DISCRIMINATING STATISTICAL DEPENDENCIES IN GLIDING-TONE SEQUENCES

DANIEL CARDOSI and GERALD KIDD, JR.

Boston University College of Health & Rehabilitation Sciences: Sargent College
Department of Speech, Language, & Hearing Sciences

I. INTRODUCTION

Solving the **"cocktail party problem"** (see Middlebrooks et al. 2017 for reviews) involves segregating and identifying multiple sources of sound (e.g. **sound source determination**; Yost 1991) and comprehending the message from the target source.

To accomplish this, listeners use various types of information, including **physical source characteristics** (e.g., transducer materials/setup) and **acoustic signal properties** (e.g., frequency, level, timbre)

However, sound sources may also be distinguished by the **statistical properties** of their output signals (Lutfi 1994, Kidd et al. 2013, McDermott et al. 2013, Masutomi et al. 2016, among others).

Past work on source determination by signal statistics has focused on distributions of amplitude envelopes, crucial for speech intelligibility.

Here we ask, "can listeners discriminate between sound sources on the basis of statistical properties of the **frequency variation** within sound sequences?" Answering this question is important because certain types of frequency variation (e.g. formant frequency transitions) are thought to bear on **sound segregation**.

II. METHODS: PROCEDURE

Participants: eight young adults (two male, six female) with normal hearing (pure-tone thresholds ≤ 20).

A two-interval, two-alternative forced-choice paradigm was used (with response feedback).

On each trial, observers were presented with two sequences and prompted to indicate which contained a stronger statistical structure.

Stimuli were generated on a subject-by-subject basis and presented diotically through headphones. Each subject performed the task individually while seated in a double-walled sound-attenuating booth.

BOSTON UNIVERSITY

Work supported by NIH-NIDCD R01 DC004545

III. METHODS: STIMULI

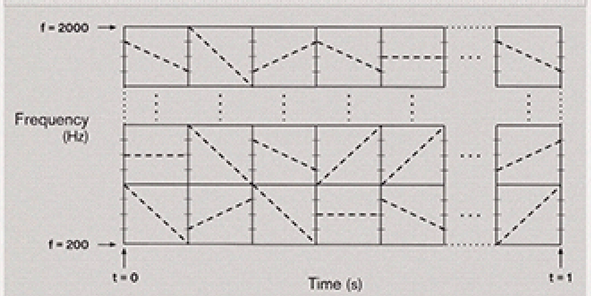
On each trial: first stimulus interval (1 sec), ISI (0.4 sec), second stimulus interval (1 sec)

In each interval: M concurrent N-tone sequences (M and N dependent on tone duration; M fixed to span 200–2000 Hz, N fixed to span 0–1 sec)

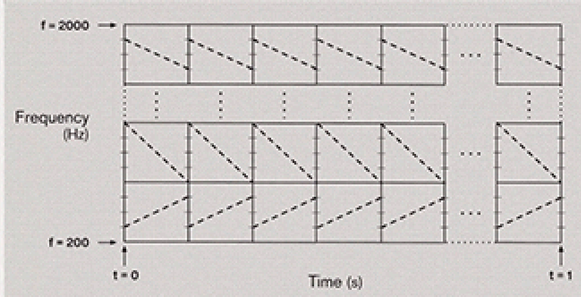
Tone slopes: determined in initial time window by random selection, post-initially by Markov matrices

1 non-target (uniform); 5 targets, ranging between the uniform and identity matrices.

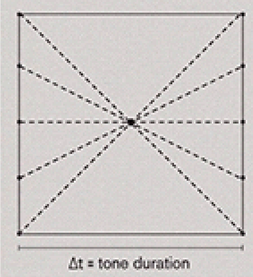
Non-Target Sequence (d = 1/5)



Target Sequence (d = 1)



DETAIL: Time-Frequency Unit



Transition Matrix

$$\begin{vmatrix} d & r & r & r & r \\ r & d & r & r & r \\ r & r & d & r & r \\ r & r & r & d & r \\ r & r & r & r & d \end{vmatrix} \quad d \in \{1.00, 0.84, 0.68, 0.52, 0.36\} \quad r = (1/4)(1-d)$$

Glide Slopes

Tone Dur. (ms)	50	100	200
Shallow Glide Mag. (oct)	19/320	19/160	19/80
Steep Glide Mag. (oct)	19/160	19/80	19/40

IV. RESULTS

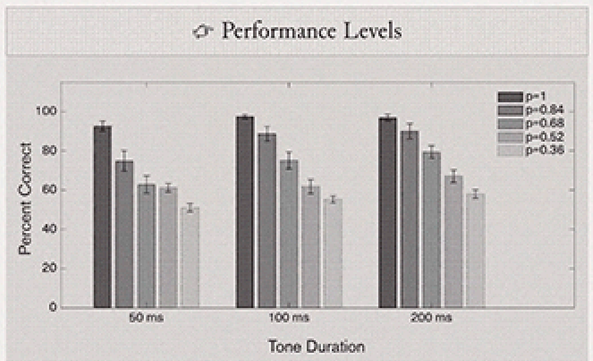
Rapid acquisition of task proficiency for all tone durations.

Averaged functions relating percent-correct discrim. performance to degree of target statistical structure were monotonic over the range tested.

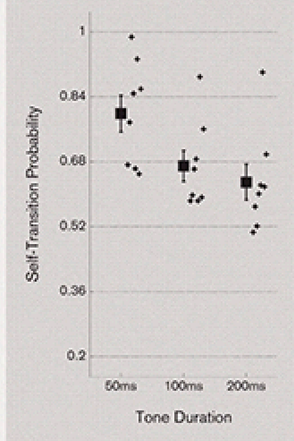
Target strength (specified by the diagonal of the transition matrix) had a signif. effect on performance levels (p < **0.001**).

Duration had a signif. effect on thresholds (p = **0.028**), which ranged from 0.79 (50 ms) to 0.63 (200 ms), but not on slopes (p = **0.34**).

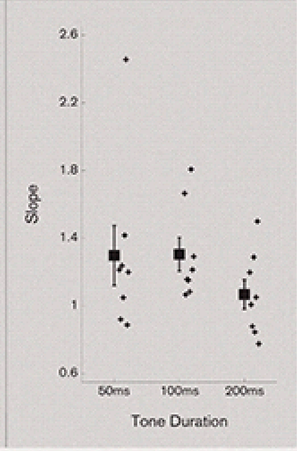
Performance Levels



Thresholds ($d^{75\% \text{ Correct}}$)



Psychometric Slopes



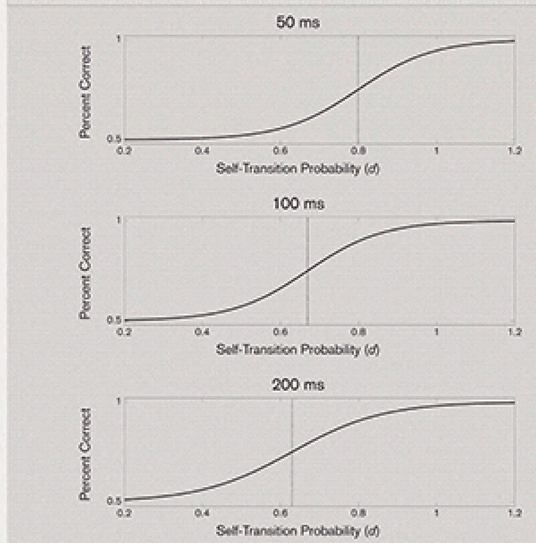
V. DISCUSSION

Listeners are able to discriminate the statistical structure in broadband, time-varying acoustic signals.

Freq.-variation sensitivity was highest for 200-ms tones (fewest streams in composite stimulus), where performance was probably governed by the better resolution of individual tone-sweeps; shorter tone durations likely yielded more of an integrated percept.

Discrim. performance was good across a range of values possibly important for segregating speech (e.g., by perceiving spectral-peak movements like formant transitions) or for identifying nonspeech stimuli from their statistical properties.

Psychometric Functions (Psignifit 4)



BIBLIOGRAPHY

Kidd *et al.* (2013): J. Acoust. Soc. Am. 134, 1215-1231.
Lindblom *et al.* (2009): Proc. FONETIK, 18-23.
Lutfi (1994): J. Acoust. Soc. Am. 95(3), 1490-1500.
Masutomi (2016): J. Exp. Psychol. 42(3), 386-400.
McDermott *et al.* (2013): Nat. Neurosci. 16(4), 493-500.
Middlebrooks *et al.* (2017): SHAR 60.
Młynarski & McDermott (2018): Conf. Cogn. Comp. Neurosci., 1135
Öhman (1965): KTH Speech Transmission QPSR 6(1), 10-13.
Yost (1991): Hear. Res. 56, 8-18.