests also expanded to include the study of reading and speech perception. Previous research had manipulated only a single variable in these fields, and our empirical work manipulated multiple sources of both bottom-up and top-down information. Gregg Oden and I collaborated to formulate a fuzzy logical model of perception (Oden & Massaro, 1978; Movellan & McClelland, 2001), which has served as a framework for my research to this day. Inherent to the model were prototypes in memory and, therefore, it was important to take a stance on perceptual units in speech and print. By this time, my research and research by others indicated the syllable and the letter as units in speech and print, respectively. Here is the logic I used.

Speech perception can be described as a pattern-recognition problem. Given some speech input, the perceiver must determine which message best describes the input. An auditory stimulus is transformed by the auditory receptor system and sets up a neurological code in a pre-perceptual auditory storage. Based on my backward masking experiments and other experimental paradigms, this storage holds the information in a pre-perceptual form for roughly 250 ms, during which time the recognition process must take place. The recognition process transforms the

Follow @GregoryHickok

Subscribe to Talking Brains

Blog Moderators

- David Poeppel
- Greg Hickok

Greg Hickok is Professor of Cognitive Sciences at UC Irvine, Editor-in-Chief of Psychonomic Bulletin & Review, and author of The Myth of Mirror Neurons. **David Poeppel,** after several years as Professor of Linguistics and Biology at the University of Maryland, College Park, is now Professor of Psychology at NYU. Hickok and Poeppel first crossed paths in 1991 at MIT in the McDonnell-Pew Center for Cognitive Neuroscience where Hickok was a post doc, and Poeppel a grad student. Meeting up again a few years later at a Cognitive Neuroscience Society Meeting in San Francisco, they began a collaboration aimed at developing an integrated model of the functional anatomy of language. Research in both the Hickok and Poeppel labs is supported by NIDCD.

Links

- iPhod
- iPhod Blog
- ResearchBlogging.org
- Science Blips

Blog Archive

▼ 2015 (50)
▼ December (1)
Max Planck Institute: position for Ph.D. candidate...
► November (2)
► October (8)
► September (7)
► August (8)

pre-perceptual image into a synthesized percept. One issue given this framework is, what are the patterns that are functional in the recognition of speech? These sound patterns are referred to as perceptual units.

One reasonable assumption is that every perceptual unit in speech has a representation in long-term memory, which is called a prototype. The prototype contains a list of acoustic features that define the properties of the sound pattern as they would be represented in pre-perceptual auditory storage. As each sound pattern is presented, its corresponding acoustic features are held in pre-perceptual auditory storage. The recognition process operates to find the prototype in long-term memory which best describes the acoustic features in pre-perceptual auditory storage. The outcome of the recognition process is the transformation of the pre-perceptual auditory image of the sound stimulus into a synthesized percept held in synthesized auditory memory.

According to this model, pre-perceptual auditory storage can hold only one sound pattern at a time for a short temporal period. Backward recognition masking studies have shown that a second sound pattern can interfere with the recognition of an earlier pattern if the second is presented before the first is recognized. Each perceptual unit in speech must occur within the temporal span of pre-perceptual auditory storage and must be recognized before the following one occurs for accurate speech processing to take place. Therefore, the sequence of perceptual units in speech must be recognized one after the other in a successive and linear fashion. Finally, each perceptual unit must have a relatively invariant acoustic signal so that it can be recognized reliably. If the sound pattern corresponding to a perceptual unit changes significantly within different speech contexts, recognition could not be reliable, since one set of acoustic features would not be sufficient to characterize that perceptual unit. Perceptual units in speech as small as the phoneme or as large as the phrase have been proposed.

The phoneme was certainly a favorite to win the pageant for speech’s perceptual unit. Linguists had devoted their lives to phonemes, and phonemes gained particular prominence when they could be distinguished from one another by distinctive features. Trubetzkoy, Jakobson, and other members of the "Prague school" proposed that phonemes in a language could be distinguished by distinctive features. For example, Jakobson, Fant, and Halle (1961) proposed that a small set of orthogonal, binary properties or features were sufficient to distinguish among the larger set of phonemes of a language. Jakobson et al. were able to classify 28 English phonemes on the basis of only nine distinctive features. While originally intended only to capture linguistic generalities, distinctive feature analysis had been widely adopted as a framework for human speech perception. The attraction of this framework is that since these features are sufficient to distinguish among the different phonemes, it is possible that phoneme identification could be reduced to the problem of determining which features are present in any given phoneme. This approach gained credibility with the finding, originally by Miller and Nicely (1955) and since by many others, that the more distinctive features two sounds share, the more likely they are to be perceptually confused for one another. Thus, the first candidate we considered for the perceptual unit was the phoneme.

Consider the acoustic properties of vowel phonemes. Unlike some consonant phonemes, whose acoustic properties change over time, the wave shape of the vowel is considered to be steady-state or tone-like. The wave shape of the vowel repeats itself anywhere from 75 to 200 times per second. In normal speech, vowels last between 100 and 300 ms, and during this time the vowels maintain a fairly regular and unique pattern. It follows that, by our criteria, vowels could function as perceptual units in speech.

Next let us consider consonant phonemes. Consonant sounds are more complicated than vowels and some of them do not seem to qualify as perceptual units. We have noted that a perceptual unit must have a relatively invariant sound pattern in different contexts. However, some consonant phonemes appear to have different sound patterns in different speech contexts. For example, the stop consonant phoneme /d/ has different acoustic representations in different vowel contexts. Since the steady-state portion corresponds to the vowel sounds, the first part, called the transition, must be responsible for the perception of the consonant /d/. The acoustic pattern corresponding to the /d/ sound differs significantly in the syllables /di/ and /du/. Hence, one set of acoustic features would not be sufficient to recognize the consonant /d/ in the different vowel contexts. Therefore, we must either modify our definition of a perceptual unit or eliminate the stop consonant phoneme as a candidate.

There is another reason why the consonant phoneme /d/ cannot qualify as a perceptual unit. In the model perceptual units are recognized in a successive and linear fashion. Research has shown, however, that the consonant /d/ cannot be recognized before the vowel is also recognized. If the consonant were recognized before the vowel, then we should be able to decrease the duration of the vowel portion of the syllable so that only the consonant would be recognized. Experimentally, the duration of the vowel in the consonant-vowel syllable (CV) is gradually decreased and the subject is asked when she hears the stop consonant sound alone. The CV syllable is perceived as a complete syllable until the vowel is eliminated almost entirely (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). At that point, however, instead of the perception changing to the consonant /d/, a nonspeech whistle is heard. Liberman et al. show that the stop consonant /d/ cannot be perceived independently of perceiving a CV syllable. Therefore, it seems unlikely that the /d/ sound would be perceived before the vowel sound; it appears, rather, that the CV syllable is perceived as an indivisible whole or gestalt.

- ► July (2)
- ► June (4)
- ► May (2)
- ► April (5)
- ► March (2)
- ► February (1)
- ► January (8)

- ► 2014 (68)
- ► 2013 (35)
- ► 2012 (41)
- ► 2011 (92)
- ► 2010 (107)
- ► 2009 (110)
- ► 2008 (158)
- ► 2007 (72)

- Home
- Announcements
- Job Postings

These arguments led to the idea that the syllables function as perceptual units rather than containing two perceptual units each. One way to test this hypothesis is to employ the CV syllables in a recognition-masking task. Liberman et al., found that subjects could identify shortened versions of the CV syllables when most of the vowel portion is eliminated. Analogous to our interpretation of vowel perception, recognition of these shortened CV syllables also should take time. Therefore, a second syllable, if it follows the first soon enough, should interfere with perception of the first. Consider the three CV syllables /ba/, /da/, and /ga/ (/a/ pronounced as in father), which differ from each other only with respect to the consonant phoneme. Backward recognition masking, if found with these sounds, would demonstrate that the consonant sound is not recognized before the vowel occurs and also that the CV syllable requires time to be perceived.

There have been several experiments on the backward recognition masking of CV syllables (Massaro, 1974, 1975; Pisoni, 1972). Newman and Spitzer (1987) employed the three CV syllables /ba/, /da/, and /ga/ as test items in the backward recognition masking task. These items were synthetic speech stimuli that lasted 40 ms; the first 20 ms of the item consisted of the CV transition and the last 20 ms corresponded to the steady-state vowel. The masking stimulus was the steady-state vowel /a/ presented for 40 ms. In one condition, the test and masking stimuli were presented to opposite ears, that is, dichotically. All other procedural details followed the prototypical recognition-masking experiment.

The percentage of correct recognitions for 8 observers improved dramatically with increases in the silent interval between the test and masking CVs. These results show that recognition of the consonant is not complete at the end of the CV transition, nor even at the end of the short vowel presentation. Rather, correct identification of the CV syllable requires perceptual processing after the stimulus presentation. These results support our hypothesis that the CV syllable must have functioned as a perceptual unit, because the syllable must have been stored in pre-perceptual auditory storage, and recognition involved a transformation of this pre-perceptual storage into a synthesized percept of a CV unit. The acoustic features necessary for recognition must, therefore, define the complete CV unit. An analogous argument can be made for VC syllables also functioning as perceptual units (Massaro, 1974).

We must also ask whether perceptual units could be larger than vowels, CV, or VC syllables. Miller (1962) argued that the phrase of two or three words might function as a perceptual unit. According to our criteria for a perceptual unit, it must correspond to a prototype in long-term memory which has a list of features describing the acoustic features in the pre-perceptual auditory image of that perceptual unit. Accordingly, pre-perceptual auditory storage must last on the order of one or two seconds to hold perceptual units of the size of a phrase. But the recognition-masking studies usually estimate the effective duration of pre-perceptual storage to be about 250 ms. Therefore, perceptual units must occur within this period, eliminating the phrase as the perceptual unit.

The recognition-masking paradigm developed to study the recognition of auditory sounds has provided a useful tool for determining the perceptual units in speech. If preperceptual auditory storage is limited to 250 ms, the perceptual units must occur within this short period. This time period agrees nicely with the durations of syllables in normal speech.

The results of the present experiments demonstrate backward masking in a two-interval forced-choice task, a same-different task, and an absolute identification task. The backward masking of one sound by a second sound is interpreted in terms of auditory perception continuing after a short sound is complete. A representation of the short sound is held in a preperceptual auditory storage so that resolution of the sound can continue to occur after the stimulus is complete. A second sound interferes with the storage of the earlier sound interfering with its further resolution The current research contributes to the development of a general information processing model (Massaro, 1972, 1975).

To solve the invariance problem between acoustic signal and phoneme, while simultaneously adhering to a pre-perceptual auditory memory constraint of roughly 250 ms, Massaro (1972) proposed the syllables V, CV, or VC as the perceptual unit, where V is a vowel and C is a consonant or consonant cluster. This assumption was built into the foundation of the FLMP (Oden & Massaro, 1978). It should be noted that CVC syllables would actually be two perceptual units, the CV and VC portions, rather that just one. Assuming that this larger segment is the perceptual unit reinstates a significant amount of invariance between signal and percept. Massaro and Oden (1980, pp. 133–135) reviewed evidence that the major coarticulatory influences on perception occur within these syllables, rather than between syllables. Any remaining lack of invariance across these syllables could conceivably be disambiguated by additional sources of information in the speech stream.

References

Massaro, D.W. (1970). Perceptual Processes and Forgetting in Memory Tasks. Psychological Review, 77(6), 557-567.

Massaro, D.W. (1972). Preperceptual Images, Processing Time, and Perceptual Units in Auditory Perception. Psychological Review, 79(2), 124-145.

Massaro, D. W. (1974). Perceptual Units in Speech Recognition. Journal of Experimental Psychology, 102(2), 349-353.

Massaro, D.W. (1975). Understanding Language: An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics. New York: Academic Press.

Massaro, D.W. and Cowan, N. (1993). Information Processing Models: Microscopes of the Mind. Annual Review of Psychology, 44, 383-425.
http://mambo.ucsc.edu/papers/1993.html

Massaro, D. W. & Oden, G. C. (1980). Speech Perception: A Framework for Research and Theory. In N.J. Lass (Ed.), Speech and Language: Advances in Basic Research and Practice. Vol. 3, New York: Academic Press, 129-165.

Movellan, J., and McClelland, J. L. (2001). The Morton-Massaro Law of Information Integration: Implications for Models of Perception. Psychological Review, 108, 113-148.

Posted by Greg Hickok at 11:01 AM 9 comments:

Saturday, March 8, 2008

Keeping up with the Jones-Hickoks (TB West)

Between the semantic dementia course -- Thank You, Greg! -- and the travelogue, I can barely keep up with all the reading ... Greg, you are a good citizen, and I have been a slacker.

One of us, DP from TB_East, attended two curious meetings recently. Here 's a little update on that.

AAAS in Boston: This meeting is largely for the media, apparently over 900 journalists attended. There were a few sessions that were relevant to our research interests. Phil Rubin from Haskins chaired a session on language technologies which included Dominic Massaro (*Talking Faces*) and Justine Cassell (Northwestern University). Massaro presented the work with Baldi, the talking head -- which is a cool tool to investigate audio-visual speech but seems a little bit behind-the-times in terms of state-of-the art animation and visualization. Given the quality of animation in current cinema, it should be possible to generate analytically precisely specified faces that give more realistic/naturalistic output. That being said, Massaro has been a leading figure in the investigation of audiovisual speech perception, and (whether one likes his Baldi figure or not) anyone studying AV speech is certainly (or should be) aware of how Massaro's FLMP model handles multi-sensory integration. Justine Cassell presented some provocative data on how children interact with avatar-style computerized friends onscreen. She applied her ideas about 'embodied conversational agents' to the interaction between autistic chuldren and the onscreen partner. A little puzzling but fascinating. The work is not yet published but bstay tuned.

I chaired a session on brain and speech that had three interesting talks. First, Pat Kuhl presented her program of research on language development/speech perception, the highlight being the new baby MEG scanner that Pat apparently convinced the Finnish MEG manufacturer to build. Pictures of babies in an MEG machine ... how can you go wrong? I am looking forward to seeing the new data coming from this approach. Jack Gandour presented a lot of data on the neural basis of tone language perception and comprehension. Jack is arguably the world's leading expert on the cognitive neuroscience of tone languages, and a 30 minute presentation cannot do justice to the huge range of data he has on these issues. Finally, former TB_East graduate student Nina Kazanina presented some of her recent work, published last year in PNAS. Nina's paper (with TB_East faculty Bill Idsardi and Colin Phillips) is called The influence of meaning on the perception of speech sounds and uses a clever cross-linguistic design (Korean, Russian) in the context of a mismatch study to test how native phonology shapes early auditory responses. Nina is now on the faculty of the University of Bristol, and we are all very proud of her.

The session in Boston that really got my blood pressure high was called *The mind of a tool maker*, and concerned -- allegedly -- the evolution of language and cognition. A very high-powered cast, a terrible session. The cast: Lewontin, Berwick, Walsh, Hauser, Deacon, as well as some other folks I did not know, and whose performance did not make me wat to run out and read their work (e.g. Mimi Lam, Dean Falk). There were, to be sure, some sensible ideas buried in there, and one genuinely good talk, by Marc Hauser. Among other reasons it stood out as good (contrast enhancement) because (a) he stayed within his alloted time (b) the talk had a point/hypothesis (c) the work actually related to the topic of the session. Berwick had an interesting idea about FoxP2, a really nice deconstruction/debunking based on a computational analysis, and Deacon presented some interesting ideas -- but too many and too scattered. But the bottom line is this: the study and discussion of evolution of cognition and language requires extreme caution, subtlety, rigor, nuance, a high-pass filter for bullshit, and so on and so forth. And, alas, the level of speculation and pure unadulterated paleo-nonsense was off the scale. This session made me appreciate why the French Academy forbade language evolution as a topic. The audience deserved better. My favorite line: the organizer of the workshop, Dr. Lam, in her opening remarks, said that one reason she wanted to have this workshop was because she had such a hard time getting her ideas on evolution of cognition published .... Yikes!

Posted by David Poeppel at 7:47 PM 3 comments:
Labels: commentary

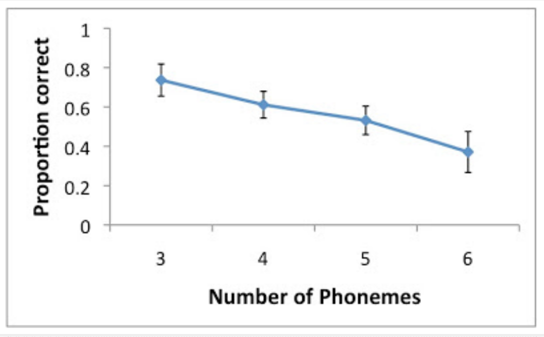
Wednesday, August 11, 2010

Importance of phonemes in speech production

In a previous post I have questioned whether we need to explicitly represent phonemes in speech perception. Massaro and others have raised this issue in the past. Phonemes, the line of thinking goes, are only really important for production. There are linguistic arguments for this that I won't detail here. There is also well-known speech error data which shows that phoneme size units can break off and dislocate themselves. Here I want to highlight some evidence from aphasia. A reviewer of one of my papers pointed me to this study by Lindsey Nickels and David Howard.

A group of aphasics who exhibited speech production errors were asked to repeat words that varied in terms of the number of phonemes, number of syllables, or syllable complexity (defined in terms of consonant clusters). These variables are, of course, highly correlated, but the stimuli were carefully designed so that the contribution of each of these factors could be examined using logistic regression analyses.

The main result was that only number of phonemes in a word predicted correct repetition (see graph below derived from their Table 4) and once this variable was taken into account, the number of syllables or syllable complexity did not explain any additional variance.



Phonemes seem to matter in speech production. I have to say, though, that I'm not fully convinced that the others factors aren't also important.

Nickels, L., & Howard, D. (2004). Dissociating Effects of Number of Phonemes, Number of Syllables, and Syllabic Complexity on Word Production in Aphasia: It's the Number of Phonemes that Counts *Cognitive Neuropsychology, 21* (1), 57-78 DOI: 10.1080/02643290342000122

Posted by Greg Hickok at 9:25 AM 7 comments:

Tuesday, August 25, 2009

"Categorical perception" in neuroscience studies of speech

> Old speech phenomena don't die they just become morphed into neuroscience studies.
> -Andrew Lotto

The phenomenon of categorical perception appears to be riding the coattails of the resurgence of interest in motor theories of speech perception. Back in the motor theory heyday, categorical perception was all the rage. Listeners appeared to perceive speech sounds differently from non-speech sounds, i.e., categorically, and this was taken as evidence for the motoric nature of the speech perception process. The argument was something like this... Acoustic signals vary continuously. Articulatory patterns are categorical (/b/ is always produced bilabially). Perception mirrors the categorical nature of articulation. Therefore we perceive speech via our motor system.

Problems with this view quickly arose. Non-human, and therefore non-speaking, animals such as chinchillas and quail, were found to exhibit categorical perception for speech sounds. Babies too, who hadn't yet acquired the ability to articulate speech, also exhibited categorical perception. Categorical perception of non-speech sounds was also demonstrated. Further, perception of speech sounds was found to be continuous if listeners were asked to rate how well a stimulus represented a given category rather than asking them to make a binary decision.

Interest in categorical perception (CP) faded -- except in neuroscience where the pace of CP studies seems to be accelerating. Here's just a few from this year:

Möttönen R, Watkins KE. Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci. 2009 Aug 5;29(31):9819-25.

Salminen NH, Tiitinen H, May PJ. Modeling the categorical perception of speech sounds: A step toward biological plausibility. Cogn Affect Behav Neurosci. 2009 Sep;9(3):304-13.

Clifford A, Franklin A, Davies IR, Holmes A. Electrophysiological markers of categorical perception of color in 7-month old infants. Brain Cogn. 2009

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. Neural correlates of categorical perception in learned vocal communication. Nat Neurosci. 2009 Feb;12(2):221-8.

I hinted previously that the failure to use signal detection analysis methods in the context of categorical perception studies may have contaminated the whole field of CP research. Lori Holt recently pointed me to a paper by Schouten et al. 2003, provocatively titled "The End of Categorical Perception as We Know It". The point of the paper is exactly was I was hinting at: perception only looks categorical because of inherent bias in the tasks used to measure it.

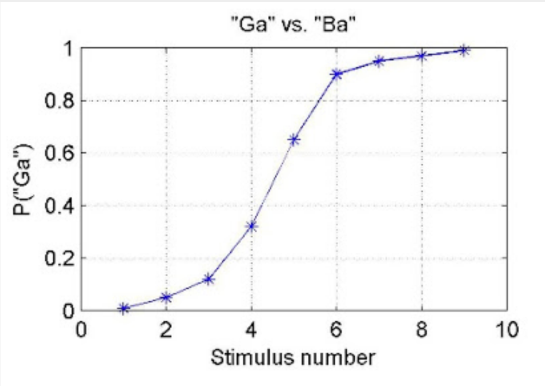
> The traditional categorical-perception experiment measures the bias inherent in the discrimination task
> (Schouten et al. 2003, p. 71)

Here's another interesting quote from this paper:

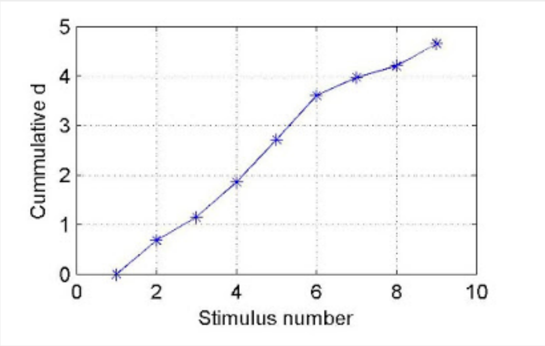
> Despite an auspicious beginning with a clear experimental definition ... categorical perception has in practice remained an ill-defined or even undefined concept, which could be used to underpin a variety of sometimes mutually exclusive claims, for example for or against the motor theory (p. 72)

This is an interesting paper that is worth a close look. But back to bias...

Let me illustrate very simply using some categorical perception data that I pulled from the literature. The graph below shows real data from a CP experiment using a GA-DA continuum. The task is explicitly categorical: subjects are asked to decide whether a stimulus is an example of GA or DA. This is not a good task to determine whether subjects perceive speech sounds categorically because it forces them to categorize. As Schouten et al. put it, "... if the nature of the task compels subjects to use a labelling strategy, categorical perception will be pretty much a foregone conclusion" (p. 77). Nonetheless, use of d-prime measures shows a rather different picture to standard measures. The vertical access is proportion of GA responses, and the horizontal axis is the various stimuli along the continuum. Perception looks nicely categorical.



Now plot the same data in d-prime units. To do this you can calculate d' for each pair of adjacent stimuli (how well are Ss discriminating Stim1 from Stim2, Stim2 from Stim3, etc.). Plotted here is cumulative d'. We should see discontinuities in the cumulative d'. Instead we see a more continuous function.



Have a look at the papers by Lori Holt and Andrew Lotto that I highlighted in a previous post as well as the Schouten et al. paper for more critical views on the nature of categorical perception. Then there's always long-time CP skeptic Dominic Massaro. His work on the topic is also worth a look.

What are the implications for neuroscience studies of speech perception? Well, if CP is nothing more than task effects and/or subject bias, then by using CP paradigms to map speech perception systems, all that is being mapped is task strategies and/or subject bias. No wonder all these studies find effects in the frontal lobe!

Schouten, B. (2003). The end of categorical perception as we know it *Speech Communication, 41* (1), 71-80 DOI: 10.1016/S0167-6393(02)00094-8

Posted by Greg Hickok at 11:48 AM 25 comments:
Labels: commentary

Home

Subscribe to: Posts (Atom)

Simple template. Powered by Blogger.