MICROPROSODY IN SEGMENT PERCEPTION

KLAUS J. KOHLER

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

ABSTRACT

For German it has been demonstrated in a number of experiments that in production as well as in perception a level and a level + falling F0 contour on a prestop vowel are cues for fortis and lenis stop, respectively. This paper reports on perception experiments that replicate the German findings for English, and relates the results to an interaction of three factors: (a) prestop microprosody, (b) poststop microporosody, (c) global utterance macroprosody.

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 /1/. 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 /2/. 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 /3/. 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/.

EXPERIMENT 1

Procedure.

The test tape of experiment 2 of /2/ was presented to a 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. Two groups of 6 and 7 British English speakers performed the same test under the same conditions, 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.

Results.

The German group replicates the results of the previous test (cf. /2/, pp. 24ff) in every respect (see figure 1). The two English groups, which do not differ from each other and are, therefore, combined in the data presentation of figure 2, also show clearly separate identification functions for level and falling F0. But they have 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 fo 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.

EXPERIMENT 2

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 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. /2/, 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 /2/. 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 center, (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 two groups of 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".

Results and discussion..

The two groups differ in their responses to the level F0 stimuli, one giving more /d/ judgements. Figure 3 presents the combined group results. They 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 differences between the two experiments are (a) somewhat more /d/ judgements in the lower half of the duration ratio scale for Experiment 2, and (b) different as against identical behaviour of the two groups in the two experiments. 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. This conflict can be solved differently, according to whether the release is weighted more highly, especially than flat F0. The two groups differ in this respect.

EXPERIMENT 3

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,

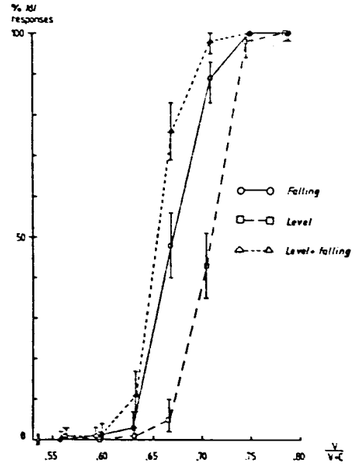

Fig. 1. Percentage /d/ responses as a function of vowel/(vowel + closure) duration ratio for the 3 F0 conditions in Experiment 1 ("leiden/leiten", German group), and binomial confidence ranges at the 5 % level; 16 listeners. At each data point N = 160.

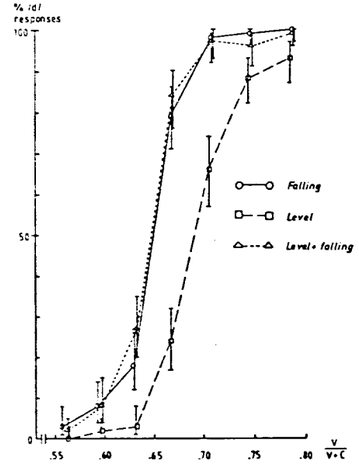

Fig. 2. Responses of the combined British English groups in Experiment 1. At each data point N = 130.

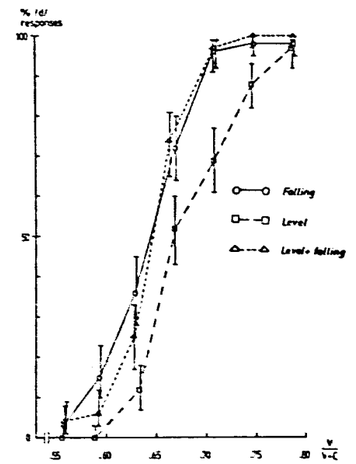

Fig.3. Responses of the combined British English groups in Experiment 2 ("Lyden/Lighton"). At each data point N = 130.

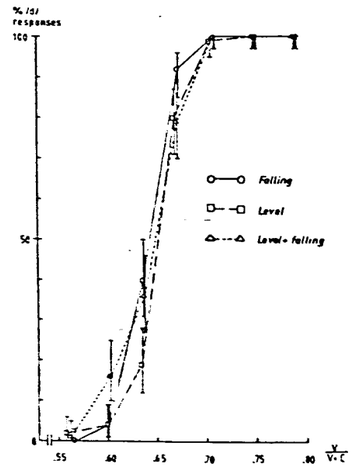

Fig. 4. Responses of the combined British English groups in Experiment 3 ("widen/whiten"). At each data point N = 110.

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 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 /3/. 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.

Results and discussion.

Figure 4 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.

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.

5. The interaction of pre- and poststop microprosody and of global utterance macroprosody explains why a prestop F0 influence on lenis/fortis perception can only arise under special circumstances and, therefore, not provide a basis for tonogenesis (cf. /1/).

REFERENCES

/1/ J.M. Hombert, J.J. Ohala, and W.G. Ewan, "Phonetic explanations for the development of tones", Language 55, pp. 37-58, 1979.

/2/ K.J. Kohler, "F0 in the perception of lenis and fortis plosives", J. Acoust. Soc. Am. 78, pp. 21-32, 1985.

/3/ K.J. Kohler, "Preplosive F0 in the perception of /d/-/t/ in English", Proc. Montreal Symposium on Speech Recognition, pp. 34-35, 1986.