Categorical Perception

- Discrimination for some speech contrasts is poor within phonetic categories and good between categories.
- Unusual, not found for most perceptual contrasts.
- Influenced by task, expectations, memory and quality of stimuli.

Categorical Perception

In "normal" perception, a perceiver can discriminate many more stimuli than they can absolutely identify. Discrimination performance is generally proportional to the absolute magnitude of the stimulus. This is Weber's Law.

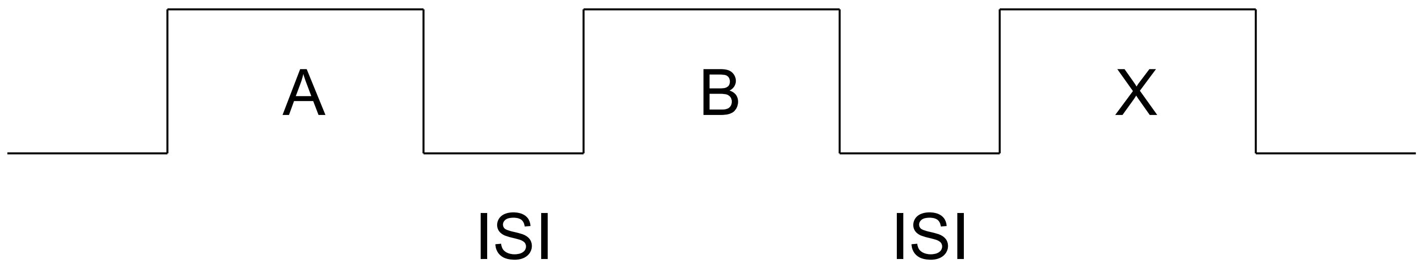
In categorical perception, a perceiver can only discriminate between stimuli to the extent that they have different labels.

Basic Tasks

Listeners perform two tasks with a stimulus continuum (e.g. a /ba/ - /pa/ VOT series): Identification and discrimination.

1. Identification - Present tokens in random order. Collect labeling responses.
2. ABX or AXB Discrimination - Each trial has three tokens. A and B are always different and X is either A or B. A and B are 1 or 2 or 3 stimuli apart (e.g. separated by 20 msec VOT).

ABX Trials



On a trial, stimuli A and B are always different. X is identical to either A or B. The ISI is typically short (50 to 500 ms). A and B change from trial to trial. The listeners task is to indicate if X was identical to A or to B.

ABX Trials - 2

To examine discrimination along a continuum, on different trials, comparison is made between A and B at different points along the continuum. For example, if our continuum is a VOT continuum from a 0 ms VOT /ba/ to a 60 ms VOT /pa/, we might compare discrimination for stimuli that are 20 msec VOT apart along the continuum.

Our comparisons would be 0 and 20, 10 and 30, 20 and 40, 30 and 50, and 40 and 60 ms VOT.

ABX Trials - 3

The order of stimuli within a trial is counterbalanced. For each comparison of A and B, there are four trials: ABA, ABB, BAA, BAB.

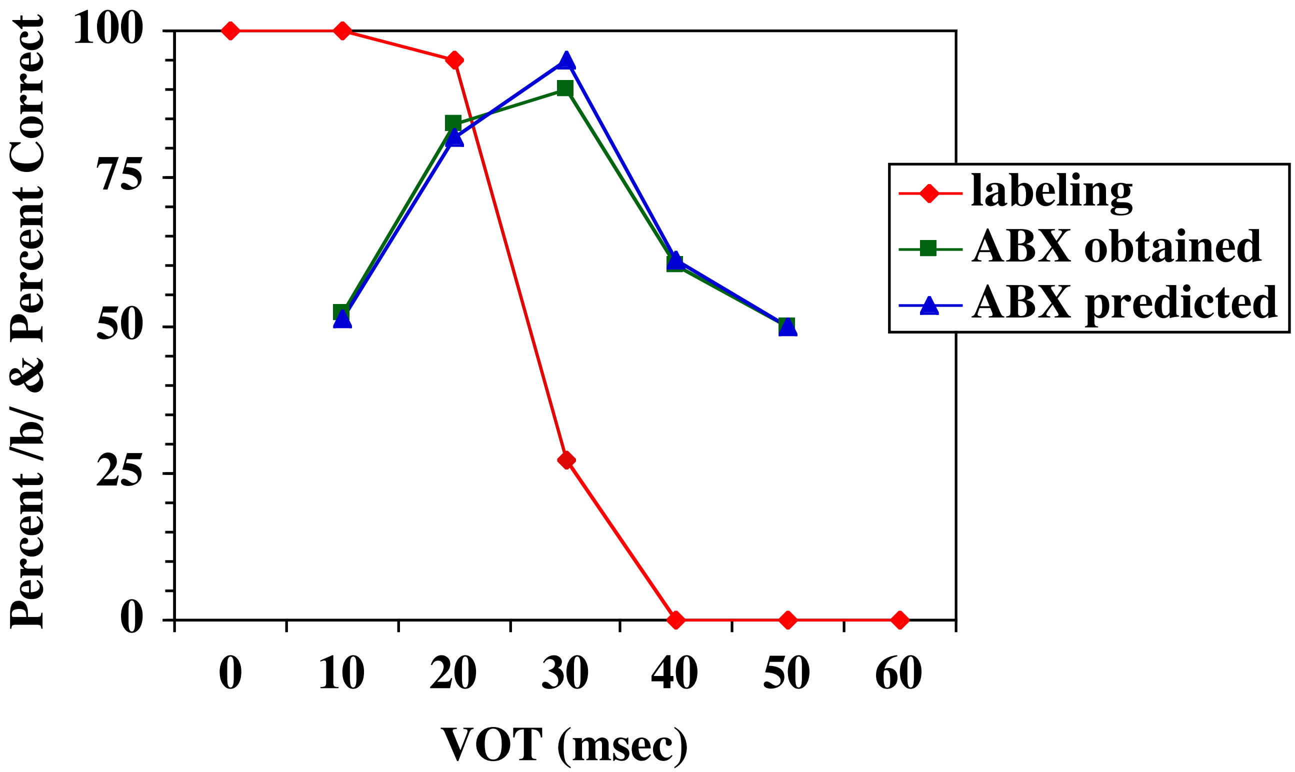
To ensure that we have stable data for each participant, we typically collect 20 or more labeling responses for each stimulus and 20 or more discrimination responses for each stimulus pair (A and B).

Basic Results

Listeners can discriminate between two different tokens only to the extent that they give them different labels.

For example, 40 and 60 msec VOT tokens are both labeled /pa/ on virtually all trials and listeners are at chance in discrimination for 40 and 60 msec pairs.

The 20 and 40 msec tokens are in different categories (/ ba/ and /pa/) and listeners are near perfect in discriminating between them.


Percent /b/ & Percent Correct
100
75
50
25
0
0
10
20
30
40
50
60
VOT (msec)
labeling
ABX obtained
ABX predicted

Prediction Formula

$0.5 + 0.5 \times (p(L_1) - p(L_2))^2$

For any two tokens L_1 and L_2, find the difference in labeling, square it, multiply by .5 and add to .5 (chance). This is the predicted correct discrimination performance.

If L_1 is 1.0 and L_2 is 0.0, the predicted discrimination performance is 1.0 while if L_1 and L_2 are the same, discrimination is predicted to be 0.5 (chance).

Further Data

Categorical perception does not happen for all speech contrasts. Stop-consonants varying in place or VOT show it. Vowels do not.

Putting the target in context makes perception more categorical. Making the targets shorter (vowels) makes them more categorical. Making the time interval between tokens in the ABX triad longer makes the results more categorical.

Further Data - 2

With training, listeners perform much better at within category discriminations (see Samuel, 1977).

Other discrimination tasks (4IAX) yield better performance (Pisoni, 1973).

What would happen if the stimuli were made more natural by co-varying multiple acoustic cues?

Further Issues

Labeling data collected with one token presented at a time are not going to be exactly the same as when listeners hear three tokens in sequence. For this reason, some researchers have used an AXB task.

In labeling and discrimination, listeners hear the same AXB sequences. In labeling, they label the middle (X) token. In discrimination, they indicate whether it matched the first (A) or third (B) token.

Further Issues - 2

Does categorical perception reflect memory limitations for rapidly changing stimuli?

Does it reflect discontinuities in perception? Put another way, could there be specialized neural coding that introduces a "threshold" into how complex information such as VOT is coded.

What does it tell us about the nature of speech perception?

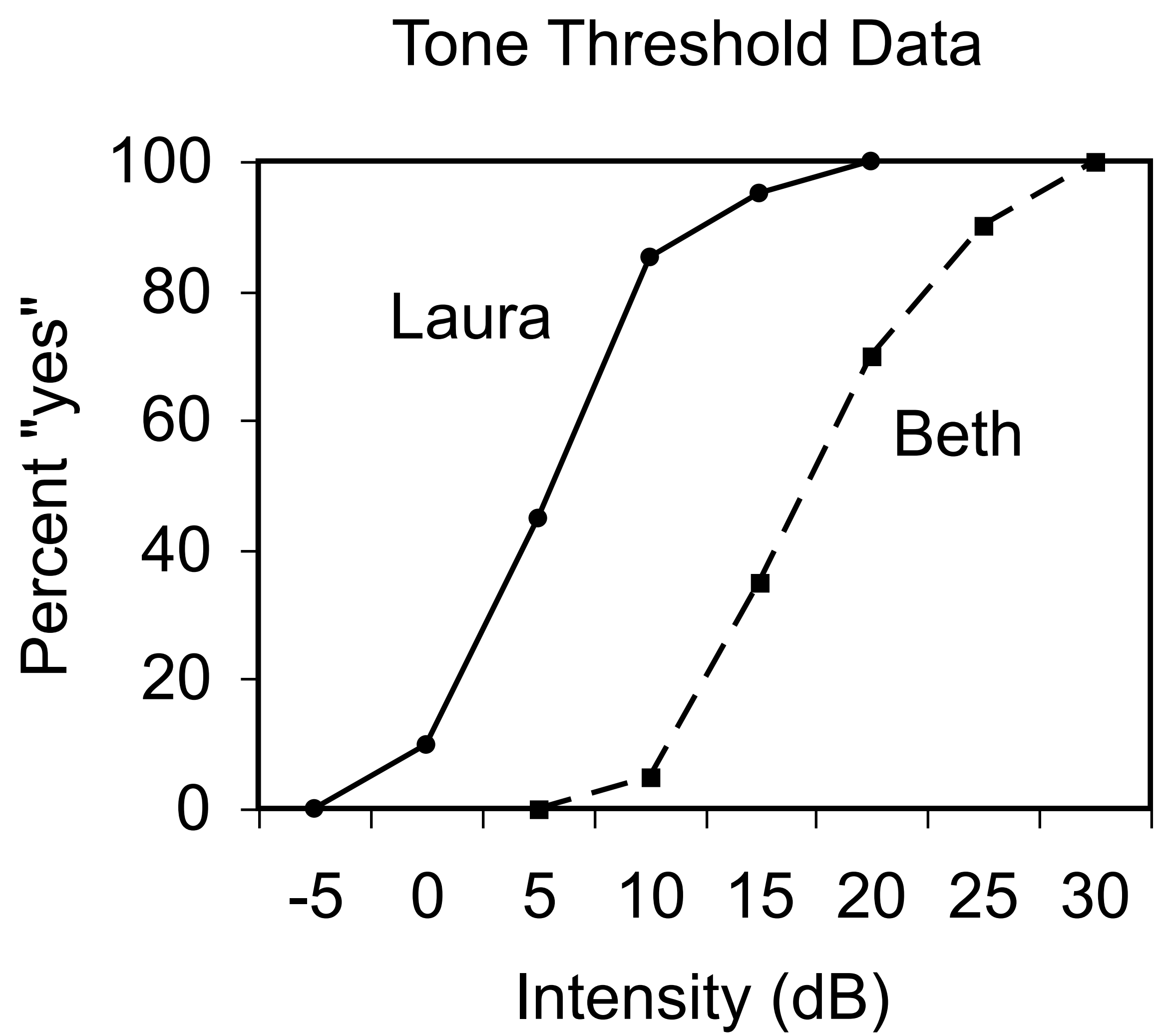
An Aside

Before going further, we need to introduce the concepts of Signal Detection Theory so that they can be used to help understand categorical perception.

Signal Detection Theory is designed to separate the performance of a perceiver in a task into two components: the underlying sensory/perceptual coding processes and response (decision) processes that map internal coding onto an external response.

To see why this is important, consider the performance of two listeners in a hearing test.

Listeners Laura and Beth.
Are their perceptual systems really as different as they appear?



Signal Detection Theory (SDT)

The Problem - There is no way to tell, using a simple threshold task, whether differences in listener data (e.g. Laura and Beth) reflect differences in perception or their interpretation of the instructions.

The Theory of Signal Detectability - TSD attempts to separate those factors that reflect decision processes (such as instructions, motivation) from those that reflect underlying perception.

Terminology

Decision Criterion (Bias) - The rule that a listener uses to map their sensory/perceptual evidence onto the response alternatives. Should be influenced by motivation, interpretation of instructions, payoffs for various outcomes.

Sensitivity - The underlying sensory and perceptual processes that transform the stimuli in the world into internal mental states. Influenced by the stimulus and basic perceptual processes.

Trial Structure

In some psychophysical methods, a stimulus is presented on every trial. In principle, the subject could say yes on every trial and be correct. The problem is that we have no objective way of knowing whether a "yes" or "no" response was correct.

We need to add trials on which "yes" and "no" responses are objectively correct. Put more generally, we need different trials with distinct, objectively correct, responses.

Trials with Different Correct Answers

		Signal	
		S_2	S_1
Response	Yes	Hit	False Alarm
	No	Miss	Correct Rejection

Laura and Beth

Return to our example of the hearing threshold. We present the -5 dB tone and silence to our two listeners.

For Laura, we get 90% hits and 40% false alarms.

For Beth, we get 50% hits and 6% false alarms.

Laura clearly has a higher *hit* percentage than Beth, but she also has a higher percentage of *false alarms*. How do we make precise comparisons between these two individuals?

An Experiment

According to the basic idea of a threshold, Laura has a lower threshold than Beth. If they are interpreting the instructions differently, then we should be able to change their performance by modifying the instructions.

Offer Beth a monetary incentive for Hits. The payoffs are:

Hit	Win $10
Miss	Lose $1
False Alarm	Lose $1
Correct Rejection	Win $1

With these payoffs, Beth has 98% hits and 70% false alarms.

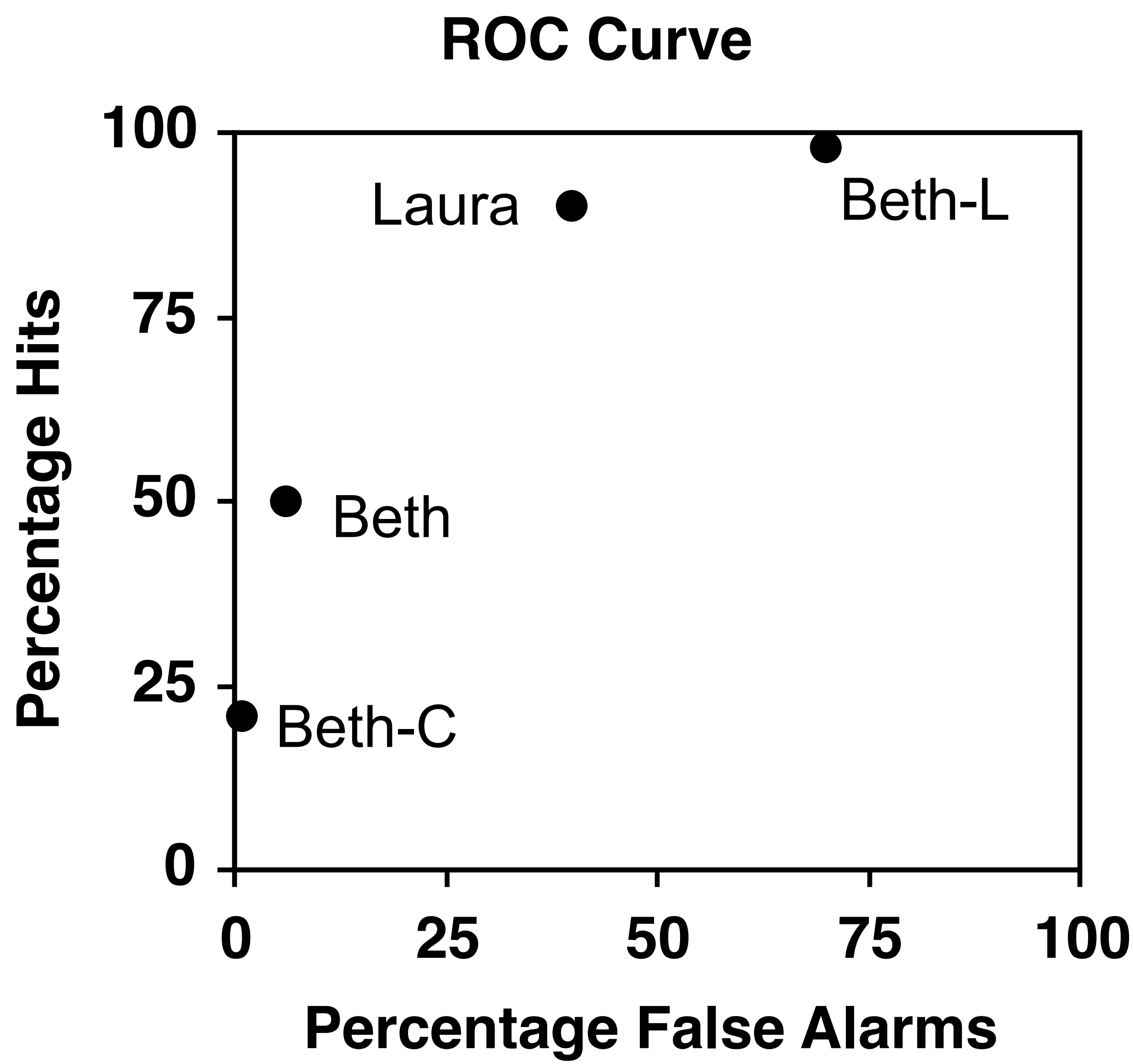
Part 2

Next, we change the payoffs for Beth to:

Hit	Win $1
Miss	Lose $1
False Alarm	Lose $10
Correct Rejection	Win $10

Now, she has 21% hits and 1% false alarms.

The first payoff matrix is called liberal since it pays Beth to say "yes" a lot. The second is called conservative (pays Beth to say "no" a lot). Beth has changed her performance across the three conditions. As the hits increased, so did false alarms. As hits decreased, so did false alarms.


ROC Curve
Percentage Hits
Percentage False Alarms
0
25
50
75
100
Laura
Beth-L
Beth
Beth-C

Interpretation

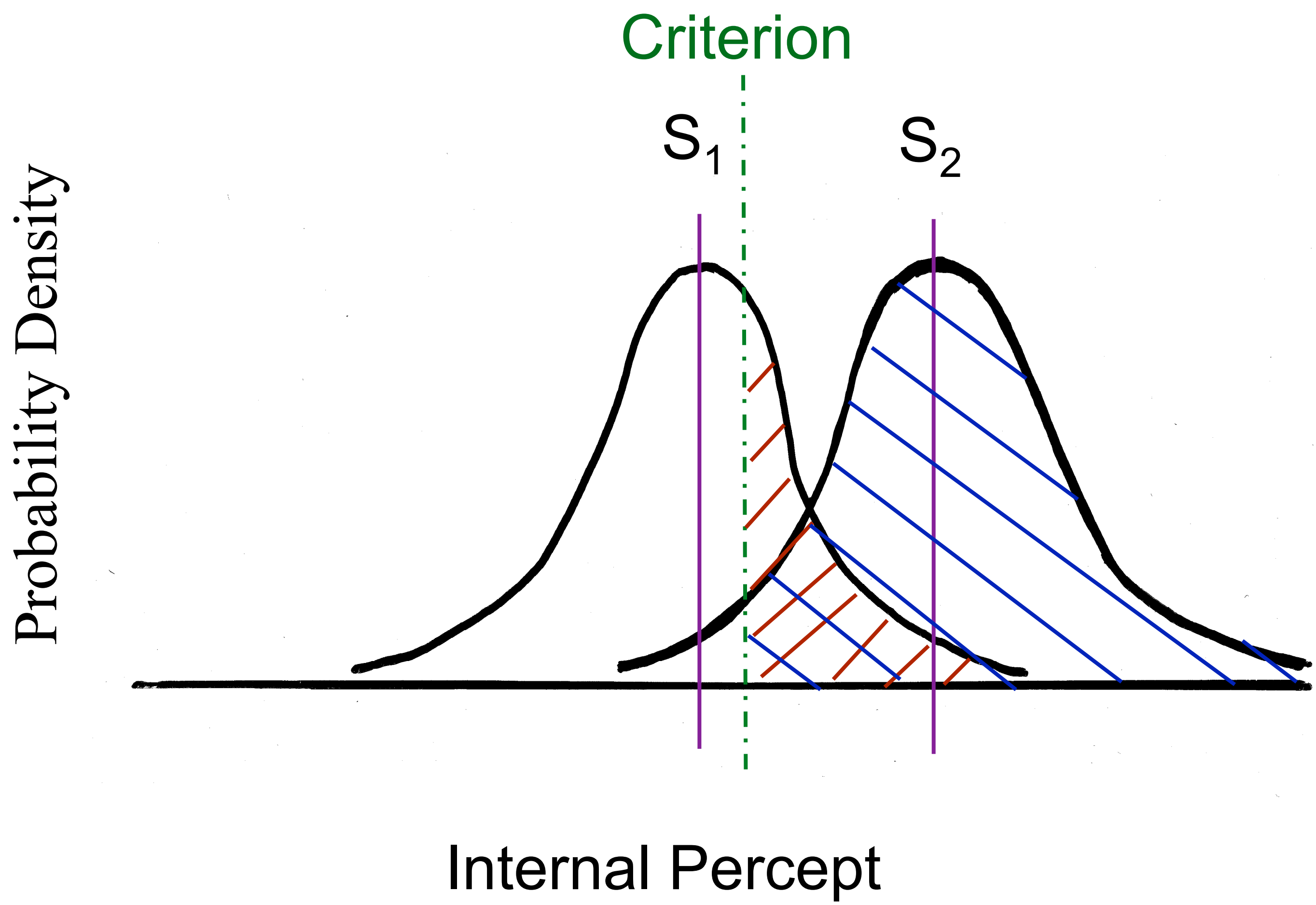
In the plot of Hits versus False Alarms, the three points for Beth's data and Laura's data seem to lie on the same smooth curve. Does this imply that the only difference between Beth and Laura is in their interpretation of the instructions, motivation, etc?

To answer, we need a quantitative theory.

The Theory

On an S_1 trial (signal absent or noise trial), the stimulus information gives rise to an internal percept of a particular magnitude. On different presentations of this trial, the internal magnitude of the sensation or percept will not always be the same (noise in the processing system). The "sensation" will be distributed normally.

On S_2 trials (signal present or signal trials), the stimulus information will also give rise to an internal percept. This will differ from S_1 in magnitude on the internal perceptual dimension. The magnitude will be variable from trial to trial and will be distributed normally.



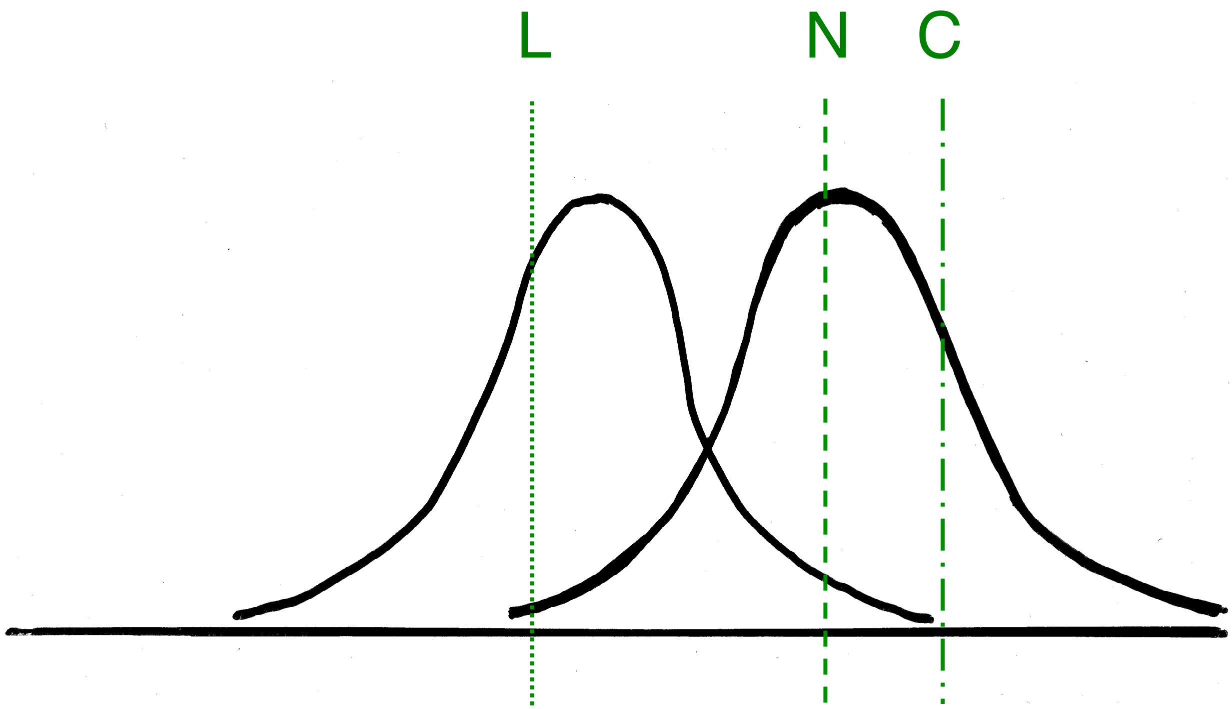
Criterion
S1
S2
Probability Density
Internal Percept

Theory and Data

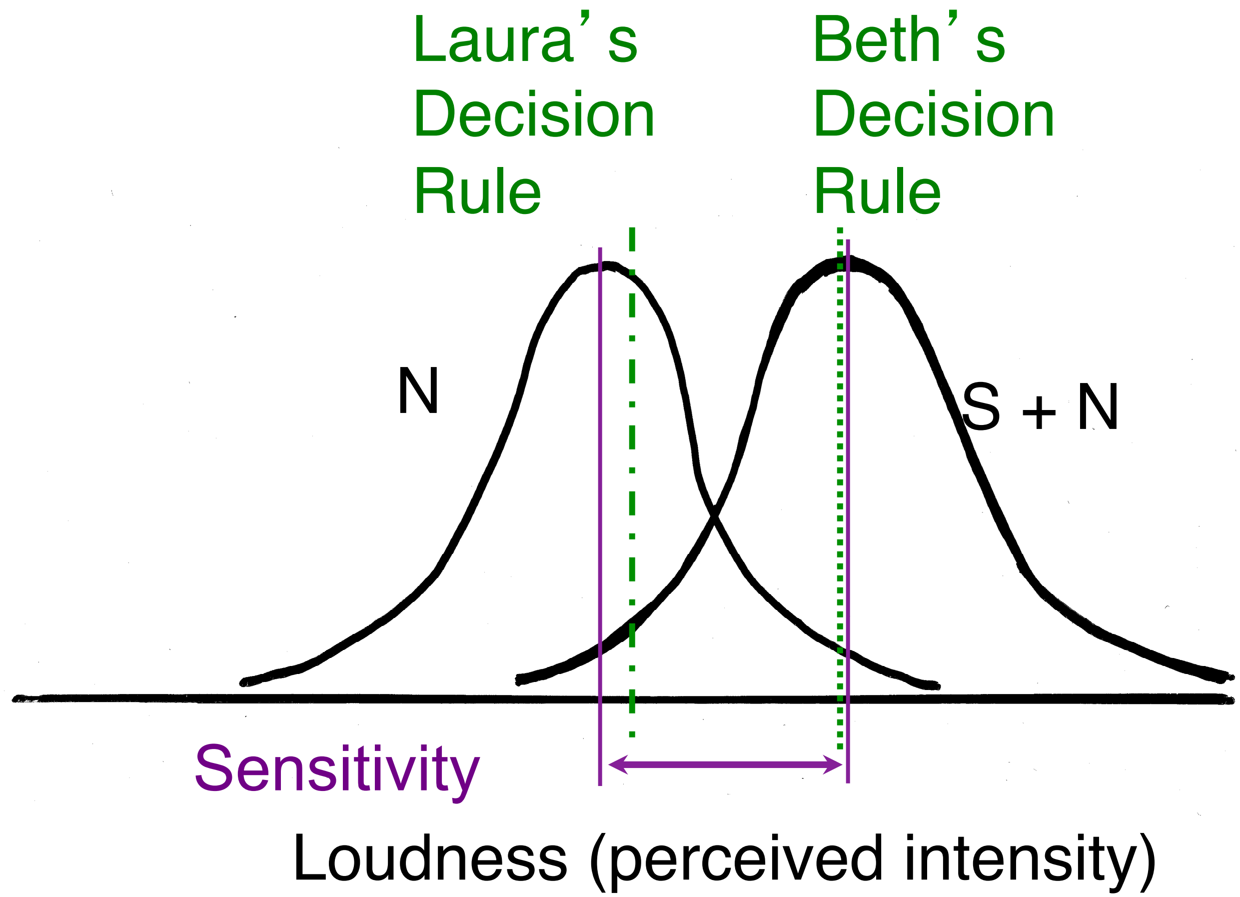
The graph shows Laura's data. A Hit occurs on an S_2 trial (distribution to the right) when the she says yes (to the right of the decision rule). The area under the S_2 distribution to the right of the criterion is the probability of a hit (shown in blue).

A false alarm occurs on an S_1 trial (distribution on the left) when the person says yes (to the right of the decision rule). The area under the S_1 distribution to the right of the criterion is the probability of a false alarm (in red).

Beth's Decision Rules



Loudness (perceived intensity)


Laura's
Decision
Rule
Beth's
Decision
Rule
N
S + N
Sensitivity
Loudness (perceived intensity)

Sensitivity

Using Hit and False Alarm data, we can compute two measures of listener performance.

Assume that the mean of the S_1 distribution is at zero. Then, using the false alarm rate, we can determine where the listener has placed her criterion (also called ß or beta) along the dimension of of the internal percept. Using hit and false alarm rates, we can determine the distance between the means of the S_1 and S_2 distributions. This is our measure of sensitivity and is called d' (d-prime).

Beth and Laura have different criteria but the same sensitivity.

Summary of TSD

Using TSD, we can assess a listener's performance in different conditions and see how both their decision rule and their sensitivity contribute to performance in different conditions.

Perceivers can differ in sensitivity, criterion, or both. We can also compare a single individuals performance in different conditions to see if differences in performance are due to criterion changes, sensitivity differences or both.

A Memory Explanation

One explanation for categorical perception is based on listeners having two memory codes: auditory and phonetic (Fujasaki & Kawashima).

The auditory code retains stimulus details. It is short lived and easily overwritten by new sounds.

The phonetic code (recode the auditory input into its language based segmental representation) is stable and easily held in working memory. It does not preserve acoustic detail.

Data

Pisoni (1973, *P&P*, *13*, 253-260; 1975, *M&C*, *3*, 7-18) has reported the results of a series of studies investigating categorical perception.

Lengthening the interval (ISI) in ABX triads makes perception more categorical. Going to a discrimination task (4IAX) that is less memory intensive than ABX makes perception less categorical. Transient, dynamic stimuli (e.g. stops varying in VOT or place of articulation) tend to be categorical. Steady-state vowels are not.

Results are consistent with memory models.

A Perceptual Discontinuity

Pastore (see Miller et al., 1976, *JASA, 60*, 410-417; Pastore et al., 1990, *P&P, 48*, 151-156) has explored an explanation based on psychophysical discontinuities.

Basically, the auditory system has coding mechanisms for complex ("pattern") properties of sound. For example, a set of neural mechanisms that detect synchrony versus asynchrony in information onset. This particular mechanism could underpin the coding of VOT. Stimuli that are coded by these mechanisms exhibit discontinuities in perception.

Data

Pastore (see above), Pisoni (1977, *JASA, 61*, 1352-1361) have reported data showing categorical perception for nonspeech sounds that have properties similar to VOT.

At first blush, the memory explanation will not work for these nonspeech data. However, if we change the two forms of memory to be auditory and “abstract” and note that the abstract does not preserve stimulus detail, then we have a memory model that will work for nonspeech.

The Role of TSD

TSD can be used to investigate the perception of speech and explanations for categorical perception. Basically, both the memory and perceptual discontinuity explanations predict that sensitivity within a category will be low and between categories will be large. That is, the influence of auditory memory or a perceptual discontinuity should show up in changes in sensitivity.

When TSD has been used to investigate phonetic continua, this is what has been found (see, e.g. Sawusch et al., 1980, *P&P, 27*, 421-434).

Summary

The predictions of memory and perceptual discontinuity models are very similar and no resolution to this debate seems satisfactory at this time.

For a wide ranging summary of views and issues, see S. Harnad (Ed.), 1987, *Categorical perception*, Cambridge Univ Press.