

CATEGORICAL PITCH PERCEPTION

KLAUS J. KOHLER

Institut für Phonetik und
digitale Sprachverarbeitung
Universität Kiel
2300 Kiel, FRG

ABSTRACT

This paper shows that the paradigm of categorical perception also applies to pitch contours. In LPC-synthesized stimuli, an F0 peak is shifted through an utterance in 30-ms steps. The stimuli of this physical continuum are identified in a contextualization experiment. The response function shows an abrupt change when the F0 peak is moved into the vowel of the stressed syllable. When stimuli from the continuum are paired with 0, 1 or 2 steps between them, the differentiation functions show maxima at the category boundary established by the identification test. The ordering in each pair has an influence on the differentiation function.

INTRODUCTION

The paradigm of categorical perception is well-known in the area of sound segments /1/. It means that a physical continuum of a sound property is partitioned into sections inside which the same category is identified and between which category identification changes. The corollary of this is that differentiation along the physical continuum is sharpest across the category boundaries and weakest inside the categories. The evidence for this phenomenon in the perception of prosodic features, e.g. word tones in tone languages, is contradictory /1,2/, and it certainly has not been demonstrated for utterance pitch contours. To show its relevance in the field of intonation the following experiments were carried out.

PROCEDURE

In the German sentence "Sie hat ja gelogen." ("She's been lying."), with focus stress on the syllable "-lo-" /lo:/, the F0 peak can be on the syllable "ge-", preceding the stress, or at the centre of the stressed syllable, or at its end (cf. /3/). This shift in the F0 peak position is correlated with a change in meaning from 'established' to 'new' to 'emphatic'. A token of this sentence was pronounced by a male speaker, LPC-analyzed, and resyn-

thesized with 11 F0 contours, in which the peak was shifted in 30-ms steps from "ge-" to "-en" (for further details cf. /3/).

Experiment 1.

The first 8 stimuli out of this series of 11 (counting from left to right) were each paired with the preceding context "Jetzt versteh ich das erst." ("Now I understand."; spoken by the same speaker, and LPC-resynthesized). This precursor sets a semantic frame of reference for something new to follow in the test utterance. Since the 8 test stimuli span the continuum of F0 peak positions from "ge-" to the centre of the stressed syllable "-lo-", they either contain the same semantic component as suggested by the context frame, i.e. 'new', or the different meaning feature 'established', which would be appropriate as a summing-up at the end of a chain of arguments, for instance after "Once a liar, always a liar; this also applies to Anne: ...". Thus the chosen context and each of the 8 test stimuli either form a semantic match or they do not. A test tape was prepared with a randomization of 80 pairings of context and test stimuli (8 stimuli x 10 repetitions) and presented to 19 listeners who had to indicate on prepared answer sheets whether context and test sentence were semantically congruous.

Experiment 2.

Stimuli from the series of 11 were paired in such a way that they differed by 0, 1, or 2 steps of F0 peak position. All 1- and 2-step combinations were formed in both orders (2x10 and 2x9, respectively), and supplemented by identical stimulus pairings at the uneven rank numbers in the series (6). Two test tapes were prepared: (I) for the ascending rank order in stimulus pairs (i.e. left-to-right shift of the F0 peak), and (II) for the descending rank order (i.e. right-to-left shift). For each test tape, the 6 identical stimulus pairs were added; the resulting 25 pairs were then repeated once and randomized.

A group of 39 subjects listened to test tape (I), a different group of 34 subjects to test tape (II). Listeners had to indicate on prepared answer sheets whether they perceived a difference between the members of a pair.

RESULTS AND DISCUSSION

Figure 1 gives the identification function for Experiment 1: it shows an abrupt change from "matching" to "non-matching" judgments in spite of the gradual change along the physical continuum, and is thus clearly categorical. The answers "matching" or "non-matching", respectively, can be interpreted as the identification of two sentence meanings: 'summing-up conclusion' (A) versus 'new point of argumentation' (B). Stimuli 1-4 represent semantic category (A), stimuli 6-8 category (B); stimulus 5 is on the border between the two. The latter is characterized acoustically by being the first stimulus in the whole series (from left to right) that has the F0 peak in the stressed vowel /o:/: approximately 30 ms after vowel onset. In the stimuli 1-4, the F0 peak precedes the stressed vowel, and there is thus only an F0 fall in it; in the subsequent stimuli, the F0 fall in the stressed vowel is prefixed by a rise of increasing extent, which at a peak position of 60 ms into the vowel has become prominent enough to signal a different category in an identification task. We thus have a time span of about 60 ms into the vowel where the F0 peak is in a boundary area between two categories, and therefore has an equivocal meaning attached to it.

Figures 2a-c provide the discrimination functions for Experiment 2. The pairs of identical stimuli show a maximum of false alarms at the category boundary found in identification, i.e. for stimulus 5. This is what one would expect if the associated meaning is equivocal: listeners overdifferentiate at the perceptual level when the semantic attribution is unclear. In the pairs of different stimuli in the ascending order, the maximum of discrimination occurs at the category boundary of the identification function, as long as one member lies outside the transition span, i.e. for the pairings 4/5, 5/6; 3/5, 4/6, 5/7. This pattern changes in the pairs with descending order; the maximum is generally shifted to the next higher rank in the stimulus series: 6/5, 7/6; 6/4, 7/5, 8/6. This finding may be related to an upward shift of the transition span, the uncertain boundary area now being around stimulus 6. Such a boundary shift can be explained by perceptual hysteresis under the special conditions of the discrimination test paradigm.

If in a sequence of two segmentally identical utterances, i.e. a repetition of the same word string, two different F0 peak positions are selected from around the category transition as established by the identification test, the listener expects a descending F0 peak order, linked to a semantic shift from the category 'new' to the category 'established', as the unmarked case; a reversal of this order constitutes the marked case in this test frame because the repetition of the sentence suggests the progression from 'new' to 'established'. In this situation, perception becomes less acute to a decrease in the extent of a rising F0 (preceding the fall in the stressed vowel) than to an increase: the category boundary is raised in a right-to-left sequence of peaks, compared to its position determined by identification. Thus, stimulus 5, which lies between the two categories in the identification test, and which seems to stay there in left-to-right discrimination, is incorporated in the 'established' category in the reversed-order discrimination.

The maximum of differentiation between stimuli from an F0 peak position continuum is thus at the transition between identification categories. Therefore, the phenomenon of categorical perception also applies to the field of prosody, in particular to global utterance intonation. At the same time, however, a strong order effect which results from the perceptual testing procedures and which disturbs the differentiation functions has to be taken into account. It is found in segment perception, too, but has largely passed unnoticed because it has not been factored out.

REFERENCES

- /1/ B.H. Repp, "Categorical perception: issues, methods, findings". In: N. J. Lass (ed.), Speech and Language: Advances in Basic Research and Practice. Vol. 10, pp. 244-335. Academic Press: Orlando, 1984.
- /2/ A.S. Abramson, "The noncategorical perception of tone categories in Thai," In: B. Lindblom & S. Öhman (eds.), Frontiers of Speech Communication Research, pp. 127-134. Academic Press: New York, 1979.
- /3/ K.J. Kohler, "Computer synthesis of intonation", Proc. 12th Intern. Congr. Acoustics, A6-6. Toronto, 1986.

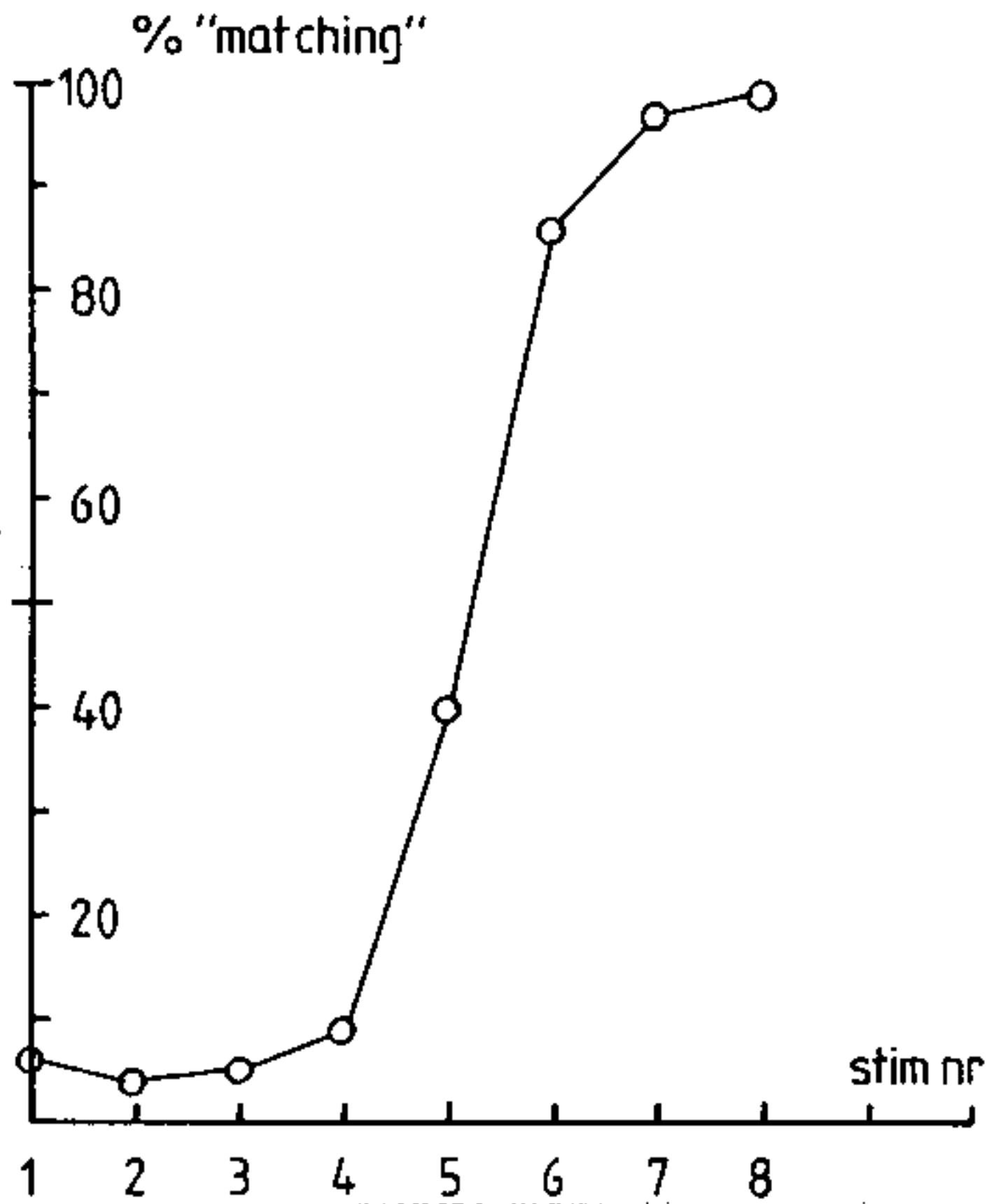


Fig. 1. Identification function in Experiment 1, showing percentage of "matching" judgements for 8 stimuli "Sie hat ja gelogen." with F0 peak shift from left to right in the context "Jetzt versteh ich das erst." 19 subjects; for each stimulus N=190.

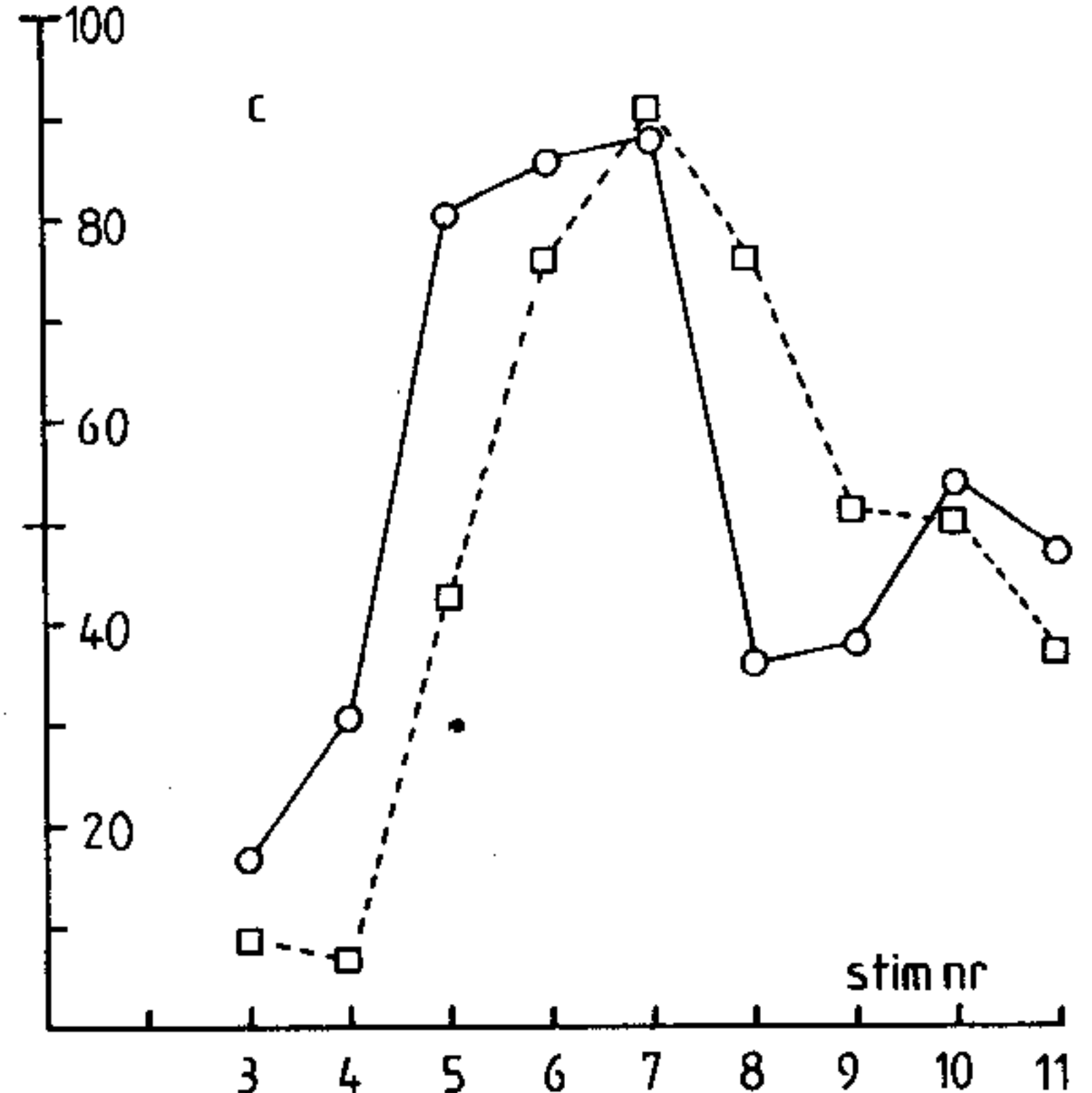
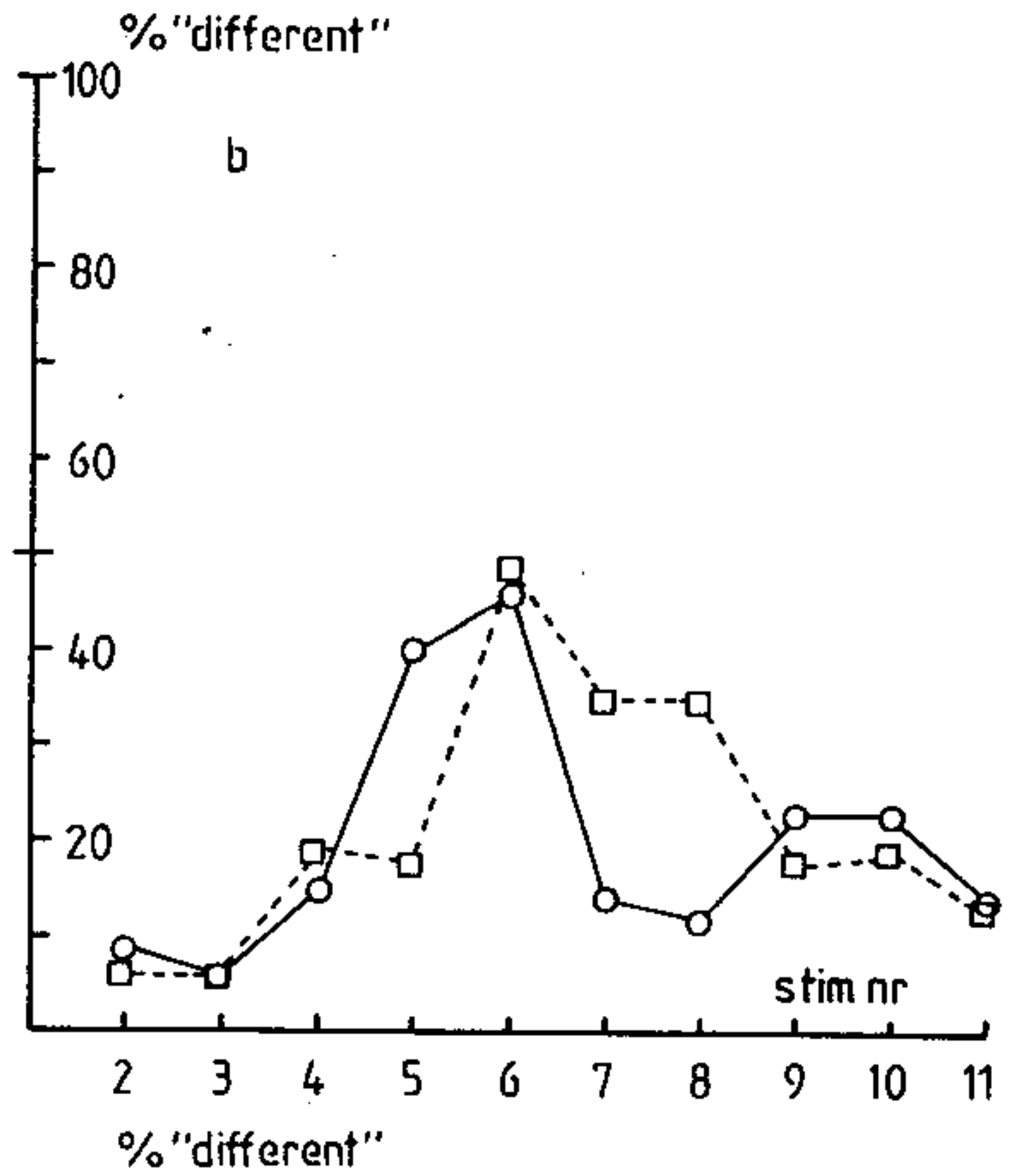


Fig. 2. Discrimination functions in Experiment 2, showing percentage of "different" judgements for utterance pairs of "Sie hat ja gelogen." with 0-(a), 1-(b), or 2-step (c) distances of F0 peak positions, in the ordering left-to-right (continuous line) or right-to-left (broken line). The stimulus numbers refer to the second stimulus in the ascending and to the first in the descending order. 73 sbs., N=146 at each data point (a); 39 sbs., N=78 in the left-to-right, 34 sbs., N=68 in the right-to-left ordering of (b) and (c).

