CATEGORICAL SPEECH PERCEPTION REVISITED

Klaus Kohler

Institut für Phonetik und digitale Sprachverarbeitung (IPDS), Kiel, Germany

THE HASKINS PARADIGM OF CATEGORICAL SPEECH PERCEPTION

- Based on linguists' concept of contrastive invariance in phonemes and conditioned variability in allophones,
- projected onto perception by psychologists: the goal of perception is the decoding of the distinctive invariance of segmental phonemes from speech variability
  - listeners categorize an acoustic continuum sharply in an identification task
  - and differentiate acutely across category boundaries but poorly inside.
- The classic identification and discrimination experiments for acoustic continua of place and VOT in plosives were taken as support of a special Speech Code in perception
  - closely linked to categorical separation in production
  - as against gradual acoustic manifestation (Liberman et al.1967)
  - resulting in the Motor Theory of Speech Perception.
- The experimental results were less clear for vowels than for consonants, and seemed to disfavour categorical tonal perception.

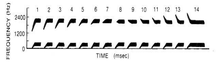

Figure 1: Continuum of 2nd formant transitions.

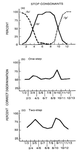

Figure 2: Identification (above), 1- and 2-step discrimination (below).

CATEGORICAL PITCH PERCEPTION

(a) The Categorization of Peak Alignment

Kohler (1987) applied CSP to the perception of F0 contours in German in a peak-shift and semantic contextualization paradigm, using the sentence *Sie hat ja gelogen.*

- There are 3 phonological peak positions
  - ***early***
    associated with finality ("knowing", "coming to the end of an argument")
  - ***medial***
    associated with openness ("observing", "starting a new argument")
  - ***late***
    associated with unexpectedness ("realising in contrast to one's expectation").

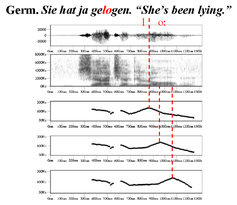

Figure 3: *Early*, *medial* and *late* peak positions.

- A series of test stimuli with peak shifts ranging from ***early*** to ***late*** was generated.
- A verbal context was constructed for the identification of stimuli from the ***early***-to-***medial*** part of the continuum as either matching or not matching.
- Discrimination was tested with 1-step and 2-step pairings in ascending and descending orders along the whole continuum from ***early*** to ***late***.

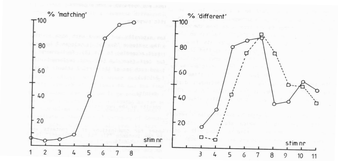

Figure 4:
Identification (left),
2-step discrimination (right):
ascending order continuous line, descending order broken line.

- Results show
  - categorical changes in the identification of ***early vs medial peaks***
  - a discrimination maximum across the category boundary
  - i.e. support for the classic Haskins paradigm.
- They have been reproduced many times, and discrimination across vs inside early/medial categories even works with speakers of diverse languages (tone and intonation), who have no knowledge of German.
- ***Psychophonetic principle of pitch discrimination, separate from functionally determined pitch identification.***

(b) The Categorization of Valley Alignment

Using the same method, Niebuhr and Kohler (2004) compared identification and discrimination of a ***valley-shift*** with a ***peak-shift*** continuum across *und wie ist dein Name*, being mirror images in semitone steps with reference to an initial F0 of 108Hz.

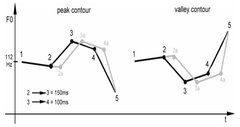

Figure 5:
Schematic representation of peak and valley shifts across *und wie ist dein Name*
"and what's your name"

- There is the usual CSP in identification and discrimination for peak shift.
- Valley shift also shows categorization of opposite ends of the shift scale. But discrimination of 2-step pairings is not significantly different from discrimination of identical stimuli in valley and peak shifts.
- The psychophonetic principle does not apply, but functional categorization does.

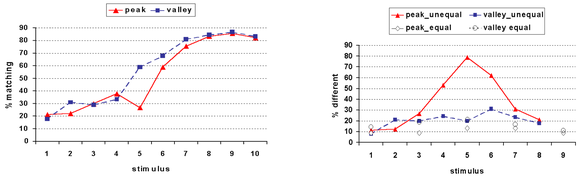

Figure 6:
Identification functions of the peak and valley stimuli (left) and discrimination functions of their (un)equal pairings (right, the number refers to the serial rank of the first stimulus in the pair).

- The psychophonetic and the functional principle can be independent, but may be linked, as in the classic Haskis paradigm, which generalised one perceptual constellation with far-reaching consequences for the theory of speech perception.

(c) The Categorization of a Phrase-final Falling-to-Rising Continuum

- The naturally produced sentence *Alle Jungen spielen Fußball*. "All boys play football." was used as the basis of stimulus generation for discrimination and identification experiments.
- It contained two peak accents, on *alle* and *Fußball* with an F0 dip between them.
- The F0 curve was stylised by 5 significant points (start, first peak maximum, minimum between peaks, second peak maximum (=110Hz), end) with linear interpolation between them.
- The first 4 points were kept constant across the series, the last one was changed in steps of one semitone, from 70Hz to 264Hz, which resulted in 24 stimuli forming a continuum from falling via level to rising pitch on the last word (see Table 1).

Table1.
Phrase-final F0 in the 24 stimuli resynthesized from the natural utterance *Alle Jungen spielen Fußball.*

Sti01	70	Sti07	99	Sti13	140	Sti19	198
Sti02	74	Sti08	105	Sti14	148	Sti20	209
Sti03	78	Sti09	111	Sti15	157	Sti21	222
Sti04	83	Sti10	117	Sti16	166	Sti22	235
Sti05	88	Sti11	124	Sti17	176	Sti23	249
Sti06	93	Sti12	132	Sti18	186	Sti24	264

- In the identification test, subjects classified each stimulus as either *final statement* or *non-final statement* or *question*.
- In the discrimination test, stimuli were paired with a step size of 2 in ascending order, and every third stimulus, starting with Sti02, was paired with itself, for judgement as either same or different.

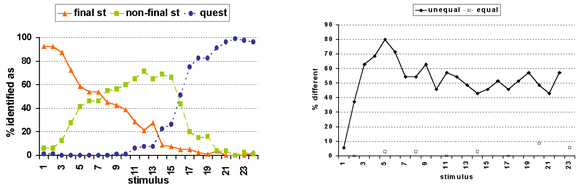

Figure 7:
Identification functions for the stimuli of the phrase-final falling-to-rising series (left) and the discrimination function of their (un)equal pairings (right, the number refers to the serial rank of the first stimulus in the pair).

- A clear categorization of the continuum via function in the identification task is not matched by discrimination maxima at the category boundaries.
- Non-identical stimuli are more often judged different than identical ones, but the discrimination function oscillates around the 50% mark from sti07-09 onwards.
- So again the psychophonetic principle does not seem to be operative in this series in spite of clear functional categorization.

EXPLAINING THE DATA

F0-peak and F0-valley continua differ in the way syntagmatic pitch and articulation contrasts are linked across the landmark of consonant-vowel transition. Early and medial peak contours are terminal F0 falls, with a high-low or a low-high trajectory across the articulatory landmark, and increased acoustic intensity heightening the pitch contrast.

In the valley shift, the decisive pitch difference between early and late final rises is confined to the vowel and thus lacks a tight link with a syntagmatic articulatory contrast. This difference may be taken to determine the different discrimination patterns.

It can be concluded that a discrimination peak is not an inherent feature in the perceptual categorization of a physical continuum but constitutes a separate psychophonetic principle, based on two features:

(1) Syntagmatic contrasts in addition to paradigmatic oppositions of pitch and vocal tract shapes lead to auditory enhancement (Diehl, 1991).

(2) Prosodic patterns are perceived in relation to the acoustic patterns of vocal tract sequencing.

These two features define a different Speech Code from the Haskins one: it transcends the segmental-phonemic orientation and the very specific CSP paradigm. The psychophonetic principle makes it possible for listeners of diverse languages to perceive categorical changes in F0 peak contour synchronization, even without a knowledge of the respective language.

In all cases where pitch patterns are not defined as syntagmatic pitch contrasts in relation to syntagmatic articulatory transitions but as pitch characteristics of syllable nuclei or phrasal position, the psychophonetic principle does not seem to operate in pitch perception, hence the negative results of discrimination tasks in valley alignment and phrase-final rising pitch, in spite of functional categorization in identification tasks. To this list may be added the discrimination of peak height for emphasis (Ladd & Morton, 1997).

Sound perception, too, would only be expected to show discrimination maxima if the definition of the sound category relies on essential syntagmatic features. This would explain the differences found in the categorization of consonants and vowels.

ACKNOWLEDGEMENTS

The author would like to thank the students who participated in his prosody class at IPDS in the summer semester of 2001, particularly Oliver Niebuhr, who supervised the data collection of falling-to-rising pitch categorization. Special thanks go to Benno Peters, who had the idea for this experiment and generated the stimuli.

REFERENCES

Diehl, R. (1991) The role of phonetics within the study of language, Phonetica, 48, 120-133.

Kohler, K. J. (1987) Categorical pitch perception, Proc. XIth International Congress of Phonetic Sciences, Tallinn, 5, 331-333.

Ladd, D. R. & Morton, R. (1997) The perception of intonational emphasis: continuous or categorical?, Journal of Phonetics, 25, 313-342.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967) Perception of the speech code, Psychological Review, 74, 431-460.

Niebuhr, O. & Kohler, K. J. (2004) Perception and cognitive processing of tonal alignment in German, Proceedings of the International Symposium on Tonal Aspects of Languages (TAL2004), Beijing.