Analogy and Disanalogy
in Production and Perception of Speech

Robert E. Remez
Department of Psychology
Barnard College, Columbia University

THE HASKINS LEGACY - The Science of the Spoken and Written Word
Acknowledgment

Generous Colleagues

Arthur Abramson
Peter Bailey
Eric Bateson
Ann Bradlow
Chris Darwin
Kathryn Dubowski
Jennifer Fellowes
Daria Ferro
Louis Goldstein
Luis Hernandez
Keith Johnson
Bob Krauss
Harry Levitt

Paul Luce
Kevin Munhall
Lynne Nygaard
Jennifer Pardo
Rebecca Piorkowski
David Pisoni
Larry Rosenblum
Philip Rubin
Michael Studdert-Kennedy
Emily Thomas
J. D. Trout
Doug Whalen
Cynthia Yang

Sponsorship

NIDCD DC000308
Research on the perception of speech was produced by a shotgun marriage of linguistics and psychology. If this union generated enormously useful empirical work that receives wide recognition across cognitive science and behavioral neuroscience, the adequacy of explanation in speech perception has been hampered by accounts that express singular theoretical devotion to one of the parent disciplines. Some research suggests that a choice between these two emphases is unwarranted. A dual focus on the abstract symbolic properties deriving from linguistic analysis and the exquisitely graded expressions observed in psychological method can set new descriptive and theoretical challenges that break the old stalemate between motoric and psychoacoustic accounts. To buttress this cheerful view, studies will be reviewed that identify instructive disanalogy between production and perception. The theoretical impact of acknowledging the importance of disanalogy permits both ends of four critical dimensions of variation to be characterized directly: speech perception is simple and complex, free and situated, conceptual and embodied, enduring and dynamic.
Language: A Local Phenomenon

Observable Universe
Production and perception of speech…

are irreducible to a fixed set of physical attributes;

are irreducible to a fixed set of physiological (auditory, motoric) attributes;

use linguistically governed phoneme contrasts which licenses phonetic variation without apparent limit;

exhibit a convergence of linguistic, personal and circumstantial causes.
A Parable of Perceptual Bistability
Reports and Discussions


Canonical Speech Cues

The Vowels of English

frequency

i l ” ø a Ø u
Perception via Canonical Speech Cues
We begin with an anomalous finding…
We begin with an anomalous finding…
We begin with an anomalous finding…
An anomalous finding…

Where were you a year ago!

Now listen to this as if it were synthetic speech…

Remez, Rubin, Pisoni & Carrell (1981)
Why are sine-waves perceived as speech?

Two differing conceptualizations of perceptual organization and analysis:

Analyze sensory elements into features, then bind features into objects.

Egon Brunswick
(1903-1955)
Why are sine-waves perceived as speech?

Two differing conceptualizations of perceptual organization and analysis:

Group sensory elements into organized streams, then analyze bound elements into objects.

Max Wertheimer
(1880-1945)
Perceptual Organization
Organizational Functions

Cocktail party setting:
  Are these $a_i$ yours?

Multimodal setting:
  Which $v_i$ go with which $a_i$?

Cocktails for two:
  Do these $a_i$ compose a set?

Which accounts of perceptual organization are adequate for speech?
Are the constituents of speech a kind of acoustic element?
A problem...

There is no closed set of $a_i$ that function as speech cues.

Sine-wave speech does not resemble natural speech in detail.

Time-varying sinusoids are improbable $a_i$ to compose speech.

Analysis is conditional on grouping.
A perceptual conclusion…

The perceptual organization of speech occurs by virtue of modulation sensitivity independent of the elements composing the carrier.

The elements need not be familiar, nor similar to natural acoustic vocal products.

Perceptual organization triggered by coordinate variation is concurrent to element-based Gestalt grouping.
Why does it matter?

This work can tell us what kind of theory of speech perception we need:

Perceptual impressions of naturalness depend on the auditory phenomenality of speech. But, phonetic perception does not depend on natural auditory quality of speech...
Why does it matter?

This work can calibrate the relevance of animal models of speech perception:

Some studies indicate close similarity in psychoacoustic function in humans and our relatives (quail, chinchilla, cat).

The value of this work depends on its relevance to phonetic perception.
Why does it matter?

This work can restrict the domain of actuarial theories of cognitive function in a principled way: Theoretical accounts that aim to tally the unelaborated incidence of sensory qualities might be adequate to explain the perception of naturalness... yet, these all entail impossibly durable sensory traces.

Accounts that invoke distributional characteristics of sensory features are unlikely to explain phonetic perception.

Why does it matter?
Sine-wave speech is conspicuously bistable…

A musical case…

simætwææle

BWV 846
Is there any independent corroboration?

Noiseband vocoded speech

Is there any independent corroboration?

Acoustic chimeras

Smith, Delgutte & Oxenham (2002)
The pattern of speechlike modulation makes its acoustic constituents effective linguistically;

Isolated acoustic elements have no linguistic causal value;

Detailed sensory properties evoke an impression of the phenomenality of speech;

The pattern of modulation evokes an impression of the linguistic properties of speech.

Auditory integration and audiovisual integration differ, blocking a strong claim of amodality or common-metric (motor) as a mediating function.
Production and perception of speech…

are irreducible to a fixed set of physical attributes;
are irreducible to a fixed set of physiological (auditory, motoric) attributes;
use linguistically governed phoneme contrasts which licenses phonetic variation without apparent limit;
exhibit a convergence of linguistic, personal and circumstantial causes.
A dynamic of abstract coherence

The production and perception of speech occurs as if the commitment to a particular motoric or sensory realization of linguistic contrasts is flexible. This readiness to create and to find functional contrasts in unexpected motoric or sensory form opposes the fixity claimed in accounts of phonetic and phonemic equivalence.