

CATEGORICAL SPEECH PERCEPTION REVISITED

Klaus Kohler

Institut für Phonetik und digitale Sprachverarbeitung (IPDS), Kiel, Germany
kjk@ipds.uni-kiel.de

ABSTRACT

CSP postulates perceptual grouping of an acoustic continuum into sharply delimited phonological categories with discrimination maxima across the identification boundaries. The experimental procedure was applied to F0 contours in a peak-shift and semantic contextualization paradigm in German and showed a categorical change from *early* to *medial* position in relation to the accented syllable. But in a comparable valley shift from *early* to *late* a discrimination maximum was not found although there was clear category formation in the identification task. In F0-peak perception a syntagmatic pitch contrast of high-low or low-high, respectively, across the syntagmatic articulatory landmark of consonant-vowel transition, preceding a final fall, is characteristic of *early* vs *medial*. 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 leads to the conclusion that perceptual categorization of a physical continuum is not tied to a discrimination maximum, unless there is an additional association with contrastive vocal tract sequencing. This can also explain differences found in the categorization of consonants vs vowels, and stresses the relevance of *syntagmatic auditory enhancement* beside *paradigmatic phonemic opposition* in speech perception.

THE HASKINS PARADIGM OF CATEGORICAL SPEECH PERCEPTION

Structural linguistics established the concept of contrastive sound units – phonemes – differentiated by distinctive features, as against the redundant features of contextually determined allophones. The psychologists at Haskins took over this segment-oriented view of language and its bipartition into contrastive invariance and conditioned variability, and projected it onto speech perception. Decoding phonemes became the task of the listener, who had to extract the distinctive features of phonemic contrasts from speech variability. This is the theoretical basis that led to the paradigm of categorical speech perception and subsequently to the development of the Motor Theory: listeners would attune to the speech parameters that distinguish phonemes, and thus categorize an acoustic continuum sharply in an identification task; at the same time they would differentiate acutely across the category boundaries, but only poorly inside them (Liberman et al.1957, 1962).

The classic identification and discrimination experiments of acoustic continua referring to place of articulation and VOT in plosives were considered supporting the notion of a special Speech Code in perception, closely linked to categorical separation in production as against gradual acoustic manifestation (Liberman et al.1967). The theory was critically reviewed by Lane (1965). The experimental results were less clear for vowels than for consonants, and seemed to disfavour categorical tonal perception.

CATEGORICAL PITCH PERCEPTION

The Categorization of Peak and Valley Alignment

Kohler (1987) applied CSP to the perception of F0 contours in German in a peak-shift and semantic contextualization paradigm, and showed categorical changes in the identification of *early* vs *medial* peaks, with a discrimination maximum across the category boundary, i.e. support for the classic Haskins paradigm. As the *early* peak was found to be associated with *finality* (“knowing”, “coming to the end of an argument”), the *medial* peak with *openness* (“observing”, “starting a new argument”), appropriate contexts could be constructed for identification of test stimuli such that their intonation either fitted or did not. Discrimination was tested with 1-step and 2-step pairings. The results have been reproduced over and over again. The discrimination pattern across vs inside *early/medial* categories even works with speakers of diverse languages (tone and intonation), who have no knowledge of German and therefore cannot provide a semantic classification, which would be different in different languages anyway. So categorical discrimination is possible in human language without being tied to semantically determined categorical identification. This points to a wide-spread, or even universal, psychophonetic principle of pitch perception in speech. It needs to be kept separate from linguistic and other functional uses of F0 peak synchronization, which vary from language to language to differentiate word tones, sentence modalities and attitudinal/expressive patterns in communication.

The shift and semantic contextualization paradigm was also applied to a continuum of valley contours from an *early* to a *late* synchronization with articulation. Several different experimental designs were used, but CSP was not confirmed by any of them. In the latest experimental series (Niebuhr & Kohler, 2004), peak and valley patterns were constructed in such a way that they were exact mirror images in semitone steps up and down with reference to an initial F0 of 108Hz. The peak shift shows the usual CSP in identification and discrimination; the valley shift also points to a clear categorization at the opposite ends of the shift scale, but discrimination of 2-step pairings is not significantly different from the discrimination of identical stimuli in peak and valley shifts. Figure 1 presents the results.

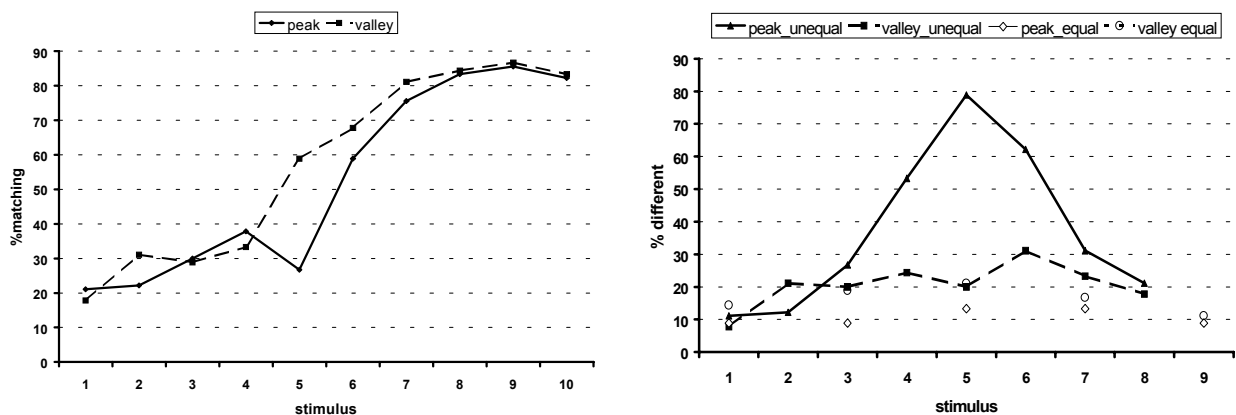


Figure 1. Identification functions of the peak and valley stimuli (left) and discrimination functions of their (un)equal pairings (right, the number refers to the serial rank of the first stimulus in the pair). 18 subjects, 5 repetitions.

So in the case of the valley shift, the psychophonetic principle does not apply, although the functional categorization is clearly there. This means that the psychophonetic and the functional principle can be independent, with either discrimination or functional identification being categorical, but they may also be linked, as in the classic Haskins paradigm and in the peak shift data. The Haskins group generalised one particular constellation with far-reaching consequences for the theory of speech perception.

The Categorization of a Phrase-final Falling-to-Rising Continuum

The absence of categorical discrimination in spite of functional category formation is also shown by results from discrimination and identification experiments with a phrase-final falling-to-rising continuum, carried out by the author with students in a course on prosody at IPDS. The naturally produced sentence *Alle Jungen spielen Fußball*. "All boys play football." was used as the basis for stimulus generation. It contained two accents, realised as peak contours on *alle* and *Fußball* with an F0 dip between them, a maximum of the second peak (on the vowel [u:] of *Fuß*) of 110Hz and a phrase-final value of 70Hz. The F0 curve was stylised by 5 significant points (start, first peak maximum, minimum between peaks, second peak maximum, end) with linear interpolation between them. For stimulus generation, the first 4 points were kept constant across the series, the last one was changed in steps of one semitone, starting from 70Hz and going up to 264Hz, which resulted in 24 stimuli forming a continuum from falling via level to rising pitch on the last word (see Table 1). The stimulus generation was done in *praat*.

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

With reference to the peak maximum of 110Hz, stimuli 01 – 08 represent a series of decreasing falls, stimulus 09 level pitch, and stimuli 10 – 24 a series of increasing rises. For the identification test, the stimuli were repeated 10 times and randomized. Subjects were asked to classify each of the 240 test stimuli as either *final statement* or *non-final statement* or *question* by pressing one of three response buttons in a computerized reaction measuring set-up. For the discrimination test, stimuli were paired with a step size of 2 in ascending order, and in addition every third stimulus, starting with Sti02, was paired with itself. This gave 30 paired test stimuli, which were repeated 5 times and randomized. Using the same equipment, subjects recorded their perception of each of the 150 test stimuli as either *same* or *different*. 8 phonetically naive subjects took part in the identification test, 7 in the discrimination test. Figure 2 shows the results of the two tests.

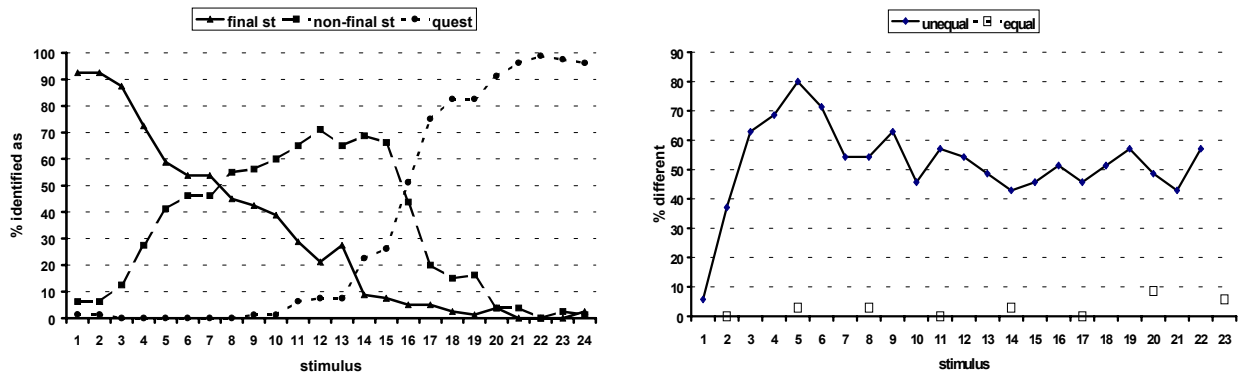


Figure 2. Identification functions for the stimuli of the phrase-final falling-to-rising series (left, 8 subjects, 10 repetitions) and the discrimination function of their (un)equal pairings (right, the number refers to the serial rank of the first stimulus in the pair; 7 subjects, 5 repetitions).

The continuum is clearly partitioned into a *statement* and a *question* section, with a further subdivision of the former into *final* and *non-final (continuation)*, which is less sharply marked, but obviously a perceived category. The first 3 stimuli of the series had enough of a phrase-final pitch fall from the accent peak to trigger an overwhelming finality judgement; thereafter there is a constant increase of the assessment as *continuation* up to stimulus 15, with *question* being quite insignificant to this point, but then dominating the response pattern. The middle range of the responses subsumes less clearly falling, level and moderately rising pitch stimuli. For a clear judgement of *question* a rise of at least 8 semitones seems to be necessary.

This 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, through the whole series, with the exception of the first non-identical stimulus pair, but the discrimination function oscillates around the 50% mark from sti07-09 onwards. From this point on pairings are first of all around level pitch (slightly falling + level, very slightly falling + very slightly rising, level + slightly rising) and then continue to be both rising from the anchor point of 110Hz, with a constant 2-semitone difference within the pair. The response pattern is different, showing a peak, when both stimuli are *falling*, i.e. sti03-05 to sti06-08. The different extents of the fall are no doubt processed by the listener as sounding *more* and *less* terminal, respectively, and for this reason they may be well discriminated. The pairings round level pitch all sound non-final, signalling continuation, and the 2-step pairings of rises cannot span the functional categories of *continuation* and *question*, and consequently all sound equally different.

So again the psychophonetic principle does not seem to be operative in this series in spite of clear functional categorization.

EXPLAINING THE DATA

The difference in F0-peak and F0-valley categorization may be explained with reference to the specific link of syntagmatic pitch and articulation contrasts across the landmark of consonant-

vowel transition. Both *early* and *medial* peak contours are defined by a terminal F0 fall, but the former is characterised by a high-low, the latter by a low-high, F0 trajectory across the articulatory landmark, where, in addition, an increase in acoustic intensity heightens the pitch contrast. This link of a reversal of a syntagmatic pitch contrast with the acoustic output of a syntagmatic articulatory contrast would thus determine discriminatory distinctivity,

In the valley shift, on the other hand, 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. So 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 Speech Code with a different perspective from the one developed at Haskins: 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. For example, a Chinese listener partitions the *early* to *late* peak sequence in the German sentence *Sie hat ja gelogen*. "She's been lying." at the same places as German listeners by referring to changes from tone 3 to tone 4 and finally to tone 2+4 (Kohler 1991, p. 156), without understanding the meaning of the sentences. This psychophonetic principle in pitch perception may be expected to be put to wide-spread use in the languages of the world for a spectrum of functions: word tone, tonal accent (e.g. Swedish), pragmatics in German, English and other languages. Moreover, as the *early* peak focuses on low pitch in the high-low transition into the accented vowel, whereas the *medial* peak focuses on high pitch across the articulatory landmark, the association of this *low* vs *high* pitch with *finality* vs *openness* in German may be seen as another aspect of the *Frequency Code* (Ohala, 1984), related to *dominance* vs *subordination*, and thus to a very general principle of human behaviour.

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 positions, the psychophonetic principle does not seem to operate in pitch perception, hence the negative results of discrimination tasks related to valley alignment and phrase-final rising pitch, in spite of functional categorization established in identification tasks. To this list may be added the discrimination of peak height for emphasis (Ladd & Morton, 1997).

Since the psychophonetic principle is conceived of as being based on syntagmatic contrast, sound perception, too, would only be expected to show discrimination peaks 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. The relevance of syntagmatic contrast is most obvious in place and "voiced/voiceless" oppositions of plosives, which formed the basis for the classic Haskins CSP and for the Motor Theory of Speech Perception. However, the fixation on the segmental phoneme made it impossible to give the syntagmatic domain a central role in speech perception theory, although the coining of the term *encoding* and the reinvention of *co-*

articulation in intensive studies a quarter of a century after it was first proposed by Menzerath and de Lacerda (1933) were attempts to blur the segmental boundaries post hoc. The time has now come to take a more radical approach to the limitations of the segment and the phoneme and to give perception research a new direction for a better understanding of speech communication (Hawkins & Smith 2001).

ACKNOWLEDGEMENTS

The author would like to thank the students who participated in his prosody class at IPDS in the summer semester of 2001, in particular Oliver Niebuhr, who supervised the data collection and report compilation of falling-to-rising pitch categorization. A special vote of thanks goes to research assistant Benno Peters, who had the idea for this experiment in the first place and who generated the stimuli.

REFERENCES

- Diehl, R. (1991) The role of phonetics within the study of language, *Phonetica*, 48, 120-133.
- Hawkins, S. & Smith, R. (2001) Polysp: a polysystemic, phonetically rich approach to speech understanding, *Italian Journal of Linguistics*, 13, 99-188.
- Kohler, K. J. (1987) Categorical pitch perception, *Proc. XIth International Congress of Phonetic Sciences*, Tallinn, 5, 331-333.
- Kohler, K. J. (1991) Terminal intonation patterns in single-accent utterances of German: phonetics, phonology and semantics, *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 25, 115-185.
- Ladd, D. R. & Morton, R. (1997) The perception of intonational emphasis: continuous or categorical?, *Journal of Phonetics*, 25, 313-342.
- Lane, H. (1965) The motor theory of speech perception. A critical review, *Psychological Review*, 72, 275-309.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries, *Journal of Experimental Psychology*, 54, 358-368.
- Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F. (1962) A motor theory of speech perception, *Proceedings of the Speech Communication Seminar*, Stockholm 1962.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967) Perception of the speech code, *Psychological Review*, 74, 431-460.
- Menzerath, P. & de Lacerda, A. (1933), *Koartikulation, Steuerung und Lautabgrenzung*. Berlin, Bonn: Ferd. Dümmlers Verlag.
- 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.
- Ohala, J. J. (1984) An ethological perspective on common cross-language utilization of F0 of voice, *Phonetica*, 41, 1-16.