24.941J / 6.543J / 9.587J / HST.727J The Lexicon and Its Features
Spring 2007

Categorical perception and its implications for spoken word recognition

24.941/6.976

10/28/04

A psychological perspective*

Acoustic phonetics: Characterization of the relevant structure of the input representation, what information is available, and why it is structured the way it is.

Phonology: ~"…is a theory of the representations in the lexicon that allow us to produce and recognize words."

-DS

*the view from the valley

By implication:

The goal of psychological research into spoken word recognition is to understand the direct mapping between acoustic phonetic structure and phonological representations.

…kind of

Defining the computational problem (Marr, 1981)

- What is the goal of the computation?
- What is the appropriate strategy?
- What is the logic of the strategy for carrying it out?

Defining characteristics of the problem

Input

- Transient, rapid, highly variable
- Small phonetic inventories, phonotactic constraints, preference for relatively short words and large vocabularies create difficult discrimination
- Highly structured due to physiological. mechanical and aerodynamic constraints on production

What is the output?

Ultimate goal is communication

- activation of correct word forms is an intermediate step, not the ultimate goal of spoken language perception
- activation of phonological representations may be mediated by lexical/semantic/syntactic/pragmatic factors

Ultimate goal is to communicate (generally contextualized)

What might be communicated?

- Meaning, identity (individual, gender, size, region, class….), affect….
- Implication: What is noise v. signal?

  It depends on what you want to get out of the signal

Relevance to phonologists

Spoken language processes probably involve a combination of perceptual and top-down processes

Perceptual processes impose direct systematic constraints on the listener that may influence the evolution of phonological systems

Top-down processes are less likely to directly pose systematic constraints (but may indirectly reflect production constraints)

Implication: Theories that ground phonology in listener phenomena must be premised on an explicit processing model

Some broad characteristics of the computation

It's fast (~200 ms from onset):

Fast shadowing

Online lexical effects (shadowing and mispronunciation monitoring)

Electrophysiological measures: N400 onset

There is parallel activation of candidates

Homophone processing: She found a *bug* on the lamp.

Fragment priming: *captain/captive*

Categorical Perception

"…is a mode in which each acoustic pattern, *whatever its context*, is always and only perceived as a token of a particular phonetic type".

Studdert-Kennedy (1971)

Categorical perception (CP)

CP empirically characterized by:

Sharp transitions in ID functions

A peak in discrimination across category boundaries (ID should predict discrimination)

Better between than within category discrimination

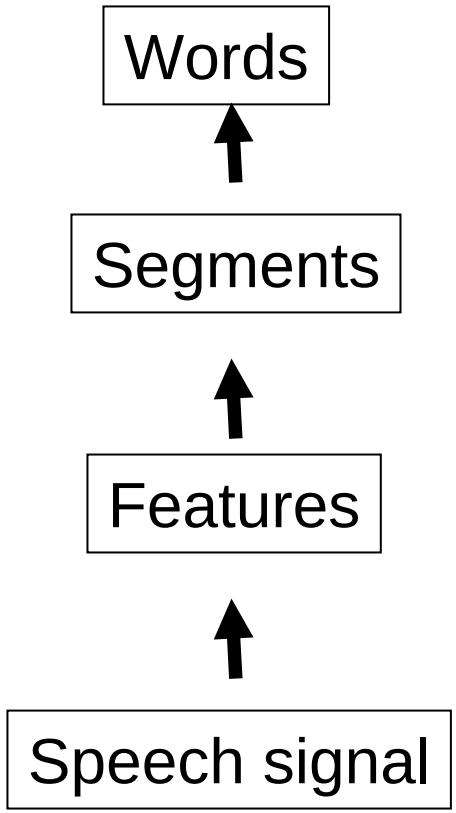
Note: The function properties used to identify CP are themselves continuous

Please also see Liberman, Harris, Hoffman & Griffith (1957).

Implications of naïve CP for models of spoken word recognition

- Speech percept is largely unambiguous
- Features (and higher levels of representation) are perceived independently
- Within category variance is lost early in processing and so cannot influence higher levels of processing

Spoken word recognition as sequential pattern mapping



Early cohort model - sequential pattern mapping structured by temporal constraints

Figure removed due to copyright restrictions.

Please see Marslen-Wilson (1984).

An evaluation of categorical perception and its limits

- CP as laboratory artifact
- Within category structure
- Non-independence

CP tasks all demand explicit binary judgments, but can we assume that such judgments are implicit in natural ASR? (Massaro)

Figure removed due to copyright restrictions.

Please see Figure 2 in Massaro.

Methodological factors shown to influence CP

Task effects: ABX data more categorical than AX

Timing: Between category discrimination improves as ISI increases from 100-2000 ms in AX

Within category discrimination decreases over the same increase in ISI (van Hessen & Schouten, 1992)

Issue: Lexical access effects show up roughly 200 ms after word onset, meaning interpretation may normally happen before perception is strongly categorical

CP effects are malleable

Identification function can be arbitrarily reshaped by feedback

Figure removed due to copyright restrictions.

Please see Carney *et al.* (1977).

The identification-discrimination relationship

CP effects in discrimination increase with memory load and task characteristics biasing subjects towards basing comparison on coded labels.

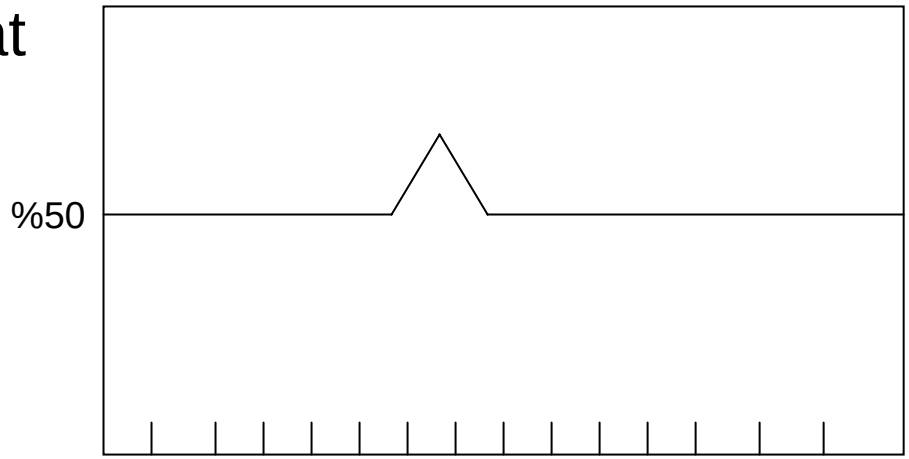
Covert labeling strategies make discrimination a variant of identification that could tap the same strategic decision mechanisms

Within Category Structure

CP as normalization

Discrimination data suggest that listeners are insensitive to within-category variation

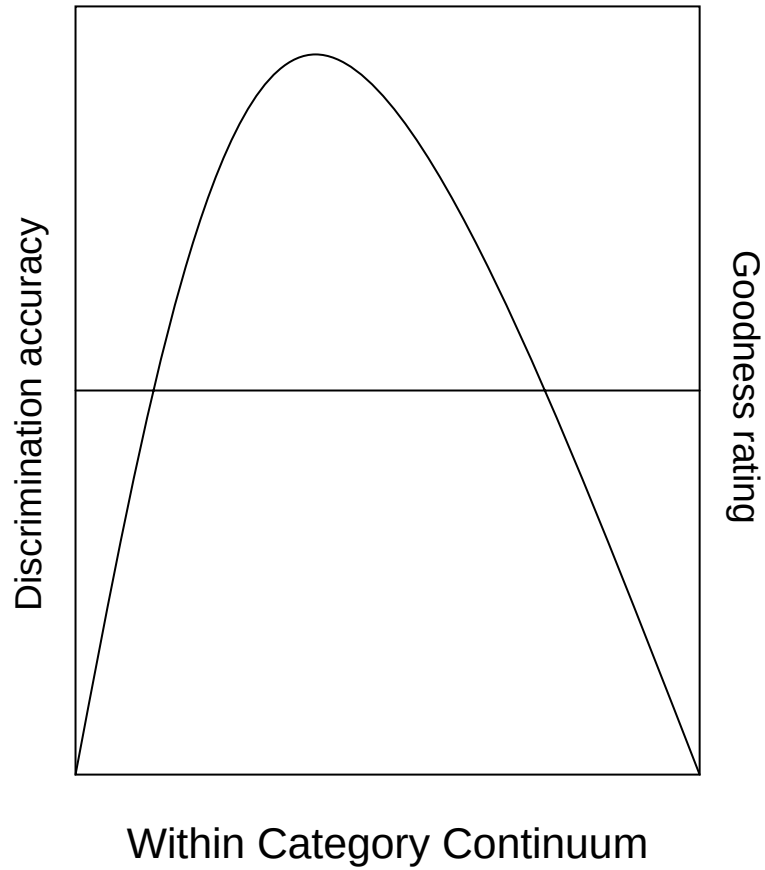
Normalization strips away within-category variation, Simplifying mapping at the possible cost of losing useful information



Within Category Structure

Discrimination performance is highly dependent on task and synthesis details

Ratings of category goodness, and RT measures in ID and monitoring tasks suggest the existence of structure within categories



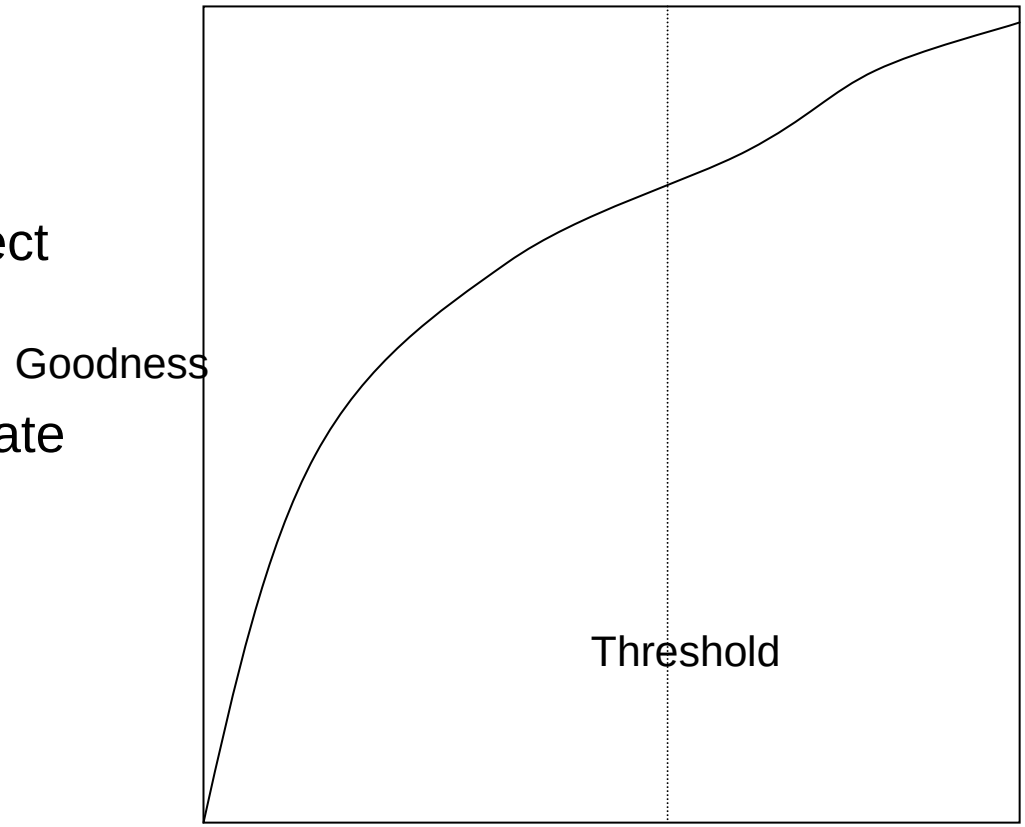
Within Category Structure

Weak exemplars slow online processing, especially near boundaries with native phonetic categories

Figure removed due to copyright restrictions.

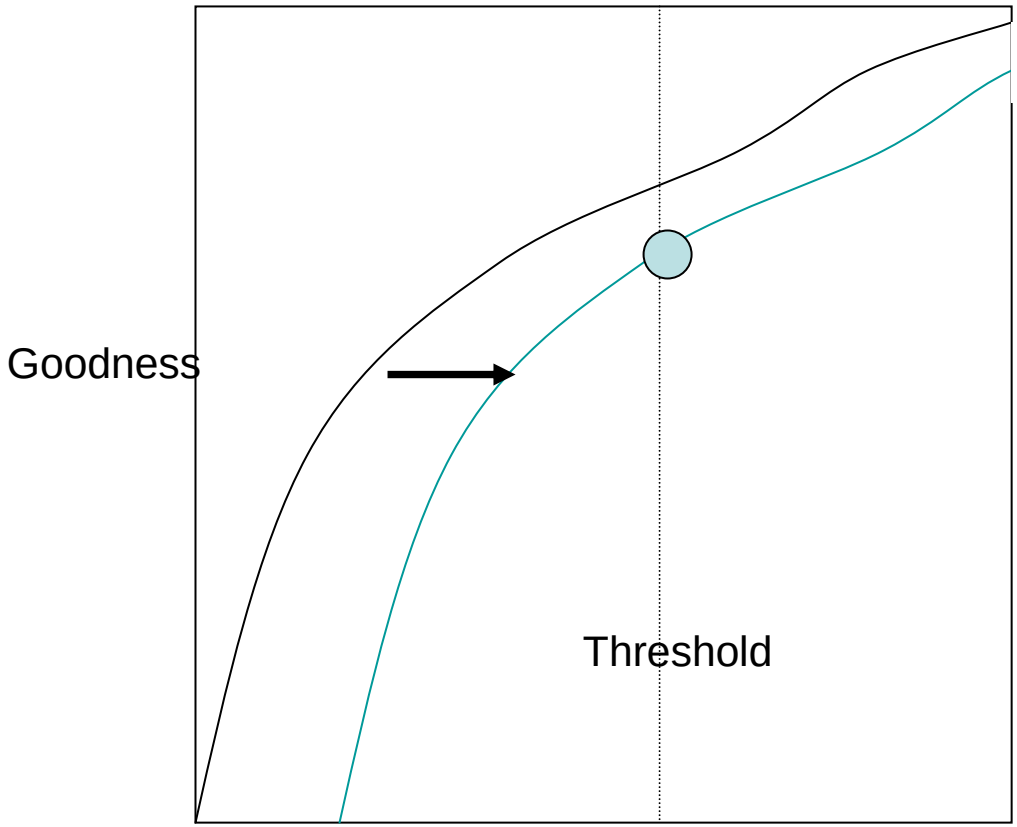
Are CP data irrelevant, if they reflect strategic, task-specific processes that may not be part of normal spoken word recognition?

Even if CP results reflect strategic decision mechanisms, these mechanisms still operate over a distribution derived from listeners' perception of stimuli



Are CP data irrelevant, if they reflect strategic, task-specific processes that may not be part of normal spoken word recognition?

Evidence for shifting goodness functions (e.g. Miller, 2001) suggest that non-independence effects reflect perceptual processes



Dependencies that influence category Structure and boundaries

Non-independence: Cue-Trading

Individual features tend to be encoded by many cues.

E.g. A subset of known voice cues

- Voice onset time
- Duration of voiced formant transitions
- F1 onset frequency
- F2 onset frequency
- F3 onset frequency
- Spectral characteristics of following vowel
- Duration of following vowel
- Duration of aspiration
- Intensity of aspiration
- Direction of F0 change at onset of voicing

Lisker (1978)

Non-independence: Cue-Trading

Multiple cues are integrated in feature perception

- F1 onset is higher, VOT longer in unvoiced velar stops
- If F1 200 Hz, crossover VOT = 34ms
- If F1 = 400 Hz,crossover VOT = 23ms
- Implication:
  1. Continuous feature values may emerge despite the use of some quantal cues
  2. Multiple cues are integrated

Figure removed due to copyright restrictions.

Summerfield & Haggard (1977)

Non-independence: Feature interaction

In production:

- VOT: unvoiced > voiced
- VOT: velar > labial

Figure removed due to copyright restrictions.

In VOT& F1 stop continua onset identification varies as a function of place (burst, F2, F3)

Benki, 2001

Non-independence: Adjacent features

1. Vowel-fricative effect

- Anticipatory lip rounding for [o] and [u] lowers the fricative noise spectrum in FV's
- [ʃ] generally has a lower frequency spectrum than [s]
- Listeners appear to compensate for round vowel contexts and show an [s] bias in F categorization before a round ([u]) v. nonround ([a]) context
- Effect broken by temporal separation, or loss of formant transitions into vowel (suggests a role of cue grouping?)
- Note: This effect operates backwards in time suggesting either reanalysis or delayed analysis operating over a buffer

Figure removed due to copyright restrictions.

Mann & Repp (1980)

Non-independence: Adjacent features

2. Fricative-stop effect

- Bias towards velar categorization of ambiguous stops following [s] v. [ʃ] with coronal bias following [ʃ]

- Similar effect of stop on fricative categorization

- Modulated by temporal separation of potential syllable boundaries suggesting role of low-level processes

Figure removed due to copyright restrictions.

Mann & Repp (1981)

Non-independence: Rate normalization
1. Local Effects

Duration cues vary with rate

/wa/ - /ba/ discrimination

- Faster transitions heard as /ba/
- Interpretation of the same transition shifts as a function of vowel duration
- Crossover point shifts from 32-47 ms. As syllable duration move from 80-296 ms
- Effect depends on syllable duration (c), not stimulus duration
- Could be construed as a trading effect

Figure removed due to copyright restrictions.

Miller & Liberman (1979)

Non-independence: Rate normalization
2. Global Effects

/ba/-/wa/ discrimination

- F1 and F2 transitions of 15-65 ms. followed with 40 ms.steady state

- Syllable preceded by 1.2 second sequence of random fast (30 ms) or slow (110 ms) tones in the F1-F2 frequency range

- Attributed to putative durational contrast effect (Weber's Law)

Figure removed due to copyright restrictions.

Holt & Wade (2004)

Non-independence: Lexical effects

Ganong effect (1980)

- 2AFC identification of segments in syllabic/lexical context
- Set up so one end of continuum yields a word, other end a nonword (e.g.*gift-*kift*)
- Function biased towards lexicality in transitional region
- Word-final effects (rug-*ruk) larger than word-initial ones, extending to unambiguous tokens

Figure removed due to copyright restrictions.

Pitt & Samuel, 1993

Non-independence: Multimodal Effects

McGurk Effect (1976)

Perception of stop place
infuenced by visual cues

- Speaker articulates [ga]
- Audio recording of [ba]
- Listener hears [da]

working demo at: McGurk Effect Demonstration

Some possible implications of non-independence in CP

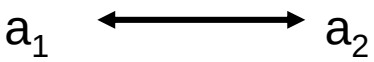
- Variance (noise) versus covariance (potential signal)

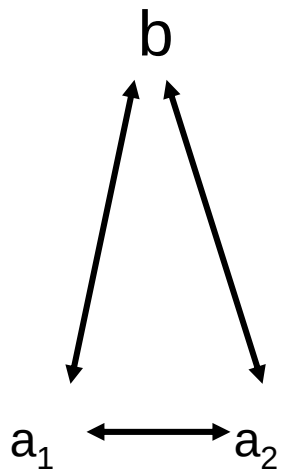
- Spoken word recognition is fundamentally a problem of integration over time, sensory modalities, and across all levels of representation that requires a high degree of interactivity

Summary of Dependencies

Phenomenon	**Dependency**
Rate normalization	nonspeech sounds-cue
Cue trading	cue-cue (within)
Compensation for coarticulation	cue-cue (between)
McGurk Effect	speech-vision
Ganong Effect	cue-lexical representation

Modeling Dependency





Parallel

Driven by perceptual Process (possibly shaped by signal constraints)

Top-Down

Driven by lexical or implicit knowledge only

Interactive

Perception, implicit and lexical knowledge all play roles

Part 2: The TRACE model

Background: Lexical Access 70's and 80's style

Logogen Model - parallel, interactive, continuous activation, vague

Autonomous Search Model- serial, bottom-up

LAFS - bottom-up, concerned with context effects, no intermediate representation

Cohort-parallel, bottom-up, early, discrete

TRACE Architecture

- Feature, phoneme and word levels
- All representations time-aligned and repeated

Representation

Localist representation of: 211 words, 14 phonemes, 9 values for 7 features

Figure removed due to copyright restrictions.

Activation Dynamics

Figure removed due to copyright restrictions.

Activation passed between levels to all consistent representations

Competition (inhibition proportionate to activation) between nodes within levels

Activation decays over time

Activation may continue to evolve after word offsets

Three Varieties of Input

TRACE I: Automatically extracted features from CV's spoken by one speaker (15 features, 5 ms windows)

TRACE II: Mock speech (11 slices/segment, mock coarticulation)

Graded mock speech input structure for categorical perception and trading relations simulations

TRACE Simulations

- Ganong effect
- Word-final lexical effects
- Phonotactic effects
- Trading relations
- Categorical perception
- Compensation for coarticulation
- Early activation
- Lexical segmentation

Ganong Effect

Lexical effects only emerge when one lexical candidate dominates. Competition at all levels magnifies activation Differences.

Word-final lexical effects

Figure removed due to
copyright restrictions.

Phonotactic Effects

Phonotactic effects produced by top-down excitation. Partial lexical matches produce gang effects in nonword stimuli.

Graded Input Needed for Category Effects

Figure removed due to copyright restrictions.

Other modifications: Lexical influences removed, phoneme to feature activation added

Trading Relations

Figure removed due to
copyright restrictions.

Additive effects of bottom-up activation produce trading

Categorical Perception

Figure removed due to
copyright restrictions.

Competition sharpens boundaries, effects increase over time.

Categorical Perception: Discrimination

Figure removed due to copyright restrictions.

Application of Luce decision rule further sharpens id functions

Phoneme-feature activation minimizes differences in activation to reduce within category discrimination

Compensation for Coarticulation (TRACE I)

Context dependent reweighing of feature cues dramatically improves performance (75% accuracy without it, 90% with it)

Simulation of data only – Does this reflect spectral contrast? Articulatory parsing? Feature parsing? Higher level perceptual units?

Early Access

Figure removed due to
copyright restrictions.

Competition suppresses candidates as mismatch emerges. Loss of competition leads to rapid rise in activation. Note the early activation advantage for short words (*priest*).

Lexical Segmentation

Figure removed due to
copyright restrictions.

Continuous access IS implicit segmentation.
Competition and decay create systematic biases

TRACE weaknesses

- No role for basic perceptual processes
- No role for learning
- Unrealistic temporal representation
- Questionable phonemic representation, arbitrary choice of feature system
- Incomplete integration (i.e. different inputs, parameters for different problems)
- No role for effects of attention or strategy
- Unrealistically small lexicon
- Little concern for biological plausibility

TRACE's strengths

- Dynamic activation processes
- Raises role of competition
- Framework useful for considering effects of graded and partial activation
- Attempts to address temporal effects including roles of decay, timecourse of activation
- Well articulated account of top-down processes
- Recognizes some role for featural representation