5aPP1. Sensitivity to Stimulus Distribution Characteristics in Auditory Categorization

Sarah C. Sullivan

University of Texas at Austin

Andrew J. Lotto

Boys Town National Research Hospital

Elizabeth T. Newlin & Randy L. Diehl

University of Texas at Austin

Introduction

A number of auditory tasks, including speech perception, require listeners to categorize stimuli on the basis of one or more features of the input. In many cases, especially speech, there is no one-to-one mapping between values along continuous features and discrete categories (e.g., phonemes). How then do perceptual systems categorize stimuli under uncertainty? One possible solution is to use probabilistic information from experienced stimulus distributions to optimize accuracy.

We propose that perceivers incorporate distributional knowledge about the acoustic environment with the information provided by the signal in order to make optimal (i.e., maximal accuracy) categorical decisions. Statistical approaches such as this are widely used in vision research but are rarely applied to auditory or speech perception. Our goal in this study was to develop a framework that will provide testable hypotheses about the nature of statistical (distributional) learning in auditory perception in general and speech perception, specifically.

Purpose

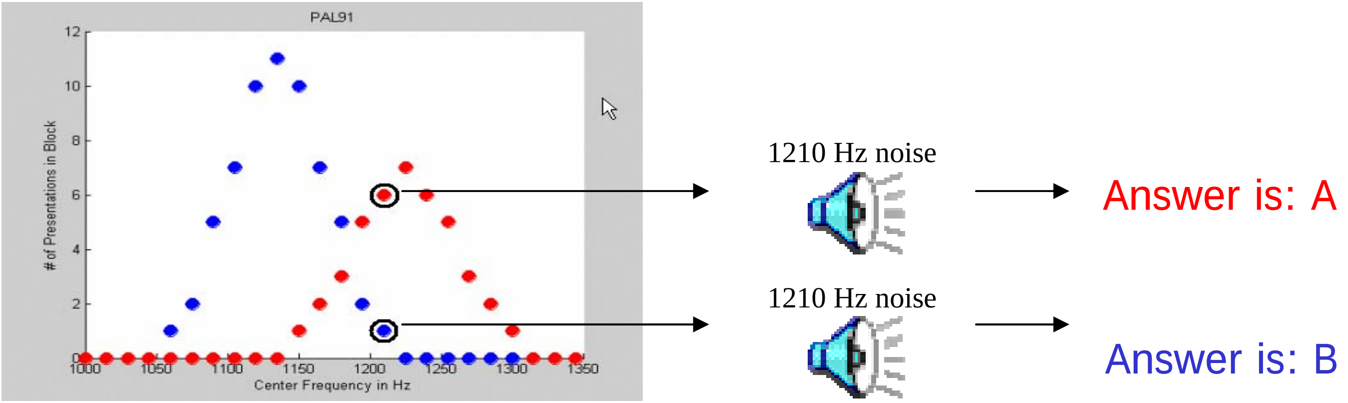
- Goal: Examine the ability of humans to categorize sounds as a function of training distribution characteristics, with future application to phonetic categorization.

- Empirical Questions:
  - Do listeners retain specific distribution information?
  - Do listeners form sharp boundaries between overlapping distributions (i.e., binary criterion judgment) or do categorization responses match the relative frequencies of the experienced distributions (i.e., probability matching)?
  - Do listeners use distributional information rationally or even ideally?
  - What are the implications for speech categories?

Stimuli

- Stimuli: 25 samples of band-pass white noise with center frequencies ranging from 1000-1360 Hz, increasing by 15 Hz steps. Band-pass filter had a bandwidth of 200 Hz.
- Non-Speech: Using non-speech stimuli allowed us to:
  - Control the listener’s distributional knowledge
  - Manipulate distribution parameters independently
- From a generalist perspective, speech and non-speech sounds are perceived the same way.

Task: Categorization Under Uncertainty



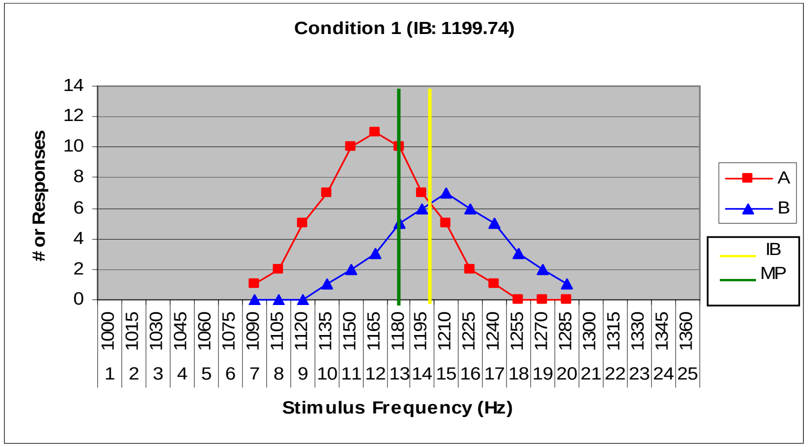
18 subjects were randomly presented noise bursts sampled from a variety of overlapping distributions. Participants were asked to identify the sounds as belonging to one of two categories (“A” or “B”). Feedback was provided. Stimuli were presented in 6 six-minute blocks.

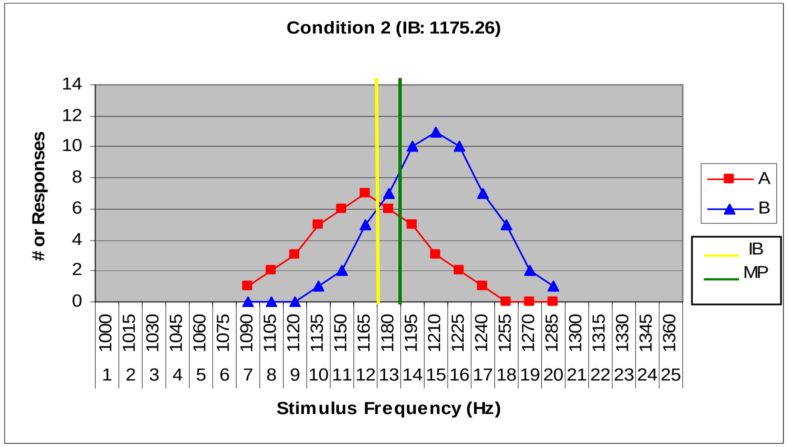
Hillenbrand, J. M. & Gayvert, R. T. (2005). Open-source software for experiment design and control. *JSLHR*, *48*, 45-60.

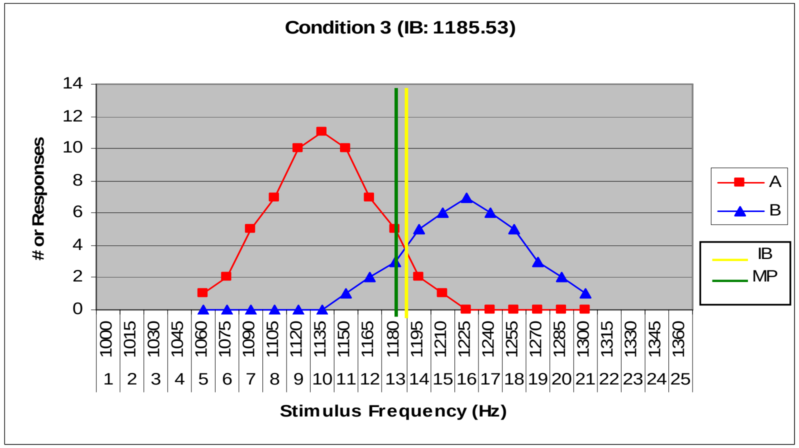
Conditions

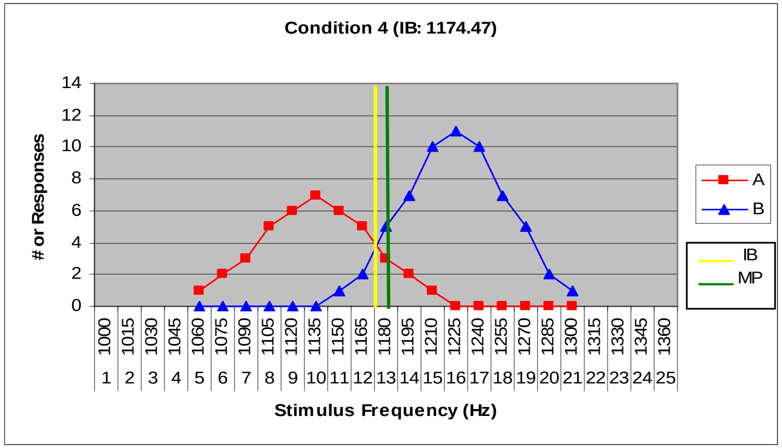
Ratio of "A" & "B" Stimuli

Overlap









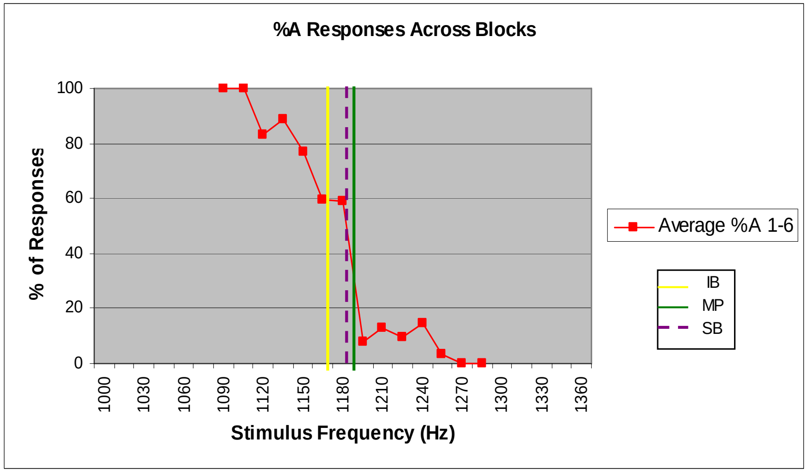
Red Lines – Category A Distribution
Blue Lines – Category B Distribution

Yellow Lines – Distribution Crossover
Green Lines – Midpoint of Range

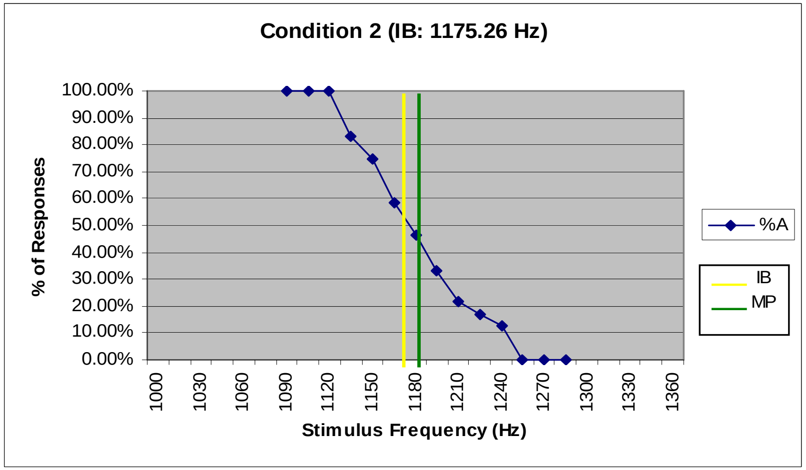
Dependent Variables

Boundary & Slope Measures:

- The number and percent of “A” and “B” responses were documented for each subject across 6 training blocks.
- Categorization functions (% A Responses) were plotted for each subject, averaged across all 6 blocks.
- Subjects’ boundaries and slope values were calculated by linear interpolation between the 2 points on either side of 50%
- Subjects’ boundary and slope values were compared to training distribution values.

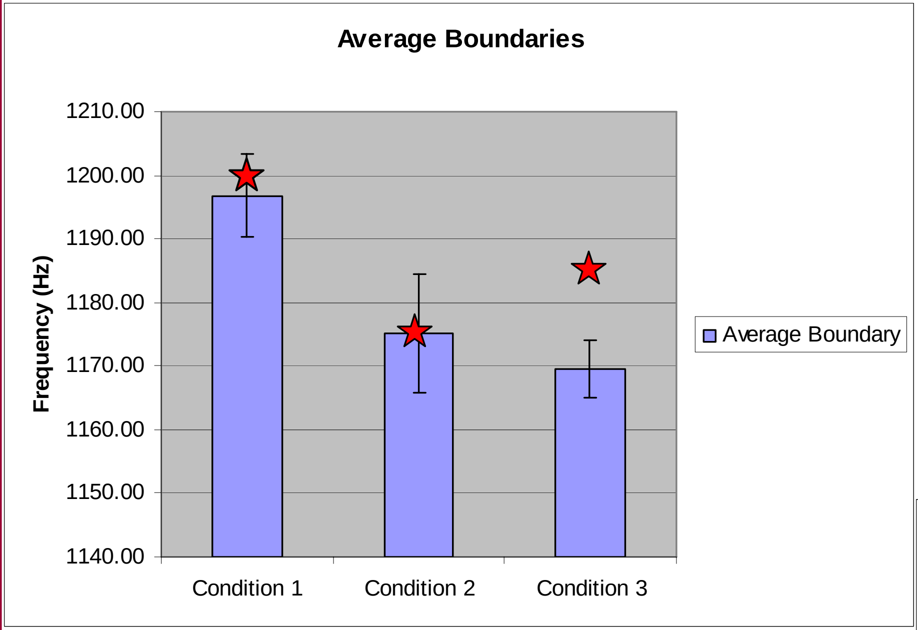


50% Boundary: 1182.63 Hz
Slope: -3.41



50% Boundary: 1175.26 Hz
Slope: -0.81

Boundary & Slope Results

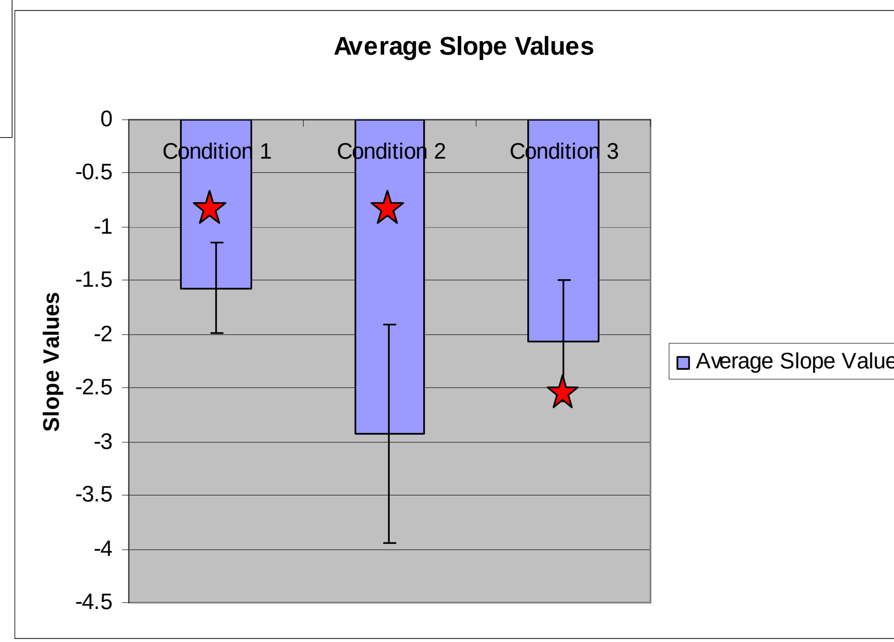


Boundary Results:

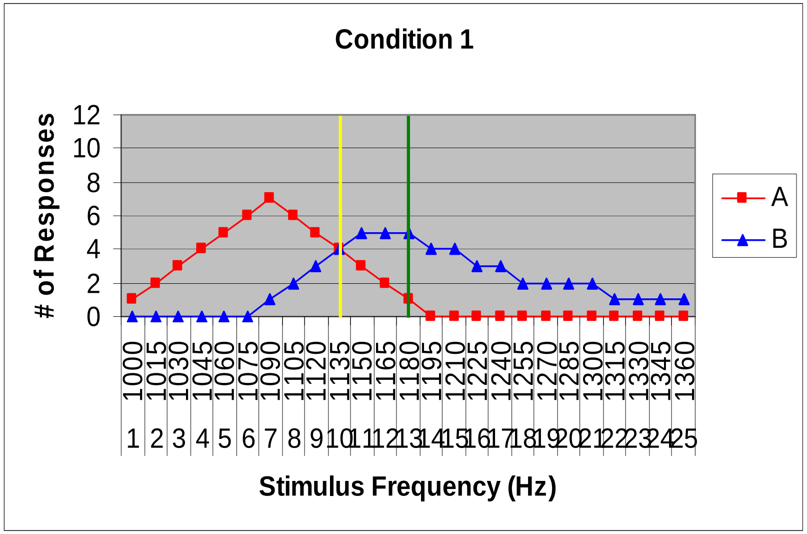
Subjects in Condition 1 and 2 (but NOT Condition 3) established boundaries that were very close to the training distribution crossover points, on average.

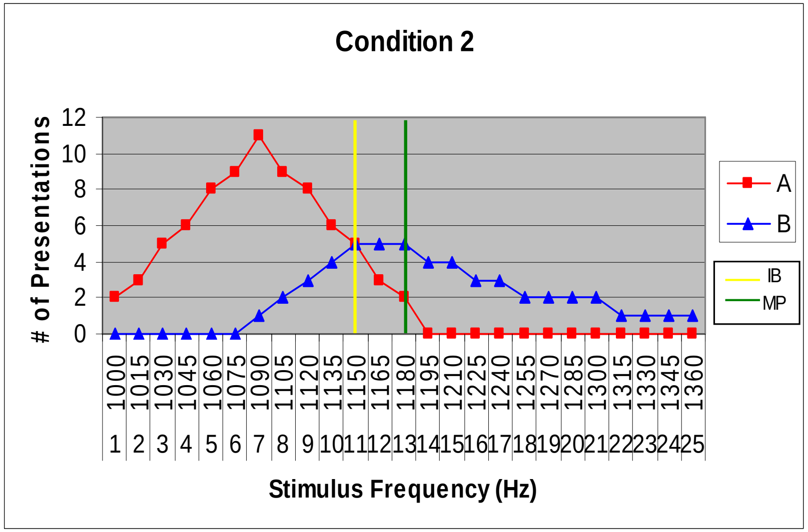
Slope Results:

Subjects in Condition 1 and 2 (but NOT Condition 3) had slope values greater than the slope values of the training distributions, on average.



New Conditions



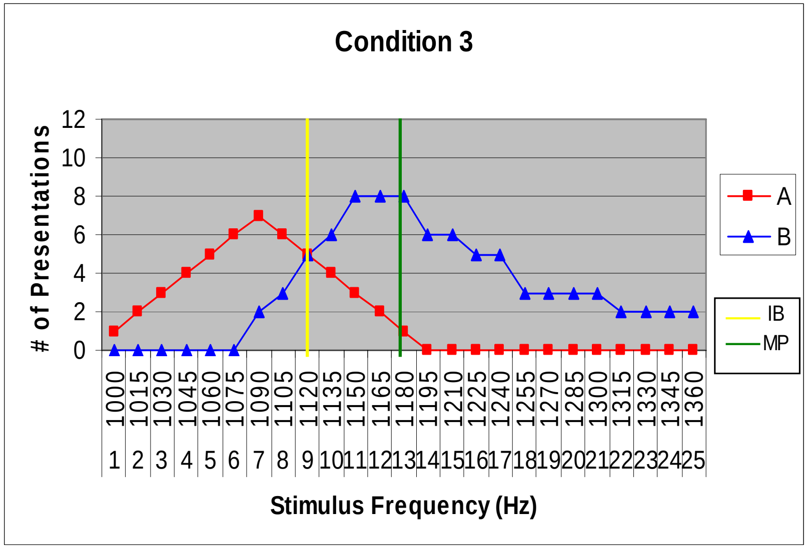


Red Lines – Category A Distribution

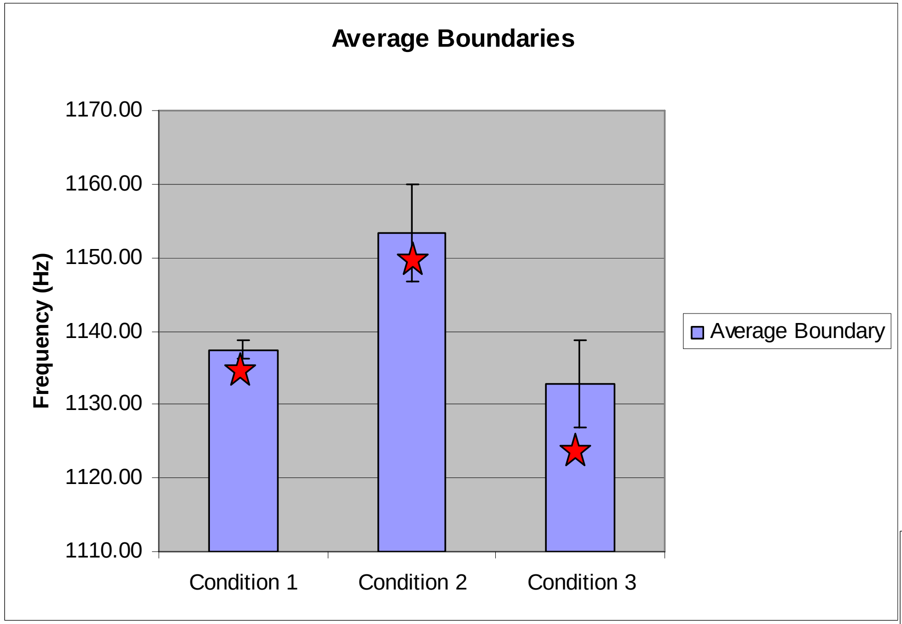
Blue Lines – Category B Distribution

Yellow Lines – Distribution Crossover Points

Green Lines – Midpoint of Range

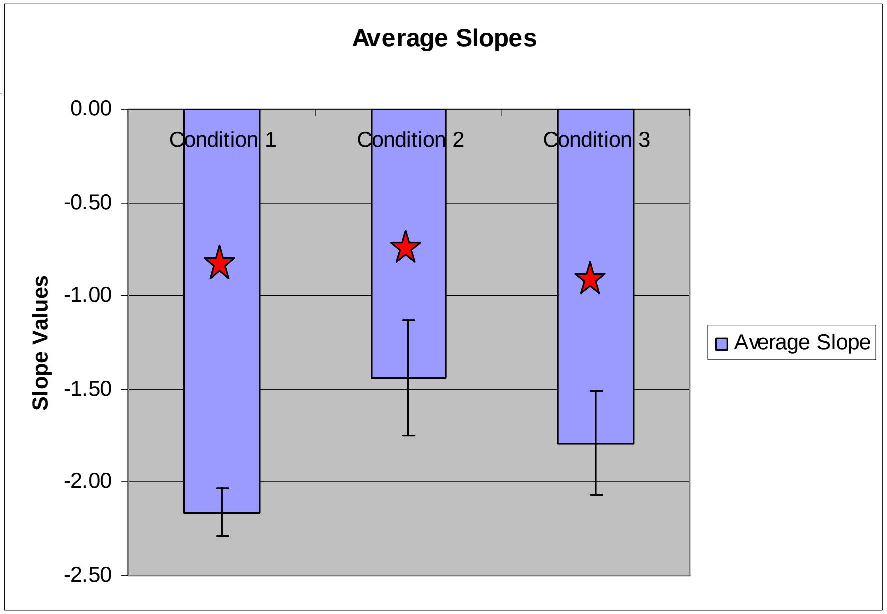


Boundary & Slope Results



Boundary Results:

Subjects in Condition 1 and 2 (but NOT Condition 3) established boundaries that were very close to the training distribution crossover points, on average.



Slope Results:

Subjects in all 3 conditions had slope values greater than the slope values of the training distributions, on average.

Answers to our Questions

Empirical Questions:

- Distribution Information: Listeners are very sensitive to characteristics of auditory distributions. Boundaries are equivalent to crossover points and there is a significant difference between conditions even though training distribution crossover points vary by only a small degree.
- Criterion Judgments vs. Probability Matching: Sharp boundaries argue for criterion-based judgments.
- Optimal Performance: Listeners’ responses appear rational and in fact, a Bayesian approach seems warranted.

Bayesian Approach

- Theoretical framework for studying perception
- Suggests that the mapping between events in the world and sensations is not deterministic but rather probabilistic
- Hypothesizes that perceivers use statistical information to make inferences as to what events are occurring in the world

Baye's Theorem

$$p(E \backslash S) = p(S \backslash E) \bullet p(E) \, / p(S)$$

Posterior Probability	Stimulus Likelihood	Prior Probability	Normalizing Factor

Where E = Event, S = Stimulus

Distribution Information

1) Prior probability = probability of an event having occurred, independent of stimulus information

Study Example: probability of a particular category being presented independent of sensory information, stimuli ratios (e.g., p (A) or p (B))

2) Stimulus likelihood = probability of a stimulus value given that a particular event has occurred

Study Example: the probability a particular frequency given a particular category (e.g., p (1210 Hz/"A"))

Speech Categories

1) **Prior probability** = probability of a particular phoneme occurring; based on lexical, semantic, and phonotactic constraints (e.g., p (/v/) following /z/ low in English but not in Russian)

2) **Stimulus likelihood** = probability of an acoustic attribute (e.g., F2 = 2500 Hz) given that a particular phoneme (e.g., /i/) was produced

Conclusions

- In an auditory categorization task, listeners require very little exposure to input distributions to establish a category boundary that is optimal in the sense of maximizing accuracy.
- The stimuli used in this task were not naturally-occurring sounds and the listeners were not instructed on what the informative cue was. Each sound was presented once and feedback always followed sound presentation and response. The fact that listeners performed near optimally under these conditions demonstrates the effectiveness of perceptual categorization.
- The categorization functions were reminiscent of those obtained for phonetic contrasts. It is plausible that this methodology allows us to tap into the same processes involved in phonetic acquisition.

Acknowledgements: Work sponsored by grants from NIH-NIDCD and NSF.