

3. Fassung

MACRO AND MICRO F0 IN THE SYNTHESIS OF INTONATION

KLAUS J. KOHLER

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

ABSTRACT

This paper distinguishes between a global utterance intonation, which is related to the meaning to be conveyed, and local F0 perturbations, which are due to articulatory constraints. It selects one problem from each field: (1) the perceptual interaction of the stress and intonation functions of macro F0 with reference to experimental data from German minimal verb pairs ("umlagern" with prefix or stem word stress in different sentence-semantic intonations), (2) the perceptual importance of level and falling F0 in prestop vowels for the distinction of fortis and lenis with reference to experimental data from German and English.

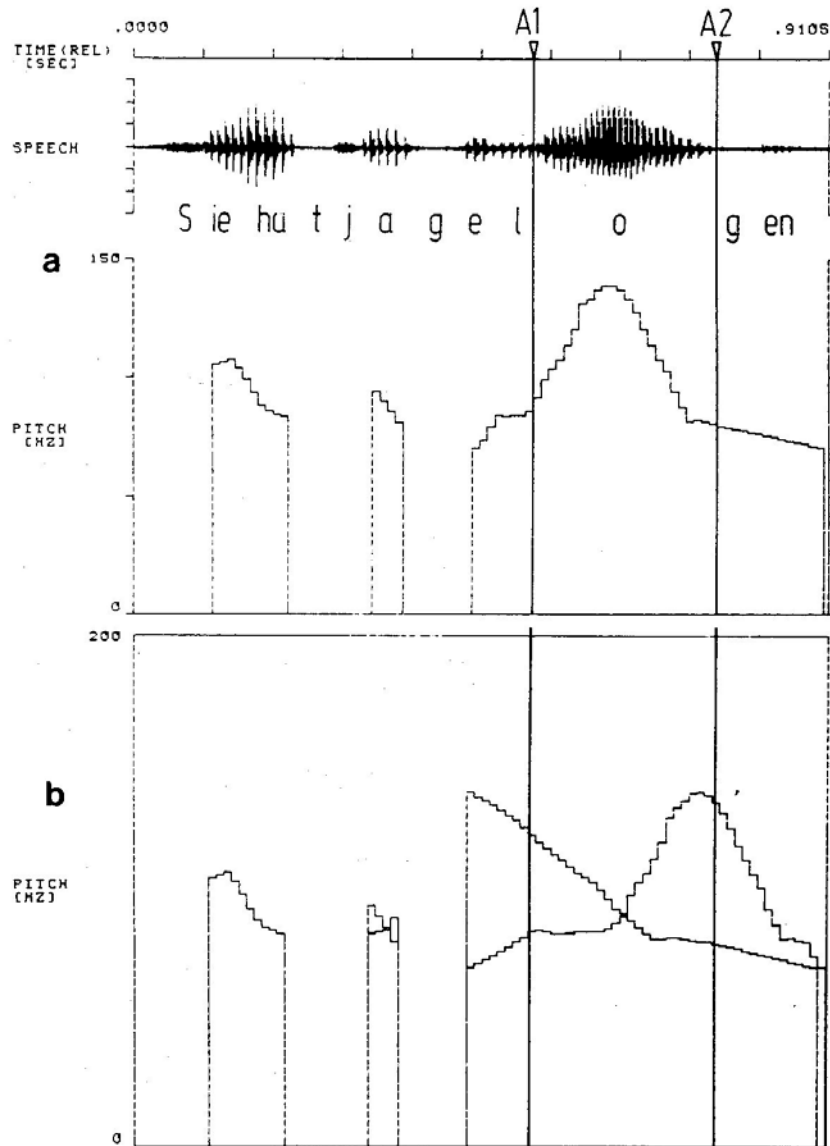
## 1. THE FIELDS OF INVESTIGATION

F0 time courses can signal stress and intonation, and they do this with a certain degree of variability within the same linguistic pattern. Two complementary questions arise in connection with this variability. First, when do changes in the F0 contour (e.g. the shift of an F0 peak) across the same segmental string effect changes in linguistic patterning (= macro F0)? Here three cases have to be distinguished.

(a) In figs. 1a,b, the position of the F0 peak in the centre of the vowel of the syllable "-lo-" or in the syllable "ge-" or late in the syllable "-lo-" of the German sentence "Sie hat ja gelogen." (= "She's been lying.") does not change the word and sentence stress, but alters the intonation with the corresponding changes of meaning from 'established' for the early to 'new' for the central to 'emphatic' for the late F0 peak. See Kohler (1986b) and Kohler (1987) for experimental data.

Figs. 1a,b about here

(b) In figs. 2 and 3a, the position of the F0 peak in the syllable "-la-" or "um-" of German "umlagern" changes the verbal stress pattern from stem to prefix stress, but represents the same late intonation peak within each word accent.



**Fig. 1**

(a) Speech wave and fundamental frequency (centre peak) in the naturally produced utterance "Sie hat ja gelogen." The end contour (on the syllable gen) was added by F0 parameter manipulation because the analysis did not provide it. The time markers A1, A2 delimit the F0 peak contour (coinciding approximately with /o:/) which was shifted left and right in (b).

(b) The left- and right-most positions of the shifted F0 peak contour on the same time scale as in (a), approximating the natural productions of early and late peaks (on ge- and -lo-), respectively.

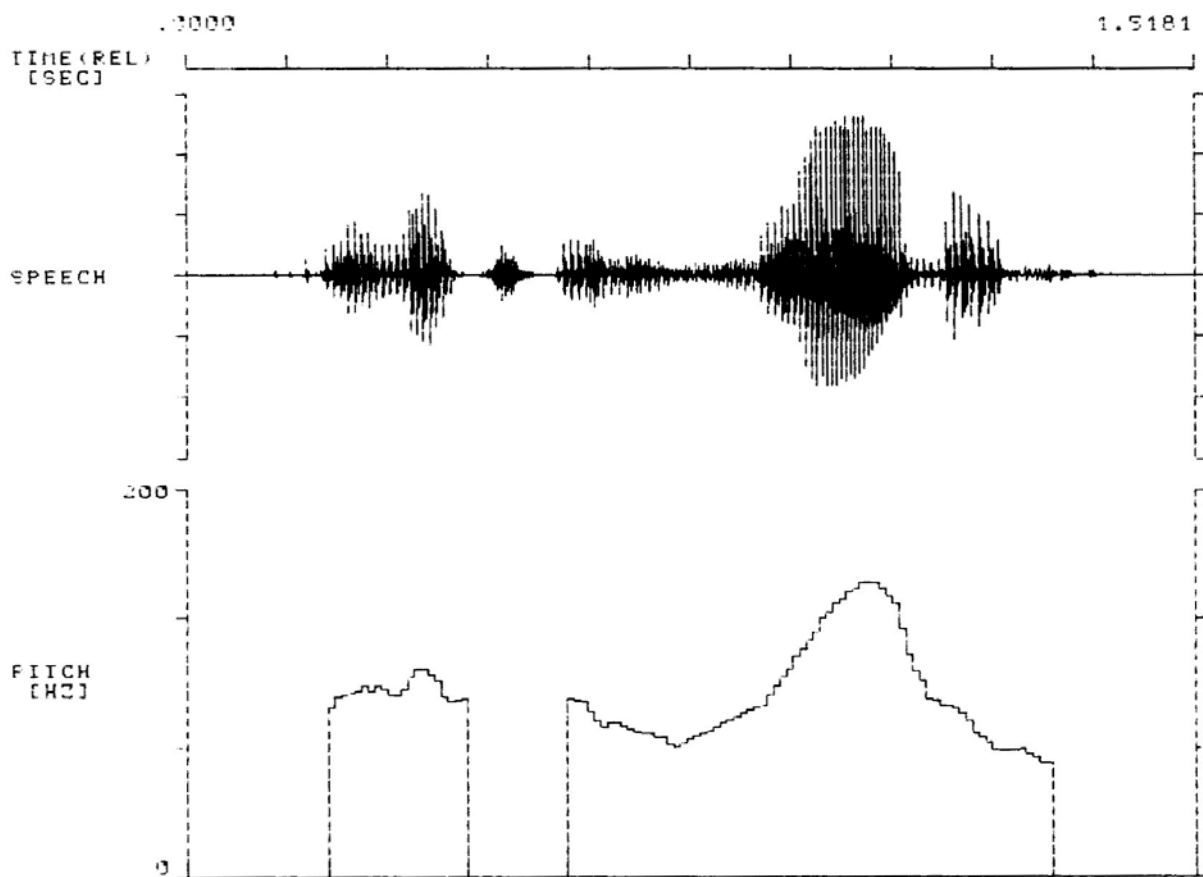
Fig. 2 about here

(c) Figs. 3a,b show changes in the stress pattern from prefix to stem combined with changes in intonation from late to early peak. This paper deals with the perceptual interaction of these stress and intonation functions of macro F0.

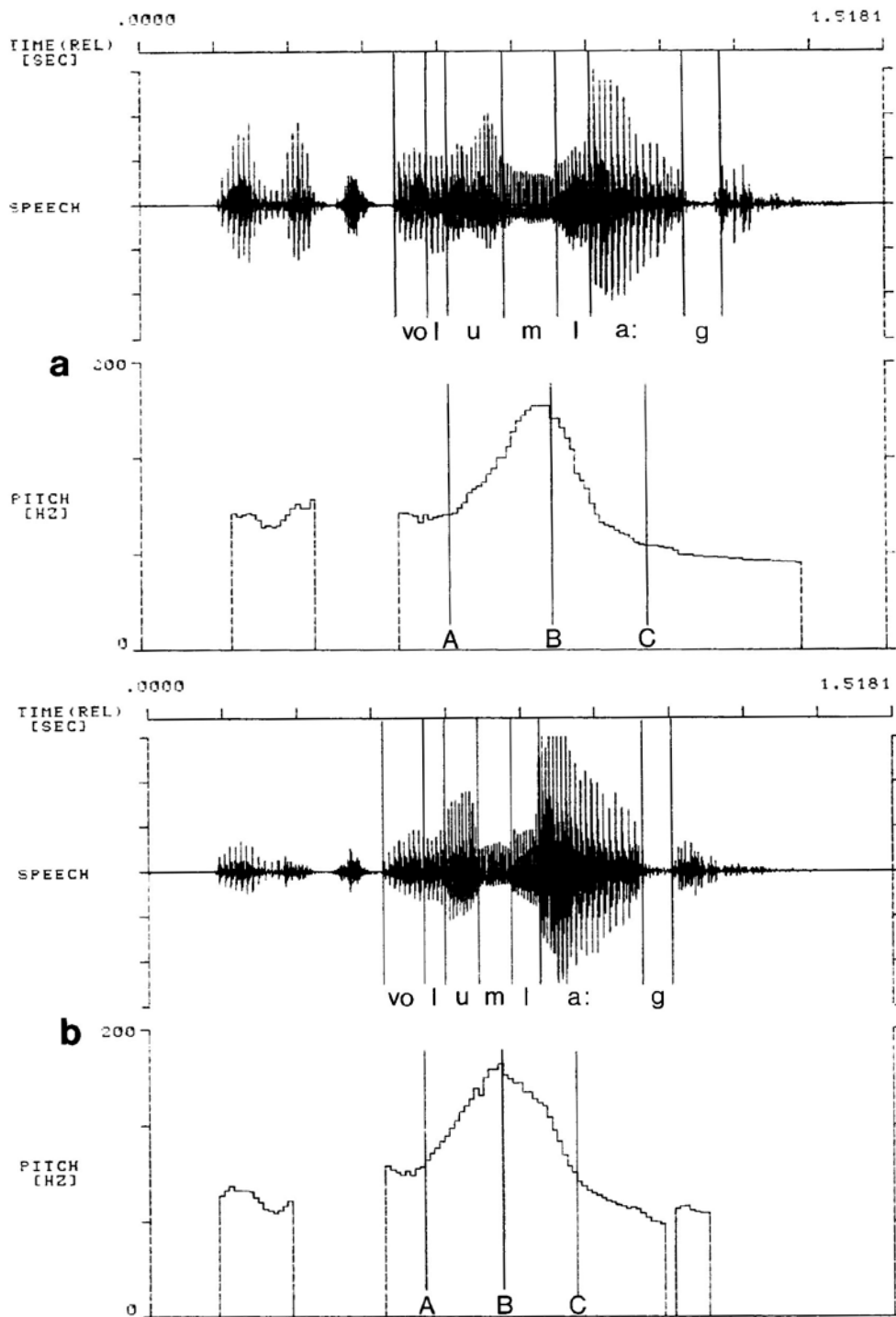
Figs. 3a,b about here

The second question concerns the changes in the F0 contour that have to take place to guarantee the identity of a linguistic stress or intonation pattern across different segmental strings (= micro F0). Thus, figs. 4a,b represent the same stress and intonation patterns (central intonation peak on the last, stressed syllable) for the German sentences "Sie malt." [zɪ 'ma:lt] (= "She paints.") and "Sie schickt." [zɪ 'ʃɪkt] (= "She sends.") in spite of the quite different F0 contours. Here articulatory constraints cause local F0 adjustments to the same underlying global intonation (macro F0), i.e., the influence of high vs. low vowels and of voiceless vs. voiced consonants. If these micro F0 differences are ignored in speech synthesis, four consequences may arise: the macro pattern changes, the linguistic identity of segments changes, the overall sound quality of the synthesized utterances changes, or there is no perceivable change at all (Silverman, 1987).

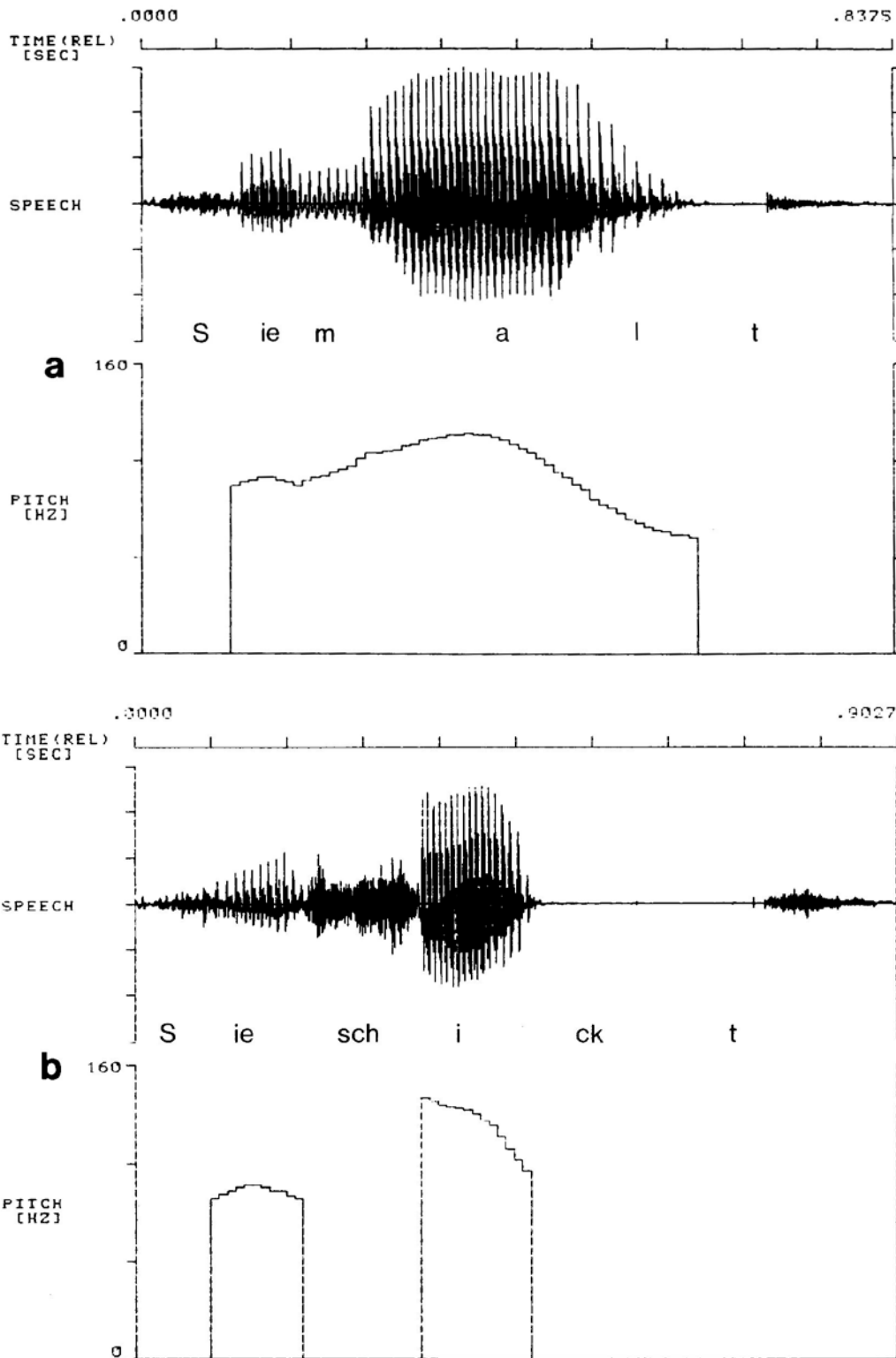
Figs. 4a,b about here



**Fig. 2.** Waveform and F0 of the stem-stress utterance of "Er wird's wohl umlagern." with late intonation peak.



**Fig. 3.** Waveform and F0 of the original prefix-stress with late intonation peak (a) and of the original stem-stress with early intonation peak (b) in "Er wird's wohl umlagern." A, B, C mark the base and peak points of the F0 peak contour for the F0 shifts.



**Fig. 4.** Waveform and F0 of "Sie malt." (a) and "Sie schickt." (b) with stress and central F0 peak on the second syllable in each case.

The two questions concerning macro and micro F0 are empirical questions which can only be solved in the laboratory, and they have to be addressed simultaneously because the measurable speech production output is the result of both factors, and it must be our aim to abstract the linguistic macro F0 patterns from this output and to provide a set of rules for microprosodic adjustments. Proceeding in this way we not only improve our knowledge of prosodic structures in different languages, but we also lay a firm foundation for the improvement of speech synthesis, e.g. in text-to-speech systems. At the same time, the use of speech synthesis as a research tool allows us to test hypotheses about the contribution of F0 to the macro and the micro patterning.

Changes in the terminal falling macro pattern may involve the following parameters:

- the positions of F0 peaks along the time scale in relation to the syllable and segment chains,
- the number of F0 peaks,
- the absolute and relative heights of F0 peaks,
- the shapes of F0 peaks: slow/fast rises/descents,
- the precursor before the first peak,
- the tail.

The first part of this paper deals with macro F0, including the variables of F0 peak alignment and of F0 peak shape. The number of F0 peaks is restricted to one and its height kept constant;



precursor and tail are dependent on the alignment variable. The two variables are studied with regard to their influence on stress and intonation perception in German.

Changes in the micro F0 may involve the following parameters:

- vowel intrinsic F0 due to vowel height,
- contextual F0 due to preceding consonants,
- contextual F0 due to following consonants,
- number of syllables,
- F0 masking in voiceless segments.

The second part of this paper looks at the preconsonantal micro F0 with regard to its influence on the perception of the following segment, an effect that has largely been ignored in microprosodic studies. It reaches its conclusions on the basis of data from German and English.

## 2. MACRO F0

### 2.1. Introduction.

In view of the fact that a shift of the F0 peak position from one syllable to another can, but need not, change the stress position in a syllable chain and can also alter the sentence intonation, two questions arise:

- (a) Under what conditions is an F0 peak shift (without concomitant changes in sound duration and intensity) sufficient to shift stress to a different syllable?

- (b) How can the stress and intonation functions of F0 peaks be differentiated, and in what ways do they interact?

To provide answers to these questions two experiments were carried out in German, which offers good examples for testing the issues because it has minimal verb pairs, with either prefix or stem stress, which can occur in the same natural sentence frame, e.g. "Er wird's wohl umlagern." (with stress either on "um-" /'um/ = "verlagern", "He is presumably going to shift it to another place."; or on "-la-" /'la:/ = "belagern", "He is presumably going to besiege it.").

## 2.2. Procedure.

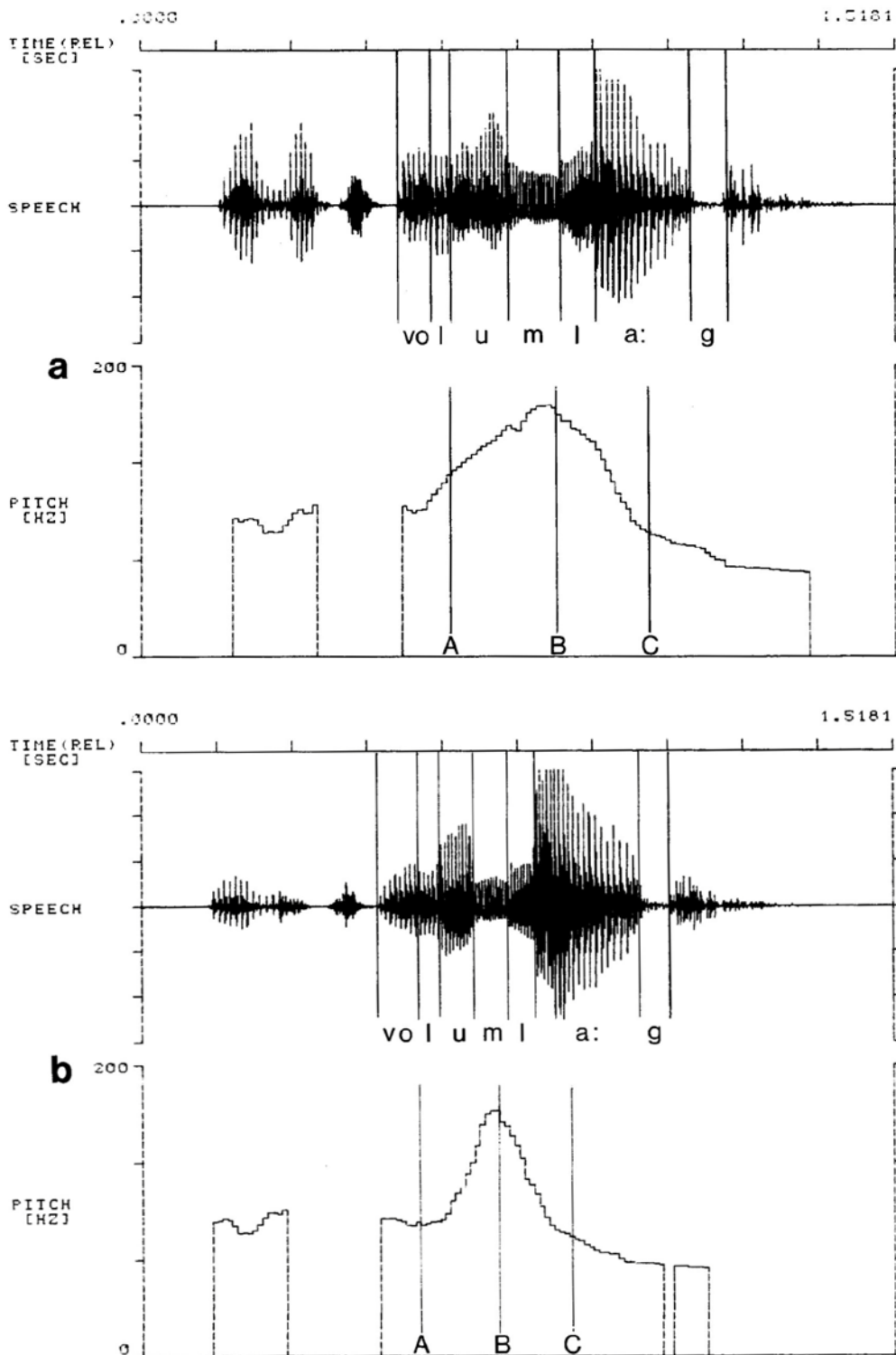
Two utterances of this sentence, (a) with stress on "um-" and a late intonation F0 peak on this syllable, and (b) with stress on "-la-" and an early intonation F0 peak, which actually falls on the syllable "um-", were selected for stimulus construction from a large corpus containing several repetitions of all the 6 combinations of 2 stress and 3 intonation positions, spoken by a trained phonetician (the author). The two tokens were analyzed using the same procedure as in Kohler (1986b). Figs. 3a,b present the waveforms together with their F0 displays. The F0 peak positions in the two utterances are practically identical in relation to the syllable structures of "umlagern": they occur at more or less the same time interval before the beginning of /l/. The differences between the two are in the shapes of the F0 peak contours and in the syllable durations. In the utterance

with stem stress the post-peak F0 descent is more gradual. Also, the "um-" is much shorter (135 ms vs. 222 ms) so that, although the initial (rising) portion of the peak (segment AB) is also somewhat shorter, its onset is much earlier relative to the "um-" syllable, occurring at the beginning of the /l/ in "wohl" rather than at the beginning of the "um-" syllable itself, as in the utterance with prefix stress. The "-la-" syllables, on the other hand, have very similar durations in the stem and prefix stress words (268 ms vs. 258 ms).

Subsequently, the F0 peak contours of the two utterances were exchanged and adjusted to the comparable points in the segmental structures. Figs. 5a,b show the waveforms of figs. 3a,b with the new F0 contours. Finally, the following F0 parameter manipulations were performed:

Figs. 5 a,b about here

- (1) In the stimulus of fig. 3a (original prefix stress), the whole peak contour between the marks A and C was shifted to the right along the time axis in 6 equal steps of 30 ms; the tail of the F0 contour beyond mark C was then time-compressed between the new time position C' and the end of periodicity, and the F0 precursor in "wohl" was time-expanded from its beginning point to the new time position A'. The left branch of the peak contour (AB) was also shifted to the left in 5 equal steps of 30 ms; the right

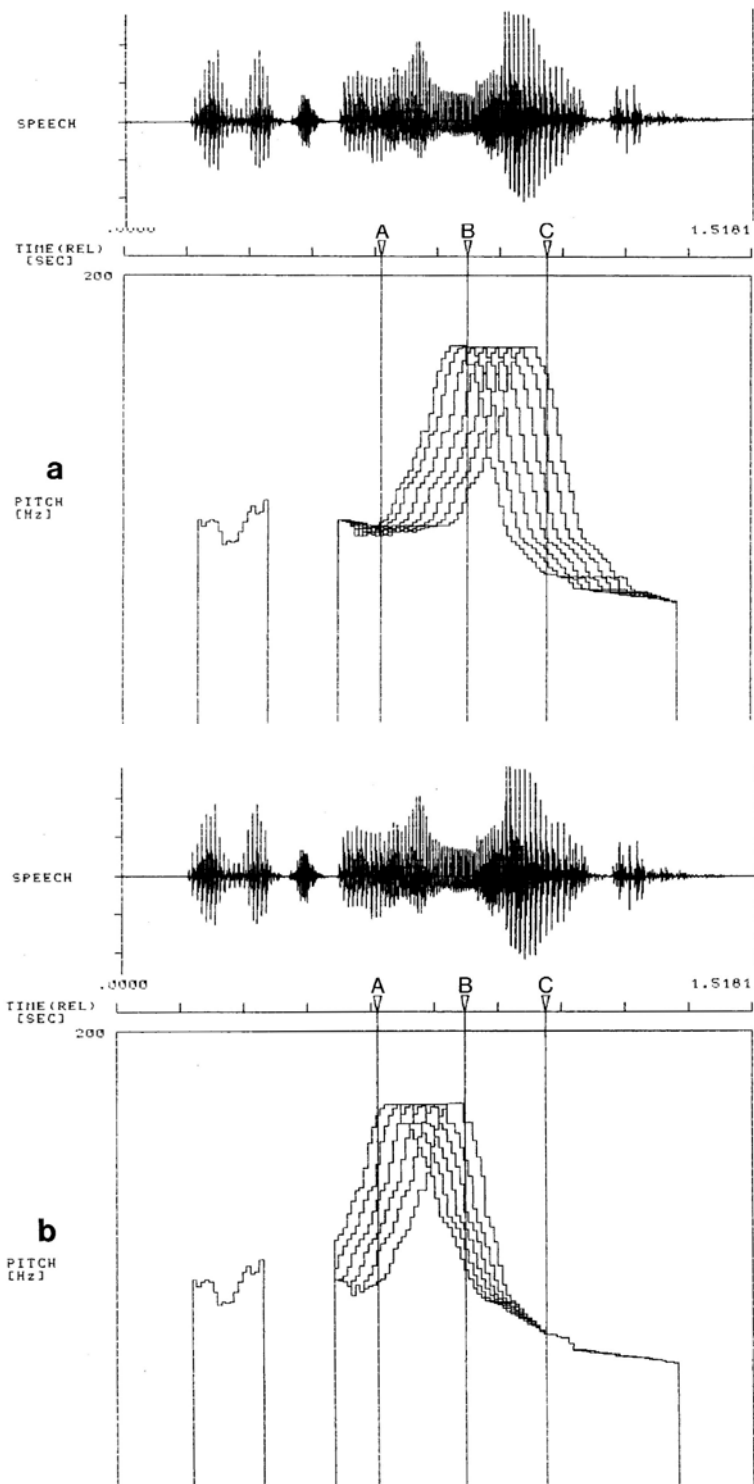


**Fig. 5.** Waveform of the original prefix-stress (a) and of the original stem-stress (b) utterance of "Er wird's wohl umlagern." in fig. 3 with exchanged F0 contours, adjusted to the different timing of the new utterance. A, B, C as in fig. 3.

branch of the peak contour was then time-expanded between the new time position B' and the time mark C, and the precursor was time-compressed between its beginning and the new time position A'. When A' fell to the left of the beginning of "wohl" the section of the contour that thus entered the voiceless stretch was masked. Figs. 6a, b illustrate the right and left shifts, respectively. In the left shifts, the fall was stretched out and thus flattened because that is what was found in the natural productions of the early-peak contours of "Sie hat ja gelogen." (see fig. 1) and of "Er wird's wohl umlagern." with stem stress (see fig. 3b). Furthermore, the flattening of the F0 descent improved the quality of the synthesized stimuli because it prevented the low F0 tail from becoming too long.

Figs. 6a, b about here

- (2) In the stimulus of fig. 3b (original stem stress), the whole peak contour between the marks A and C was shifted to the left in 8 equal steps of 30 ms; the tail of the F0 contour beyond mark C was then time-expanded between the new time position C' and the end of periodicity. As regards the left-branch adjustment the same procedure was followed as in the left shifts of (1). In the construction of this stimulus set, the post-peak F0 fall was not stretched out (as it was in (1)), because a pilot test had shown that a further flattening of the already less abrupt F0 fall in the



**Fig. 6.** The original F0 peak position of the prefix-stress utterance with late intonation peak (fig. 3a) together with the 6 rightward shifts (a) and together with the 5 leftward shifts (b) in the construction of test series (1).

original stem-stress utterance with an early peak does not effect a shift to prefix stress. So, if a shift was to occur at all the F0 descent would have to take place outside the stem syllable.

(3) In the stimulus of fig. 5a (original prefix stress with transferred F0 peak shape), the same F0 peak shifts were carried out as in (1).

(4) In the stimulus of fig. 5b (original stem stress with transferred F0 peak shape), the same procedure was followed as in (2).

From these parameter manipulations, there resulted 12 F0 contours, with peak positions from near the beginning of "um-" to the second half of "-la-", in (1) and (3), and 9 F0 contours, with peak positions from the beginning of "wohl" to near the end of "um-", in (2) and (4). These F0 contours entered into a stimulus synthesis with the LPC-derived formant and volume values of the original prefix-stress utterance in (1) and (3), and with the corresponding data of the original stem-stress utterance in (2) and (4). In each case, two test stimulus sets were thus generated, with a slowly and an abruptly falling F0 peak contour, respectively: (2), (3) vs. (1), (4). The slow fall is characteristic of the early-peak intonation illustrated in fig. 3b. Series (2) is based on the utterance in fig. 3b, i.e. the stem-stress pattern with early peak, whereas series (3) is

based on the utterance in fig. 5a, which has the prefix-stress segments and duration pattern combined with the slowly falling early-peak shape transferred from the utterance in fig. 3b. The abrupt fall was characteristic of the late-peak intonation illustrated in fig. 3a. Series (1) is based on the utterance in fig. 3a, i.e. the prefix-stress pattern with late peak, whereas series (4) is based on the utterance in fig. 5b, which has the stem-stress segments and duration pattern combined with the more steeply falling late-peak shape transferred from the utterance in fig. 3a.

In (1) and (3), the F0 peak positions straddle the syllable structures where a change from prefix to stem stress is to be expected if F0 is a sufficient cue. The two sets differ in that the peak shape of (3), but not of (1), approximates the configuration found in the early peak of the original stem-stress utterance (see fig. 3b). It is hypothesized, therefore, that if stress is perceptually shifted at all in (1) and (3), there will be a more clear-cut change in (1) because there is a higher probability in (3) that an F0 peak position on "um-" is not only perceived as a central or late peak with prefix stress but also as an early peak with stem stress. The same would apply to (2) as against (4).

To check these hypotheses two test tapes were compiled: (I) containing the 12 stimuli of (1) and the 9 of (4), (II) containing the 12 stimuli of (3) and the 9 of (2). (I) was



produced in a short version with 5 repetitions of the 21 stimuli, and in a long version with 10 repetitions, with separate randomizations of the 105 and 210 test stimuli, respectively. (II) was only produced in a short version. Each stimulus sentence was preceded by a bleep and followed by a 4-s pause in which subjects were to answer, by ticking the appropriate boxes on prepared response sheets, whether the meaning of the perceived stimulus was "belagern" or "verlagern". 18 subjects did test (I) in its long version, 9 in its short one. 4 of the 18 deviated in their responses by judging the 9 stimuli of (4) exclusively as "verlagern". They were, therefore, dealt with separately and not included in figs. 7 and 8. 16 subjects, some of whom had done test (I), took test (II) in later sessions. The subjects listened to the test tapes in several subgroups via a loudspeaker in a sound-treated room of the Kiel Phonetics Institute.

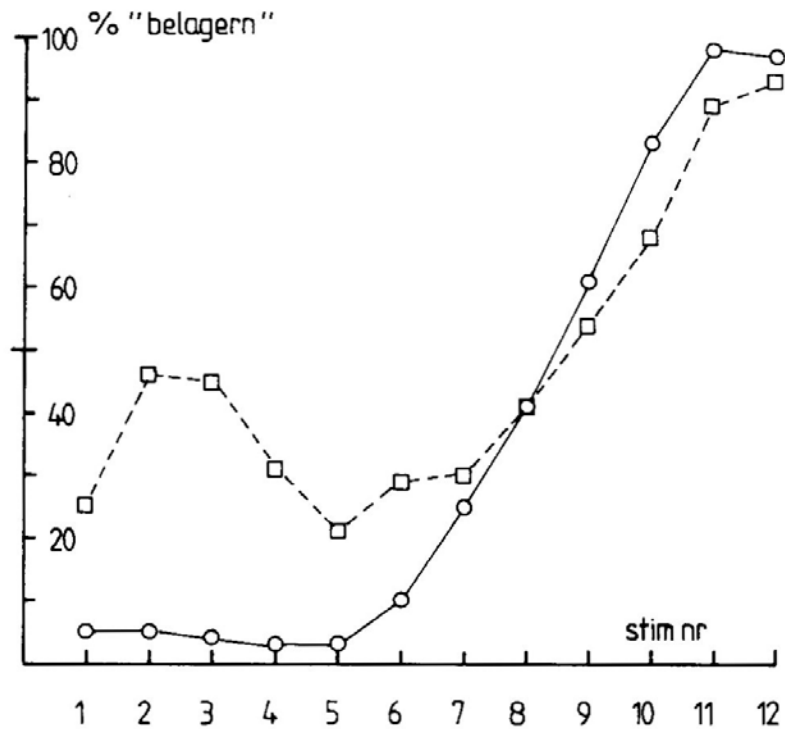
### 2.3. Results and discussion.

Figs. 7 and 8 present the results from these experiments for the 12-stimulus sets with the original prefix-stress duration pattern (series (1), (3)) and for the 9-stimulus sets with the original stem-stress duration pattern (series (2), (4)), respectively. In the shift of the more sharply falling (original) F0 peak contour through the original prefix-stress utterance (see the unbroken line in fig. 7), there is a clear change from initial to stem stress, in spite of the duration of "um-" pointing to the former. F0 can thus override duration,

particularly since the duration of the unstressed "-la-" syllable in the original utterance is very close to its duration under stress. In stimulus 10, which is the first in the ordering from 1 to 12 to yield an unequivocal stem-stress categorization with over 80% positive responses, the F0 peak position is 30 ms into the vowel of the syllable "-la-". This corresponds to the data, found in Kohler (1986b) and Kohler (1987), concerning the change from an early to a central intonation peak on the stressed syllable. The fact that the change from one stress category to the other is gradual rather than categorical can be related to some interaction of the stress and intonation functions of F0 because the more sharply falling F0 peak assumes positions before the beginning of the syllable nucleus /a:/ of "-la-" which can simultaneously function as the central or late intonation peak in stressed "um-" and as the early intonation peak related to stressed "-la-". When the more slowly falling F0 peak is substituted (see the broken line in fig. 7) the initial-stress category is not clearly represented: the interpretation of an early intonation peak for stem stress is never completely precluded.

Fig. 7 about here

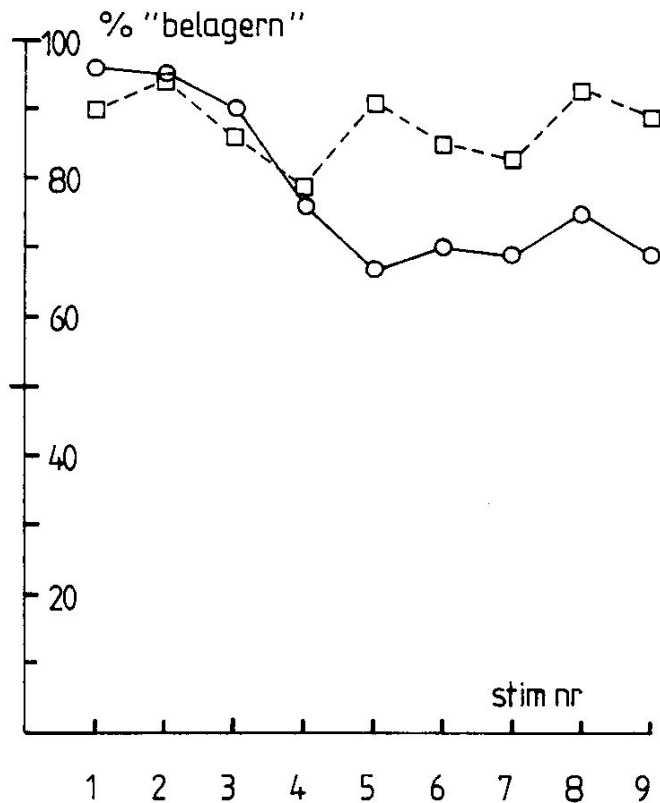
When an F0 peak contour is shifted through the original stem-stress utterance (see fig. 8), there is no change between the stress categories: the answers are predominantly in favour of stem stress. In this case, F0 can thus not override the



**Fig. 7.** Percentage stem-stress responses for "umlageren" (= "belagern") in the series of 12 F0 peak positions (from left to right) combined with the original prefix-stress utterance of "Er wird's wohl umlagern."; original, sharply falling peak contour (continuous line, at each data point  $N=14 \times 10 + 9 \times 5 = 185$ ), and slowly falling peak contour, transferred from the original stem-stress utterance (broken line, at each data point  $N=16 \times 5 = 80$ ).

duration cue completely because "um-" is too short in relation to "-la-" to signal initial stress. There is some effect of F0 when the more sharply falling F0 peak (circles connected by solid lines in fig. 8) occurs actually within the syllable "um-". That is, in stimuli 1 to 5, the F0 peak has been shifted leftward all the way into the preceding syllable "wohl", whereas in 6 to 8 it has been moved only as far back as some point within the prefix syllable "um-", and in these stimuli there are up to 30 % judgements of prefix-stress. This pattern can be interpreted as meaning that the overriding salience of the duration cue is checked somewhat when the characteristic peak contour occurs in the relevant syllable, allowing the interpretation of the transferred abruptly falling late-peak intonation contour as a central or late peak on "um-". In the other series (squares connected by dashed line in fig. 8), however, the slowly falling F0 contour reduces the possibility of interpreting the peak as a central or late peak for a prefix-stress pattern, because of the interference of the original early peak intonation type. These response curves can be contrasted to the four subjects that behaved differently and were not included in the figure. Those subjects seemed to have been guided only by the F0 without any reference to the short duration of the unstressed prefix syllable, and therefore perceived prefix stress on all 9 stimuli because of the presence of the F0 peak before the beginning of the "-la-" syllable.

Fig. 8 about here



**Fig. 8.** Percentage stem-stress responses for "umlagern" (= "belagern") in the series of 9 F0 peak positions (from left to right) combined with the original stem-stress utterance of "Er wird's wohl umlagern."; original, slowly falling peak contour (broken line, at each data point  $N=16 \times 5=80$ ), and sharply falling peak contour, transferred from the original prefix-stress utterance (continuous line, at each data point  $N=14 \times 10+9 \times 5=185$ ).

The hypotheses that led to the experiments discussed above have thus been confirmed, and the questions asked initially can be answered as follows:

- (a) An F0 peak shift by itself is sufficient to bring about a clear change from one stress position to another, provided the duration of the stressed-syllable-to-be is not too short. But even when it is there is a residual F0 effect.
- (b) The intonation function of F0 interferes with its stress function if the latter is not supported by duration. This finds its expression in a gradual change from one stress category to another over a stretch of utterance where the positions of a central intonation peak in one stressed syllable and an early intonation peak related to a stressed syllable following can coincide. This interaction is strengthened when the shape of the F0 peak contour approximates the more slowly falling one of the early intonation peak of a later stress.

### 3. MICRO F0

#### 3.1. Introduction.

The importance of F0 after stop release as an acoustic cue for the lenis/fortis categorization of stop consonants has been

known for a long time (Hombert et al., 1979). F0 preceding the stop closure, on the other hand, has not been attributed a similar cue value. For German it has been demonstrated in a number of experiments with the utterances "Diese Gruppe kann ich nicht leiden/leiten." ("I cannot stand/lead this group.") that in production as well as in perception a level and a level + falling F0 contour on the prestop vowel are cues for /t/ and /d/, respectively (Kohler, 1985). These results have been only partially replicated for English in the utterances "I am telling you I said widen/whiten." with very much smaller effects (Kohler, 1986a). This difference was related to the fuzziness of the segment boundary in /w/ + /ae/ as against /l/ + /ae/ and to the fact that long initial formant transitions have been found to increase the perceived duration of a following vowel. To test this hypothesis, three perception experiments were carried out. In the first one, the previous German test was repeated (a) with another German group in order to demonstrate the generalizability of the discovered signal/perception link for German, (b) with a group of British English speakers in order to show up any perceptual differences due to language background, and to establish a base-line for the other two experiments, which (1) replicated the segmental chain and the F0 patterns of the German test items (/ 'laedn/ - / 'laetn/) in an English sentence frame, and (2) compared its results with those for / 'waedn/ - / 'waetn/.

### 3.2. Experiment 1.

3.2.1. Procedure. The test tape of experiment 2 of Kohler (1985) contained a randomization of 10 repetitions of 21 stimuli "Diese Gruppe kann ich nicht leiden/leiten.", being three sets of the same 7 complementary vowel and following silence durations from clear /d/ to clear /t/ in the utterance-final word, combined with three F0 patterns across the stressed vowel /ae/: level + falling, level, and continuously falling. This tape was presented to a new group of 16 native speakers of German (students of phonetics and languages), in several subgroups, via a loudspeaker in a sound-treated room of the Kiel Phonetics Institute. They classified the stimulus utterances as "leiden" or "leiten" sentences by ticking the appropriate boxes on prepared answer sheets. 13 British English speakers performed the same test under the same conditions in two subgroups, but they gave their answers by pressing one of two buttons at the recording stations of a reaction-time measurement system. They were students of German spending 6 months in Kiel to improve their proficiency in the language.

3.2.2. Results and discussion. The German group replicates the results of the previous test (cf. Kohler, 1985, pp. 24ff) in every respect: fig. 9 again shows that level F0 introduces a /t/, level + falling F0 a /d/ bias, compared with the continuously falling pattern. As can be seen in fig. 10, the English group also shows clearly separate identification functions for level and falling F0. But it has a higher

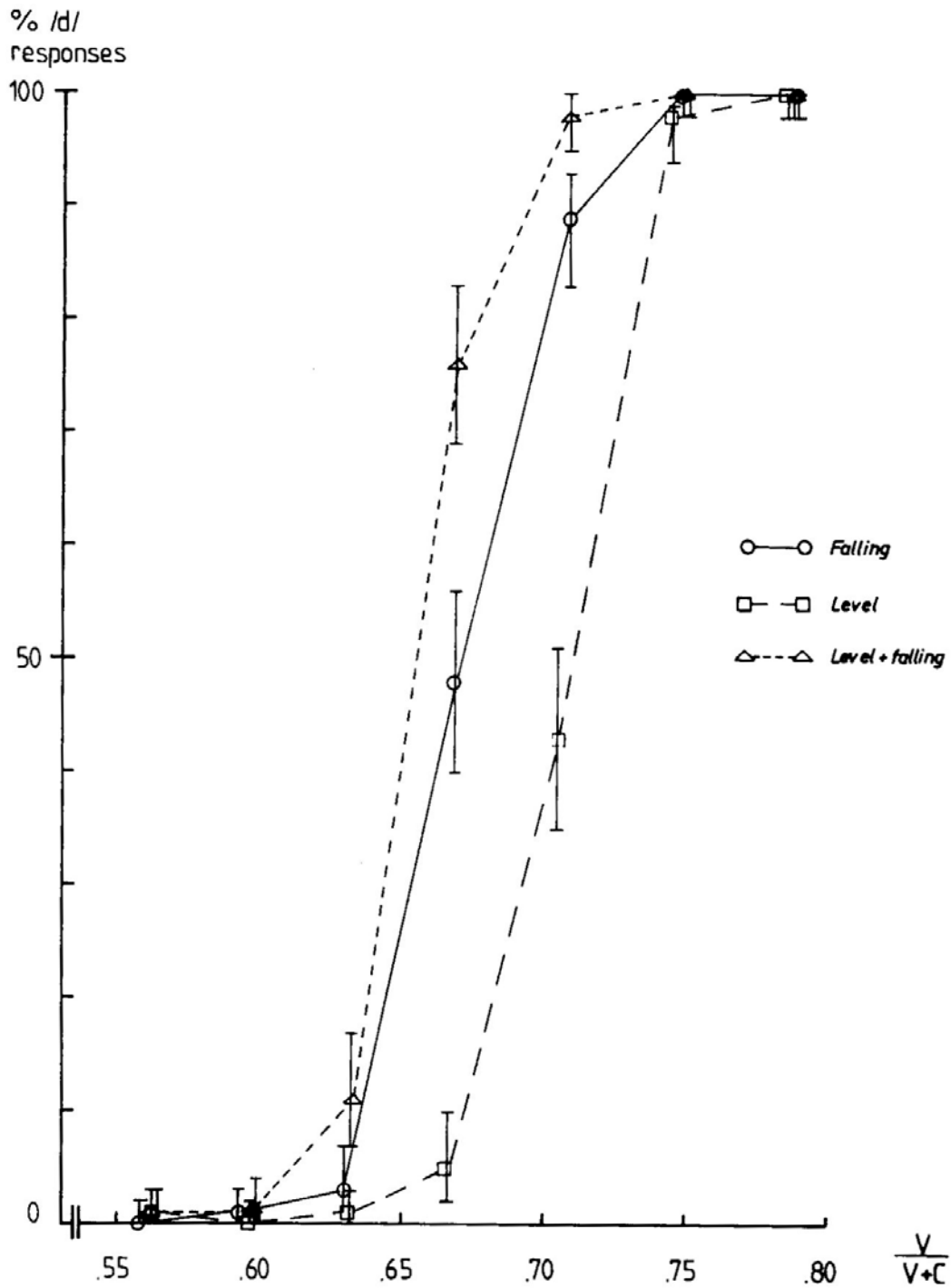


percentage of /d/ responses in the middle of the duration ratio range for both level and continuously falling F0, and the response curves for falling and level + falling F0, which are already close together in the data of the German group, coalesce in this upward shift of two of the identification functions. This means that the English subjects show the same perceptual effects with regard to level F0 as against the other two F0 patterns, but that they nevertheless locate the duration ratio boundary at a lower value than the German listeners. The reason for this difference may be that because English speakers generally devoice the nasal plosion after fortis stops, the absence of this feature in the German test stimuli biases English listeners towards /d/ in the middle of the duration ratio range.

Figs. 9 + 10 about here

### 3.3. Experiment 2.

3.3.1. Procedure. Two English sentences were constructed that replicate the focal and utterance-final position as well as the segmental structure and the phonetic context of the German test words in Experiment 1. The two family names "Lyden" and "Lighton", which are of equal (low) frequency in Britain, were inserted in the sentence frame "I think you'd have to ask ...". They contain the same phoneme sequences as the German words and can also be realised with nasal plosion. They, too, occur after a voiceless consonant cluster that interrupts the F0 glide from



**Fig. 9.** Percentage /d/ responses as a function of vowel/(vowel + closure) duration ratio for the 3 F0 conditions in Experiment 1 ("leiden/leiten", German group), and binomial confidence ranges at the 5% level; 16 listeners. At each data point N = 160.

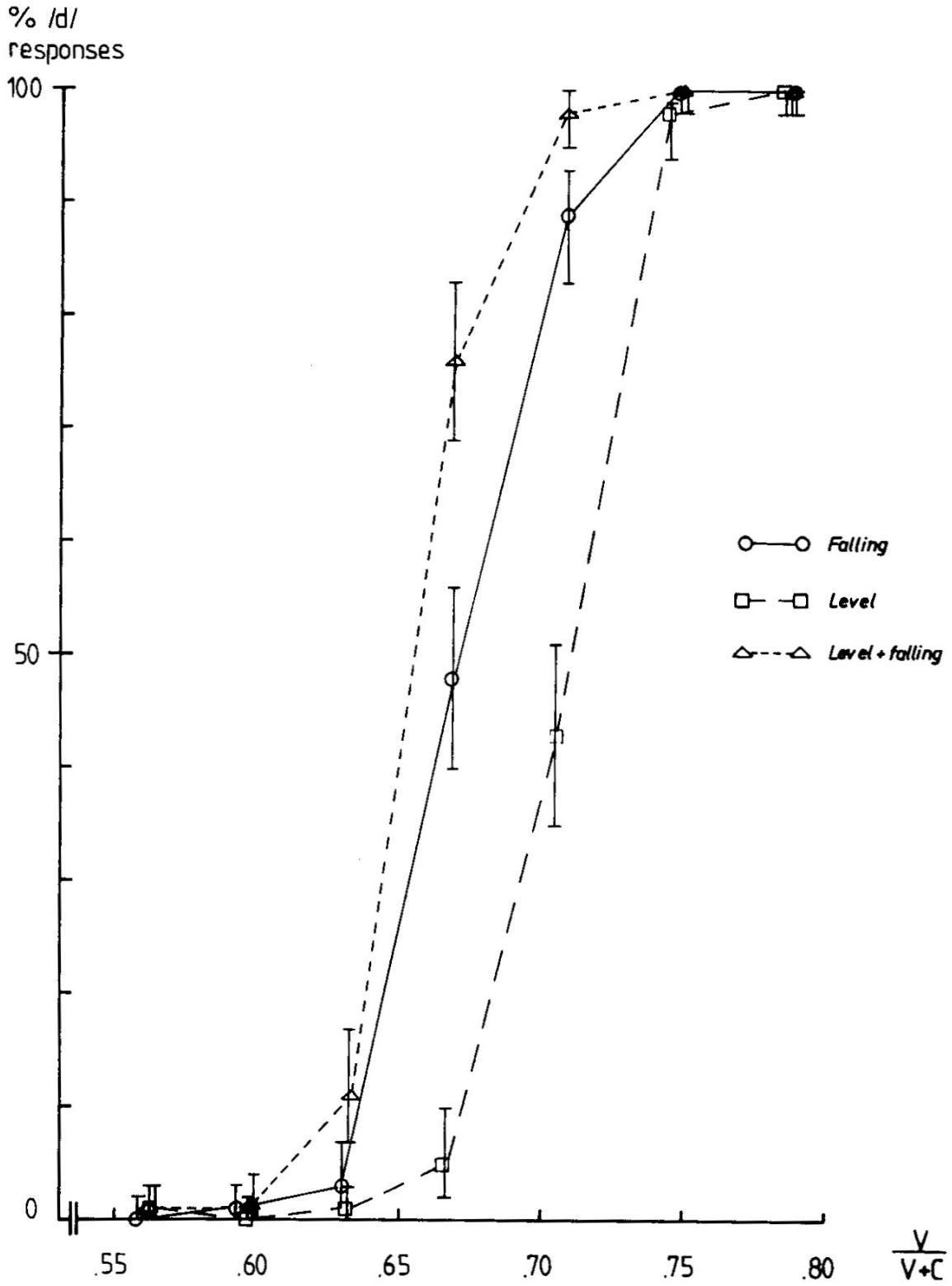


Fig. 10. Responses of the combined British English groups in Experiment 1. At each data point  $N = 130$ .

a low value on "ask" to a high one in the contrastively stressed name so that F0 has practically reached its peak value when it sets in again at voiced /l/ onset.

These sentences were pronounced several times by a native speaker of Southern British, with focus stress on the name, elicited by the context "Who do you think would know about this, Lyden or Lighton?" The F0 contours across the names were very similar to those found in the German sentences of Experiment 1 (cf. Kohler, 1985, p. 24): before the lenis stop F0 drops much further in the stressed vowel than before fortis. One token of a "Lyden" sentence was selected for the test stimulus generation, which followed the principles laid down in Kohler (1985). The stressed vowel measured 289 ms, its closure duration 46 ms and its stop release 24 ms.

Three F0 patterns were generated across the stressed vowel: (a) Level + falling (122-120-75 Hz) with the fall beginning at the vowel centre, (b) level (122-120), (c) linearly falling throughout (122-75 Hz). These F0 contours were combined with 7 rate-manipulated vowel durations, from 260 ms down to 200 ms in 10-ms steps. The closure voicing and release were excised and replaced by silence, which was increased from 70 ms up to 160 ms in 6 equal steps, complementary to the vowel shortening. The 21 vowels produced in this manner, together with the complementary closure pauses, were spliced into the carrier utterance. Thus the durations and F0 patterns of the resulting 21 "Lyden/

"Lighton" stimuli were fully comparable to those generated in the German test, the only difference being that after the silence F0 set in at 70 Hz (instead of 66 Hz) and that the periodicity of the nasal was more regular and of much greater amplitude than in the German "leiden/leiten" stimuli, i.e. there was proper and strong voicing instead of creak.

Since the frame was not synthesized, the stimuli sounded completely natural, and no "synthetic" quality was detectable in the synthesized vowel sections either. The 21 stimuli were copied ten times and randomized to give a test of 210 stimuli, following the same procedure as in the German test. The same group of 13 native British English speakers as in Experiment 1 acted as informants under the same listening conditions in separate sessions. They classified the stimulus utterances as "Lyden" or "Lighton".

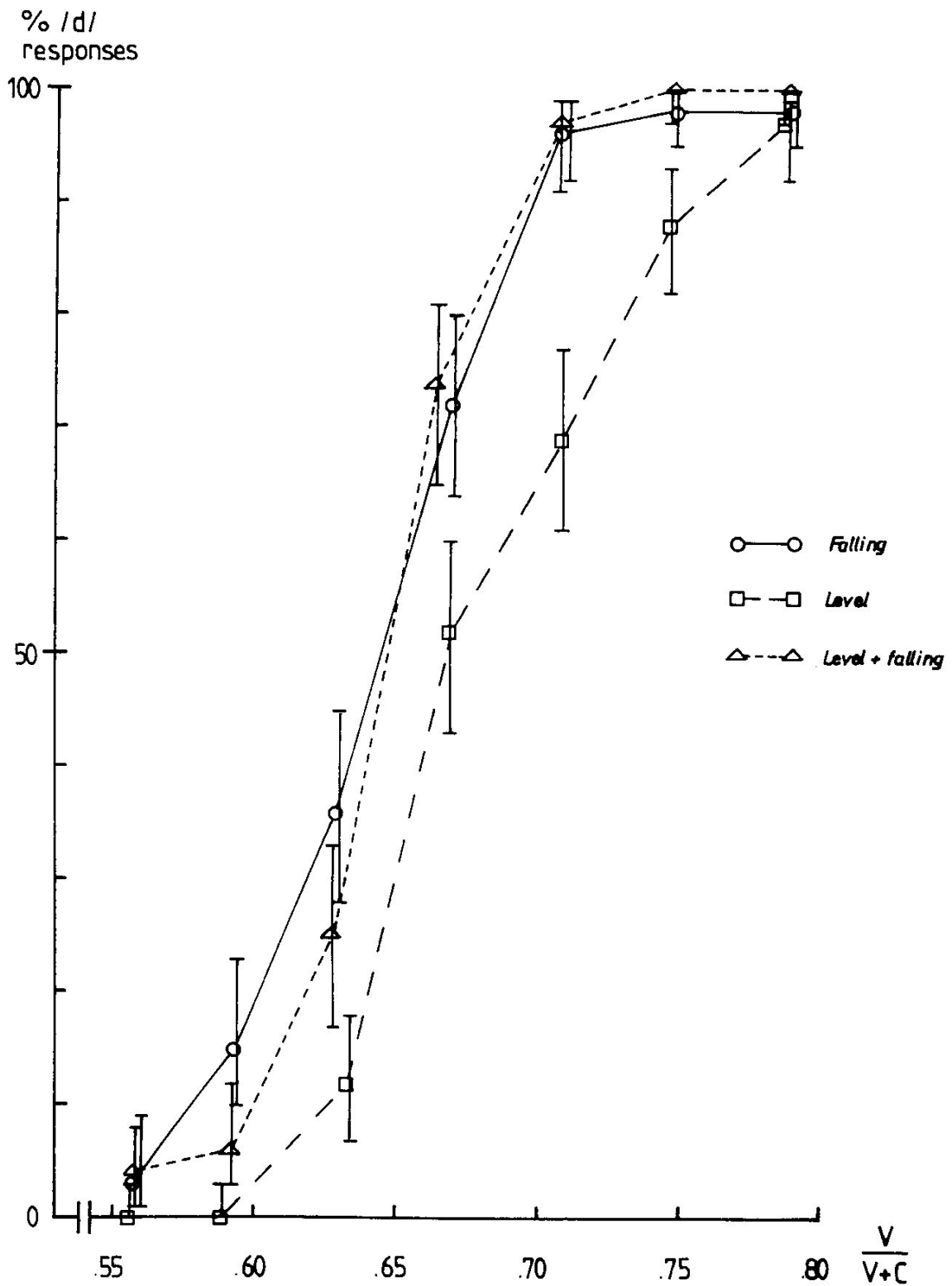
3.3.2. Results and discussion. The results in fig. 11 are basically congruent with the English group results of Experiment 1: the identification curves occupy more or less the same positions along the duration ratio axis, the functions for the two falling F0 sets are again not differentiated from each other, but are clearly separate from the function for level F0, which yields significantly more /t/ responses. The difference between the two experiments lies in somewhat more /d/ judgements in the lower half of the duration ratio scale for Experiment 2. So there must be some essential acoustic difference between the

English "Lyden/Lighton" and the German "leiden/leiten" stimuli. The obvious candidate is the strong voicing instead of creak in the final nasal of the English utterances. It provides a more prominent release cue for /d/, which may enter into conflict with the fortis cues and weaken their effects, i.e., the effect of flat F0 generally and the effect of duration in the lower range. It has also been shown by Kohler and van Dommelen (1987) that different voice qualities affect the perception of lenis and fortis consonants.

Fig. 11 about here

### 3.4. Experiment 3.

3.4.1. Procedure. The sentences "I am telling you I said widen/whiten." were pronounced several times with focus stress on the final word and with nasal plosion by the same native Southern British speaker that produced the utterances for Experiment 2. One "widen" token was selected for constructing 21 test stimuli according to the same principles as in Experiments 1 and 2. The vowel durations ranged from 265 ms to 205 ms, the silence durations from 70 to 160 ms. Again 3 F0 patterns were generated with each vowel duration. In the level + falling F0 pattern the level section was represented by the naturally produced fluctuation between 119 and 123 Hz over the first 100 ms of the original vowel, followed by a linear fall to 85 Hz, the proportion of level and slope sections staying the same in all 7 stimuli. The first 100 ms of the level F0 were identical



**Fig. 11.** Responses of the combined British English groups in Experiment 2 ("Lyden/ Lighton"). At each data point  $N = 130$ .

with the level section of the level + falling pattern in the longest vowel and changed proportionally with the vowel duration; the remainder descended to 122 Hz. In the third pattern, F0 fell linearly throughout from 119 to 85 Hz.

The original /d/ release was again eliminated, and the 21 synthesized vowels + closure pauses were spliced into the sentence frame. F0 at voice onset of the final nasal was 89 Hz, descending to 69 Hz. The very large amplitude of the regular periodicity in /n/ was adjusted to the one found in "Lyden" by applying the reduction factor .35. The durations and the F0 patterns were comparable to the ones in the test stimuli of Experiments 1 and 2, but with important differences in the height of the pre- and postconsonantal F0 ending and starting points.

The test tape construction and the running of the experiment followed the same lines as in Experiment 2. A previous run of the test was reported in Kohler (1986a). It was repeated here by the same two British English groups as in Experiments 1 and 2. In a pretest, each of the 13 subjects was examined as to whether they distinguished "wh" from "w". Two informants did and were, therefore, excluded from the test because their expectations for "whiten" would have been different.

3.4.2. Results and discussion. Fig. 12 provides the data for the combined group. There are no inter-group divergencies: The



differences between the three F0 patterns have practically disappeared. The effect of flat F0, which was still slightly present in the previous run of the same test, has been levelled out. Otherwise the two test runs provide corresponding locations of the identification functions. Since it is only the response curve for flat F0 that is positioned differently in the "Lyden/Lighton" and the "widen/whiten" data, the initial consonant /w/ cannot be responsible for the increase of /d/ judgements. It must be an acoustic feature difference that is peculiar to the flat F0 stimuli. In "Lyden/Lighton", F0 is flat across the stressed syllable, and a rise from the preceding syllable is masked by voicelessness; after the closure silence, F0 resumes at its low utterance-final value. The flat F0 contour is thus bounded by voiceless stretches on both sides, with low F0 preceding and following. In this environment, the high flat F0, i.e. the fortis cue, becomes perceptually salient. In "widen/whiten", on the other hand, there is an upward F0 glide from the low value of the preceding syllable right into the stressed vowel, and it is only the final 130 - 160 ms that are actually flat. After the closure pause, there is a substantial F0 fall of 20 Hz. In this context, the high flat F0 is integrated into a macroprosodic rise-fall pattern and is, therefore, perceptually far less salient, thus losing its fortis cue strength.

Fig. 12 about here

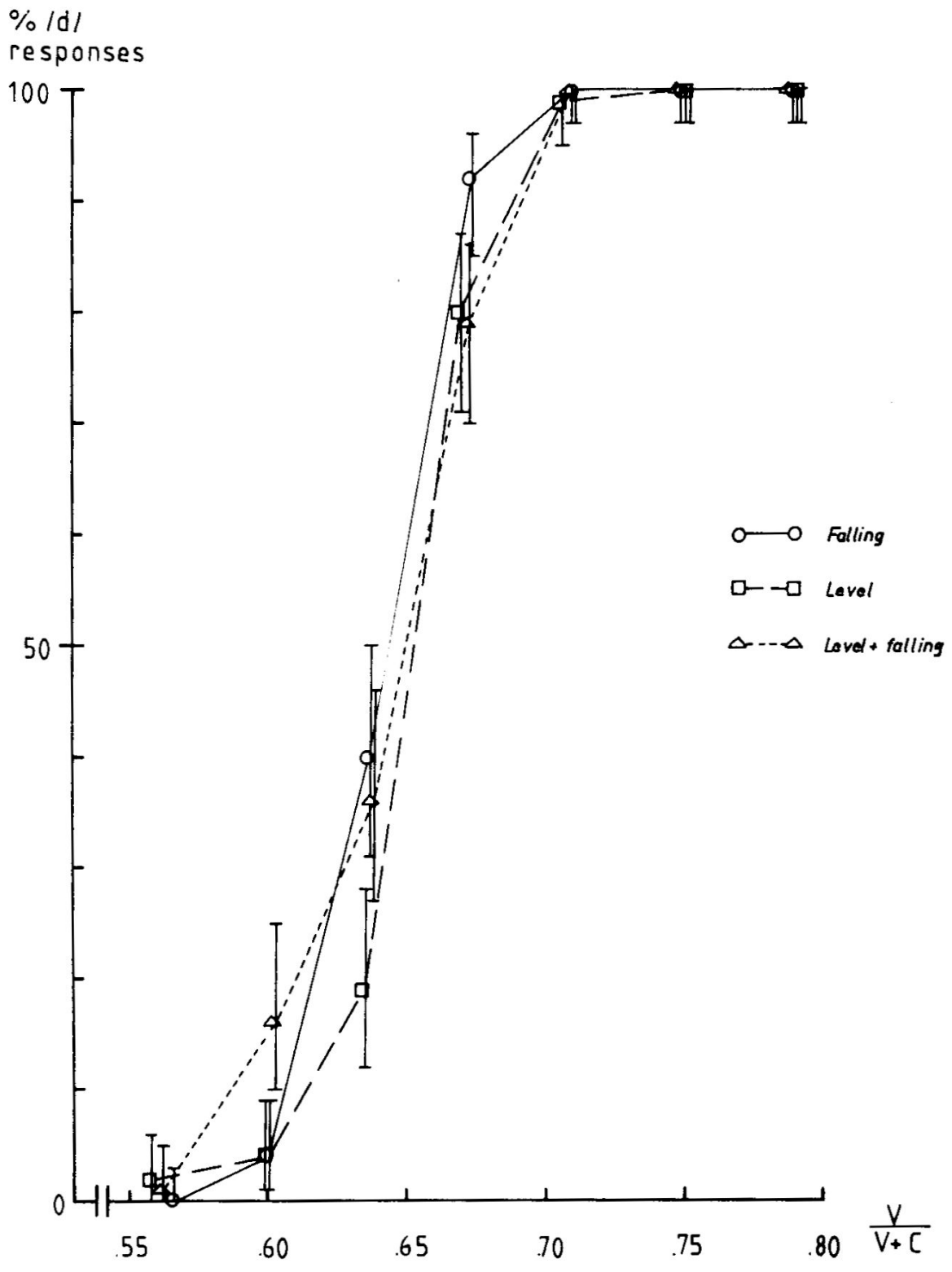


Fig. 12. Responses of the combined British English groups in Experiment 3 ("widen/whiten"). At each data point N = 110.

3.5. General Discussion.

The results of the 3 experiments point to the following prosodic influences on lenis/fortis stop perception in German and English.

1. A flat F0 across a stressed prestop vowel in a focused utterance-final disyllable is a fortis cue, compared with falling F0 patterns, in both German and English, as long as the flat F0 is clearly detachable from a macroprosodic utterance intonation as a microprosodic manifestation. In German, a flat + falling F0 is also differentiated from a continuously falling F0 as a stronger lenis cue.
2. In English, the category boundary between lenis and fortis is located at lower duration ratios. This leads to a coalescence of the identification functions for flat + falling and continuously falling.
3. A stop release with regular voicing of high amplitude and an F0 fall (below the focus peak) weakens the preconsonantal microprosodic fortis cue.
4. The microprosodic effects of prestop flat and flat + falling F0 are obliterated when they are integrated into macroprosodic utterance pitch patterns.

## 4. CONCLUSIONS

As regards the global utterance F0 of such languages as German and English, it is necessary to distinguish between an intonation and a stress function at this macro F0 level. In German, the shift of an F0 peak contour, over a total time stretch of 300-400 ms, from the centre of a syllable nucleus to its right-hand boundary or to a syllable preceding it, which is unequivocally signalled as unstressed by its segmental qualities and quantities, does not change the stress position but cues different intonations related to the semantic features 'established' vs. 'new' vs. 'emphatic'. On the other hand, an F0 peak shift by itself is also sufficient to bring about a clear change from one stress position to another, provided the duration of the stressed-syllable-to-be is not too short. The shape of the F0 peak, in addition to its temporal location in the syllable structure, is a further cue for the signalling of the stressed syllable. This stress function of F0 shows an interference from its intonation function (i.e. the early, central or late peak in relation to the same stress position) if it is not supported by duration. Similar findings regarding the stress and intonation functions of F0, as established for German, are to be expected for English and other languages that differentiate stressed from unstressed syllables and associate meaning-related utterance intonations with stressed syllables. I would even hypothesize that the use of early vs. later F0 peaks to signal the 'established' vs. 'new' dichotomy is quite

wide-spread. If, however, the contrast of temporal peak alignment has already been assigned a function at some other linguistic level, as in the Scandinavian word accents (cf. Gårding, 1982), its intonational use for the utterance-semantic categories is either precluded or has to undergo some language-specific modification. It is an interesting empirical question to study the phonetic realization of the intonational categories described for German in a language such as Swedish.

Beside these language-specific macroprosodies of stress and intonation, there are the universal microprosodic F0 adjustments due to the articulatory constraints at each point in the utterance. Thus, central peak contours in stressed vowels of German and English show perceptually relevant microprosodic influences from following as well as preceding fortis or lenis consonants, respectively. In both cases, F0 is raised by fortis consonants. But the perceptual effects of a prestop F0 raising only shows up when the pattern is clearly detachable from a macroprosodic utterance intonation as a microprosodic manifestation. Although microprosodic influences arise both from prevocalic and postvocalic consonants and can become perceptual cues to segmental identity their effects are more easily overridden by the utterance prosody in the preconsonantal position. When the segmental cues to the lenis/fortis distinction (voicing, aspiration, duration) disappear through sound change the microprosodic differences may be preserved or even heightened and thus lead to tonal distinctions. But as the

preconsonantal F0 effect is better controlled by the global F0 than the postconsonantal one, such tonogenesis is hardly attested in this context.

REFERENCES

- E. Gårding, "Swedish prosody. Summary of a project", *Phonetica* 39, pp. 288-301, 1982.
- J.M. Hombert, J.J. Ohala, and W.G. Ewan, "Phonetic explanations for the development of tones", *Language* 55, pp. 37-58, 1979.
- K.J. Kohler, "F0 in the perception of lenis and fortis plosives", *J. Acoust. Soc. Am.* 78, pp. 21-32, 1985.
- K.J. Kohler, "Preplosive F0 in the perception of /d/-/t/ in English", *Proc. Montreal Symposium on Speech Recognition*, pp. 34-35, 1986a.
- K.J. Kohler, "Computer synthesis of intonation", *Proc. 12th Intern. Congr. Acoustics*, A6-6, Toronto, 1986b.
- K.J. Kohler, "Categorical pitch perception", *Proc. 11th Intern. Congr. Phon. Sc.*, Vol. 5, pp. 331-333, Tallinn, 1987.
- K.J. Kohler and W.A. van Dommelen, "The effects of voice quality on the perception of lenis/fortis stops", *Journal of Phonetics* 15, pp. 365-381, 1987.
- K. Silverman, "The structure and processing of fundamental frequency contours", Ph.D. thesis, Cambridge University, 1987.