LAST COPY
XEROX

746

Phonetica
Editor: K. Kohler, Kiel

Separatum
Publisher: S. Karger AG, Basel
Printed in Switzerland

Phonetica 1990;47:36–49

Gradient Effects of Fundamental Frequency on Stop Consonant Voicing Judgments

D. H. Whalen[a], *Arthur S. Abramson*[a,b], *Leigh Lisker*[a,c], *Maria Mody*[a,d]

[a]Haskins Laboratories, New Haven, Conn.; [b]University of Connecticut, Storrs, Conn.; [c]University of Pennsylvania, Philadelphia, Pa.; [d]City University of New York, N.Y., USA

Abstract. The post-stop-release rise or fall of fundamental frequency (F_0) is known to affect voicing judgments of syllables with ambiguous voice onset times (VOTs). In 1986, Silverman claimed that the critical factor was not direction of F_0 change but rather its direction relative to the intonational contour. He further claimed that only F_0s that start above and fall to the contour have an effect proportional to the size of the frequency change; F_0s that rise to the contour by different amounts were claimed to be equivalent. In our first experiment, we examined the effect on voicing judgments of five onset F_0s preceding a single, flat contour. Only falling F_0s were differentiated in the first set of judgments, but after increased exposure to the syllables, even F_0s below the contour differentially affected the voicing judgment. In a second experiment, the contour of the final part of the syllable was flat, rising or falling. F_0 contour affected the judgments, as did onset F_0s, but the two factors did not interact, indicating that the onset values were not being judged by reference to the contours. However, the contour which was predicted to result in more voiceless judgments also ended at a higher F_0 in the vowel, and another effect of voicing is that the F_0 is higher throughout the vowel after voiceless stops. In a third experiment, F_0 contours were created to contrast contour and mean F_0. The effect of the F_0 during the vocalic segment appeared to be attributable to the average F_0 rather than the contour. In all three experiments, the F_0 onset values contributed to the voicing judgment whether they were above or below the putative intonation contour. The contribution of the lower F_0s, while significant, was not as great as that of the higher F_0s, which argues for a noncategorical contribution of intonation.

Introduction

After voiceless stops in a variety of languages, fundamental frequency (F_0) has been found to fall, while after voiced stops, it has often been found to be level or slightly rising [House and Fairbanks, 1953; Lehiste and Peterson, 1961; Hombert, 1975; Löfqvist, 1982; Hombert et al., 1979]. Other studies find that F_0 after voiced stops falls, but less steeply than after the voiceless stops [Haggard et al., 1981: Ohde, 1984]. So while voiceless stops clearly raise F_0, the role of the voiced stops is less clear. Since the size of this change in F_0 quickly diminishes, it has been called a perturbation of the F_0 by the stop voicing. In part to resolve the conflict among studies of the F_0 perturbation after voiced stops, Silverman [1986] examined the intonational F_0 as well as the F_0 at stop release. He noted that the inconsistencies in the literature about whether F_0 rose or not after voiced stops could be accounted for if the F_0 at stop release was determined relative to the intonation pattern rather than by absolute measurements. In the earlier studies, if the intonational pattern itself was rising, a voiced stop which did not affect the F_0 would be measured as rising, since, physically, F_0 is rising after stop release. Silverman calls this lack of a perturbation due to voiced stops the 'no-rise' hypothesis, in contrast to the 'rise-fall dichotomy' account. His own studies of produced perturbations [Silverman, 1987] did in fact show a fall even after voiced stops and were also well described as displaying the unperturbed intonation contours. The discrepancies in the earlier findings were therefore attributed to a lack of control of the intonational contour across different studies.

Perceptually, this F_0 perturbation can cue the voicing distinction [Haggard et al., 1970; Fujimura, 1971; Haggard et al., 1981; Abramson and Lisker, 1985]. While Haggard et al. found that large F_0 shifts affect the categorization of all values of voice onset time (VOT), Fujimura, and Abramson and Lisker, using a more natural range of F_0s, found effects only with ambiguous VOTs. Silverman [1986] studied the effect of 3 post-stop-release F_0 contours (falling, flat, rising) on the voicing distinction as cued by closure duration. When the F_0 after the stop-related contour was level, the 2 onset contours associated with the voiced stop (flat and rising) were not distinguished, while the falling contour gave more voiceless judgments. However, if the contour in the syllable was changed to a rising one, the flat onset was then appropriate for a voiceless stop. In this case, the falling contour again resulted in more voiceless judgments than the flat one, but the flat in turn had more voiceless judgments than the rising. Thus Silverman concluded that a 'no-rise' explanation is better able to account for the results than a 'rise-fall' dichotomy. While this set of results applies quite nicely to the previous production studies, it leaves several unexplored areas, which we will address in this paper.

First, only 3 onset values of F_0 were directly compared in Silverman [1986], which allows only a minimal test of the 'no-rise' hypothesis. His assertion [p. 87] that '[r]ises from below the intonation, however, do not decrease the probability that the stop was voiceless. They are simply ignored' rests on only one critical case, hardly enough for such a strong conclusion. Partial confirmation appears in Abramson and Lisker [1985], where the 3 F_0 onsets at or below the steady-state F_0 give indistinguishable identification functions, but the 2 higher onset values give more 'p' responses depending on their value. The first experiment described here explores this in more detail, using the F_0 values of Abramson and Lisker [1985] but more VOT values. As it turns out, the simple change in procedure of requiring more judgments per stimulus has an effect.

Second, it is of interest to see whether the F_0 perturbations are related to the intonation contour only when the closure duration is the acoustic aspect which is manipulated, or whether VOT boundaries are also affected relative to the intonation. Silverman [1986, p. 88] implies that he has tested the interaction of F_0 and VOT, but this is problematic. Indeed, he goes on to say that the fact the endpoints of his continua were affected by F_0 contradicts assertions by Abramson and Lisker [1985] to the effect that F_0 can affect only ambiguous values of VOT. However, we see no contradiction between the two sets of results. If we interpret VOT in its common usage as the time between release of the stop and the onset of voicing in the syllable of concern, we understand that in initial position the turbulence of the aspirated release is itself a very powerful acoustic cue, one not manipulated in Silverman's stimulus array. Of course, his series of synthetic closure durations can be seen as an increasingly long set of breaks in voicing and thus a manifestation of voice timing. Although voice *onset* time should be seen as simply one of a number of forms of voice timing [Abramson, 1977], Silverman's voicing breaks in fact cooccur with the stop closure. The two kinds of timing should not necessarily be expected to give the same results.

Abramson and Lisker [1985] found it convenient to test VOT in utterance-initial position, but the intonation pattern of such monosyllables may be difficult to assess. Yet, even if Silverman is correct in saying that the F_0 perturbation due to stop release can only be perceptually interpreted in relation to the intonation contour, his complaint that the use of isolated monosyllables in Abramson and Lisker did not 'control what underlying intonation contour their listeners may have perceived or imagined' [Silverman, 1986, p. 88] loses force. While the intonation was not manipulated in that study, if the perturbation is nonetheless referenced to the contour, then the listeners *must* have made use of that contour. Similar objections could be raised for Silverman's study itself. Although the F_0 values used in Silverman's studies clearly had patterns, he provides no evidence that those were the intonation contours that were actually perceived by his subjects. This is particularly problematic in his first experiment, in which the F_0 pattern was not modeled on English intonation but was flat across three syllables. The second experiment reported here begins to address this issue by using an isolated monosyllable with a very schematic intonation contour, namely, one that either rises, falls, or remains flat from the point at which the F_0 perturbation ends. If such a manipulation does not have an effect, it may only indicate that isolated monosyllables do not have a sufficient intonation contour for segmental effects to appear. If, on the other hand, a comprehensible pattern emerges even in this case, then we can conclude that the intonation contour does not, in all utterances, require more than one syllable's worth of support.

Finally, although Silverman's [1986] stimuli clearly had an intonational pattern which represented a possible utterance, the rising intonation also had an effect on the mean F_0 of the syllable containing the stop. Voiceless stops are often followed by higher F_0 even in the middle of the vowel [Ohde, 1984; Whalen, 1990; see Löfqvist et al., 1989, for an instance where no difference appears]. The rising intonation supported voicelessness in the stop, but so did the mean F_0. Therefore Silverman's results are ambiguous with respect to which factor was responsible. Our third experiment will take a step towards resolving this ambiguity.

Experiment 1

The first experiment explored a 5-member F_0 onset continuum crossed with a 7-member VOT continuum. The subjects made identifications of an initial stop as 'b' or 'p'. After one unspeeded identification task came a speeded identification task (which addressed a separate question and will not be reported here). Finally there followed an unspeeded identification task essentially identical to the first.

Method

Stimuli

The syllables used were modeled after those of Abramson and Lisker [1985]. They were created with the Haskins software serial synthesizer designed by Ignatius G. Mattingly. The formant values gave rise to a /ba/ or a /pa/ percept for English speakers. The vowel steady-state formants were centered at 730, 1,250, and 2,440 Hz with bandwidths of 100, 100, and 125 Hz. The formant values at the beginning of the syllable were 450, 1,080, and 2,300 Hz for F_1, F_2 and F_3 respectively, and changed linearly to reach the steady-state values after 75 ms. The steady-state F_0 was 114 Hz. Vowel amplitude was level until the last 30 ms of the syllable, at which point it decreased linearly to 0. Total duration of the syllables was 250 ms.

The VOT values were 5, 10, 15, 20, 25, 35, and 50 ms. These were obtained by turning off the voicing component (AV) and introducing aperiodic hiss (AH) for the appropriate number of synthesis frames.

The onset F_0 values were 98, 108, 114, 120, and 130 Hz. The onset value was used at the onset of voicing, and was linearly interpolated to reach the steady-state value of 114 Hz 100 ms later. For a 5-ms VOT, then, the end of the F_0 contour occurred 105 ms into the syllable, while for the 50-ms VOT, it occurred 150 ms into the syllable.

Each VOT was paired with each F_0, giving 35 unique stimuli.

Procedure

The first and second conditions of the experiment consisted of 5 repetitions of the 35 stimuli randomized and recorded on audio tape. There was an inter-stimulus interval of 2.5 s, with a longer pause (6 s) after every 10 trials [corresponding to a line on the answer sheet]. Subjects heard the stimuli over headphones (TDH-39) in a quiet room and indicated which stop they heard by writing 'b' or 'p' on the answer sheet provided. The first condition was followed by a speeded response task consisting of 15 repetitions of the same stimuli (for a total of 525 judgments). This set of results will not be reported here. The final part of the session was the second condition which was, like the first condition, an unspeeded identification of a new randomization of 5 repetitions of the 35 stimuli.

Subjects

Twelve Yale University undergraduates and graduate students who were native speakers of American English with no reported hearing problems were paid to participate.

Results and Discussion

The percentage of 'p' responses averaged over subjects is presented in figure 1, with Condition 1 (the first presentation) in the upper panel und Condition 2 (the second presentation) in the lower. The upper panel replicates the results of Abramson and Lisker [1985] quite well, with the ambiguous VOTs being most affected by the F_0. Also, there was a straightforward effect of the high onset F_0s and a clustering of the three lower onset values. By the time the second condition was run, however, the pattern was even better articulated, with all five values being separated.

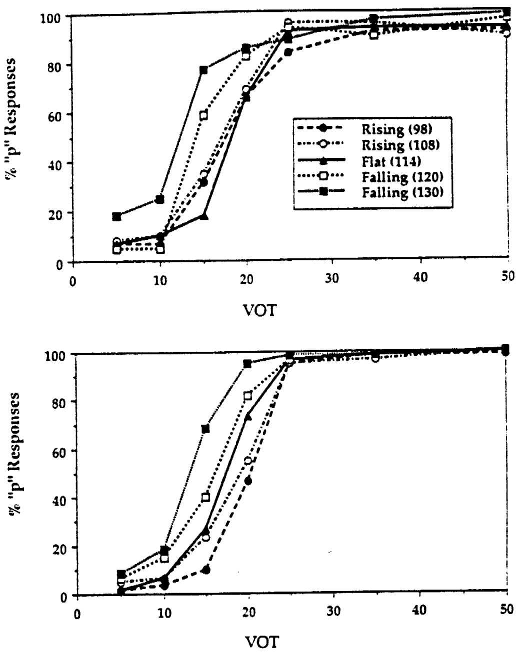

Fig. 1. Identification functions for Experiment 1, collapsed across 12 subjects. The upper panel represents the 5 judgments per stimulus in the first condition, and the lower panel the 5 judgments in the second.

For the statistical analysis, it was found that fitted sigmoids (PROBIT) were overly sensitive to noise at the endpoints, so the total number of 'p' responses was used instead. This has the benefit that noise at the 'b' end offsets noise at the 'p' end, since the first gives an extra 'p' and latter loses a 'p'. An analysis of variance was conducted with the factors Condition (1 and 2) and F_0 Onset (5 levels). Condition had no effect ($F(1,11) < 1.0$), while F_0 Onset was highly significant ($F(4,44) = 14.11$, $p < 0.001$). The two factors did not interact ($F(4,44) = 1.74$, n.s.). However, the two predictions we were testing relied primarily on those stimuli which had flat or rising F_0 contours at the beginning of the vocalic segment. The 'no-rise' view predicts that those will be equivalent, while our own expectation was that lower F_0s would give more voiced judgments even if they are below the intonation contour. A separate analysis with 3 levels of F_0 Onset showed no effect of Condition ($F(1,11) < 1.0$), a marginal effect of F_0 Onset ($F(2,22) = 2.67$, $p < 0.10$), and a marginal interaction ($F(2,22) = 2.50$, $p < 0.15$). As an exploration of this marginal effect, a final analysis was done with just the 3 lower onsets of the second condition. This resulted in a significant effect of F_0 Onset ($F(2,22) = 4.34$, $p < 0.05$). The comparisons between the 2 adjacent pairs of F_0 onset values were not significant by a Newman-Keuls test, indicating that the values of 98 and 114 were distinct.

As in Abramson and Lisker [1985], there was little effect of F_0 onset on voicing judgments for unambiguous VOTs. It is of interest that while the sensitivity to F_0 in the ambiguous region increases across conditions, it decreases in the unambiguous region.

In the course of an hour's experiment, subjects became increasingly sensitive to the nuances of the synthetic stimuli used. The 2 F_0 onset values above the level F_0 of the vocalic segment, which is the most likely intonation for listeners to perceive, resulted in more voiceless judgments. The effect was gradient in that it was more pronounced for the higher of the two. For the 2 onsets below the putative intonation, the effects were less clear cut. In the first block, they were not distinguished at all, while there is a weak statistical implication that they were distinct in the second. This difference might be due to the pressure of responding to the stimuli in the (unreported) reaction time condition, but it could also be due simply to hearing them more frequently. Silverman [1986] had only 2 repetitions of his stimuli, and Abramson and Lisker [1985] had 5 (as in the first condition here), while the total in this experiment was 10. Subjects therefore may be increasingly sensitive to the differences among the lower onsets as they listen to the stimuli involved. However, it is also clear that the effect of the lower onsets is much less pronounced than that of the higher onsets. From a psychoacoustic viewpoint, this is unusual, since a change of a particular magnitude will be better discriminated at lower frequencies than at higher [e.g., Moore, 1973]. These considerations may indicate that Silverman's [1986] account is partly correct, that is, although there is a difference between the different rising patterns, this difference is not as great as it is for the falling patterns. It will be further explored in the second experiment.

Experiment 2

The second experiment had two aims. First, we wanted to test in more detail the notion that the relation of the F_0 onset perturbation to the intonation was the critical factor. We thus changed the F_0 pattern of the vowel's steady-state to create different intended intonations. Although these were necessarily brief and not modeled on patterns bearing intonational meaning, they could nonetheless provide a direct test of Silverman's intonational hypothesis, if a difference in voicing judgments emerges between them. (If intonation type did not affect voicing judgments at all, we would not have an appropriate test.) Second, we wanted to determine whether the apparent difference between the two blocks in the first experiment was due to simple exposure to the stimuli or depended critically on the (unreported) speeded identification task that came between these blocks. Thus in the second experiment, the subjects' increased exposure to the stimuli came from sheer repetition of similar stimuli for one kind of judgment, namely, unspeeded judgments.

Method

Stimuli

A subset of the synthetic stimuli of Experiment 1 was used once again, along with some new variants. Only 5 of the original VOT values were included (5, 10, 15, 20, 35 ms) to reduce the demands on the subjects. In addition to the original, flat F_0 offset, another pattern ended at a high value and another one ended at a low value (fig. 2). The contour for the high offset began at 114 Hz at the point where the onset contour ended (or would have ended, in the case of the 114-Hz onset), and then changed linearly to a final value of 138 Hz. The low offset began at the same point but fell linearly to 90 Hz. The presumed underlying intonation patterns were linear extensions back in time of the rise or fall. The combination of 5 VOT values, 5 onset F_0 values, and 3 offset F_0 values yielded 75 unique stimuli.

Procedure

The first condition consisted of 5 repetitions of each stimulus randomized and then recorded on audio tape as in the first experiment. Subjects heard the stimuli on the equipment used in Experiment 1 and wrote their responses as 'b' or 'p'. The second condition was simply another randomization of 5 repetitions, and was given to the subjects after a brief break.

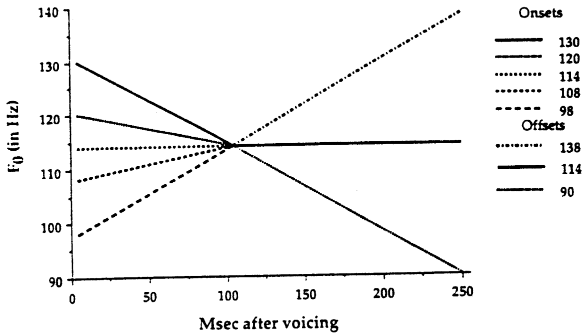

Fig. 2. F_0 patterns for Experiment 2, showing the 5 onsets and the 3 offsets. The timing shown is for the 5-ms VOT. For larger VOTs, the inflection point in the pattern is correspondingly later.

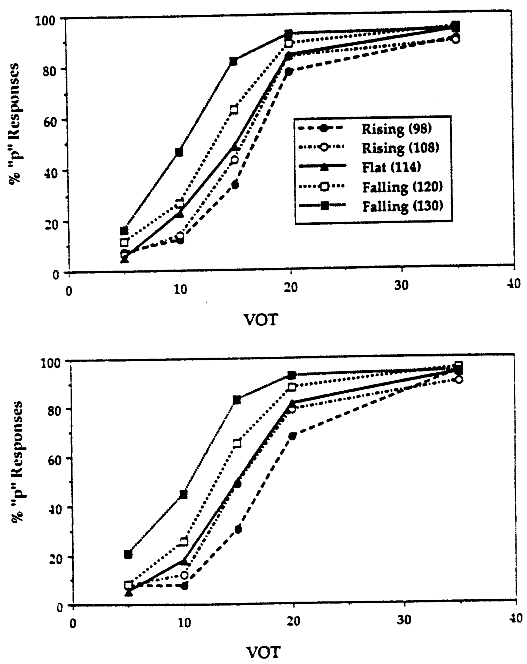

Fig. 3. Identification functions for Experiment 2, collapsed across 12 subjects and the 3 F_0 offsets. The upper panel represents 15 judgments per F_0/VOT combination in the first condition, and the lower panel those judgments in the second condition.

Subjects

Twelve other Yale University students with no reported hearing difficulties were paid to participate.

Results and Discussion

Figure 3 shows the identification functions for the 5 onset F_0s collapsed across offset F_0, with the first condition in the top panel and the second condition in the lower. It is clear that in both conditions the 5 onset F_0s were distinguished in the expected way. Statistical analysis of the total number of 'p' judgments with the factors F_0 Onset, F_0 Offset, and condition (first or second presentation) confirmed that F_0 Onset was a significant factor ($F(4,44) = 35.77$, $p < 0.001$), with more 'p' responses for high Onsets. F_0 Offset was also a significant factor ($F(2,22) = 4.71$, $p < 0.05$), more 'p' responses for high offset. There was no difference between conditions ($F(1,11) < 1$, n.s.). None of the interactions reached significance. This is particularly damaging for the 'no-rise' account, since all the values for the low offset should have been the same (assuming the underlying intonation pattern was continuously falling), while all the onsets should have been distinct for the high offset (assuming the underlying intonation pattern was continuously rising). The lack of an interaction shows that this prediction is not supported.

To test the results in a fashion similar to that of the first experiment, we analyzed only the lower 3 onset F_0 values separately. In this experiment, this analysis mirrored the full analysis completely. F_0 Onset was significant ($F(2,22) = 11.96$, $p < 0.001$). F_0 Offset was also significant ($F(2,22) = 4.74$, $p < 0.05$). There was no difference between conditions ($F(1,11) < 1$, nonsignificant). None of the interactions reached significance.

Table 1 shows the responses for each of the 5 onset F_0s comparing the 3 offset F_0s. For each of these, the prediction based on Silverman's [1986] account is that high offsets should have more 'p' responses and low offsets should have fewer. This turns out to be correct, as seen in the main effect of F_0 offset reported above. However, the interaction of F_0 onset and offset was not significant. Such a lack of an interaction is incompatible with the 'no-rise' account. That is, if the cue value of a particular onset changes when the slope of the intonation changes, there should be a difference across the 3 offset values used here. There was not, indicating that the influence of onset F_0 was gradient and independent of the intonation to which it was referenced. A possible alternative explanation is that the subjects were perceiving an intonation other than the one intended. Our own listening to these stimuli gave us no impression of complex intonation patterns. In the absence of other evidence to the contrary, we feel that such arguments are not compelling.

Table 1. Average number of 'p' responses (out of a possible 25) across 12 subjects and five VOTs for the different F_0 combinations of Experiment 2

Offset	High	H–M	Mid	M–L	Low
Onset					
98	11.0	0.1	10.9	0.6	10.3
108	12.6	1.0	11.6	0.3	11.3
114	12.8	0.4	12.4	0.1	12.3
120	14.8	0.9	13.9	0.0	13.9
130	16.9	0.5	16.4	0.0	16.4
Average difference		0.6		0.2	

The 3 offsets are shown in the 1st, 3rd, and 5th columns. The differences between these columns are shown in the 2nd and 4th column, with an effect of intonation showing up as a positive difference.

It is clear that the use of a speeded task in Experiment 1 was not necessary for the improvement in sensitivity to F_0 onset cue value. In contrast to Experiment 1, in which a speeded task intervened between the two blocks, here simply tripling the number of syllables heard in the session (though each individual stimulus was heard the same number of times) was enough to reveal an effect of the lower onset values. Since the separation appeared even in the first condition, this may give us some indication of the critical number of trials for seeing this effect. Alternatively, the existence of F_0 changes in the remainder of the syllable may have increased the subjects' sensitivity to F_0 cues from the start. The results do not distinguish between these possibilities.

The F_0 of the 'steady-state' portion of the vocalic segment (that is, that part in which the formant values were constant) influenced the voicing judgment. This could have been due to the effect of intonation contour or of the vowel's mean F_0. If the intonation pattern was responsible, then even extremely brief, stylized patterns are sufficient to change the effect of the onset F_0s in the way predicted by Silverman [1986]. In this case, the pattern was not supported by a preceding environment, as it was in Silverman's case, yet subjects were able to respond to it in a coherent manner. There may, in fact, be other tonal patterns of English which could give rise to different predictions for the perturbations. However, it seems unlikely that subjects would infer more complex patterns for these monosyllables, in the same way that it seems unlikely that subjects in Silverman's experiment would infer an unusual implementation of the tonal sequences given. Although the present study is not definitive, it seems likeliest that subjects were perceiving a simple intonation, one in which the perturbations should have been separate for the rising intonation and indistinguishable in the falling.

These results therefore lend further support to the notion that even short intonations can be taken into account in perception. However, it is also possible that the changes over time were perceived as simple differences in the mean F_0. Since F_0 is higher throughout the vowel after a voiceless stop than after a voiced one [Ohde, 1984], this could also account for the significant but small effects found here. For patterns which begin at the same value, there cannot be a rising F_0 pattern without also entailing a higher mean F_0, so the two explanations are difficult to dissociate. The next experiment was an attempt to provide such a dissociation.

Experiment 3

In the final experiment, different F_0 patterns were used to dissociate the direction of the initial F_0 change from the mean F_0. The F_0 varied rapidly up and down through the syllable, giving a mean value in the middle of the range of variation. The initial direction of the contour could be either up or down, changing the relationship of the initial F_0 value to the underlying contour. Two more versions were created to have an overall higher mean F_0, or lower. In 2 of the 4 cases, the cue value of the F_0 onset relative to the contour and that of the mean F_0 were in opposition. While it is not necessarily the case that an intonational account could not also predict an effect of overall F_0, the present stimuli set the two in opposition.

Method

Stimuli

The formant values of the synthetic stimuli of Experiment 1 were used once again, but the F_0 pattern and duration were different. Only 2 of the original VOT values were used (15 and 20 ms). These were the most ambiguous in the previous experiments. The same 5 F_0 onset values were used (98, 108, 114, 120, and 130 Hz), and the F_0 again went linearly from each of those values to 114 after 100 ms of voicing. From that point, the F_0 contour took on a quasi-sinusoidal shape (fig. 4). This pattern either fell to a low of 98 Hz and then rose to a high of 130 (upper panel), or rose and then fell (lower panel). Two versions of each of these main contours were made. In one, the pattern after the second extreme was continued (see the solid func-

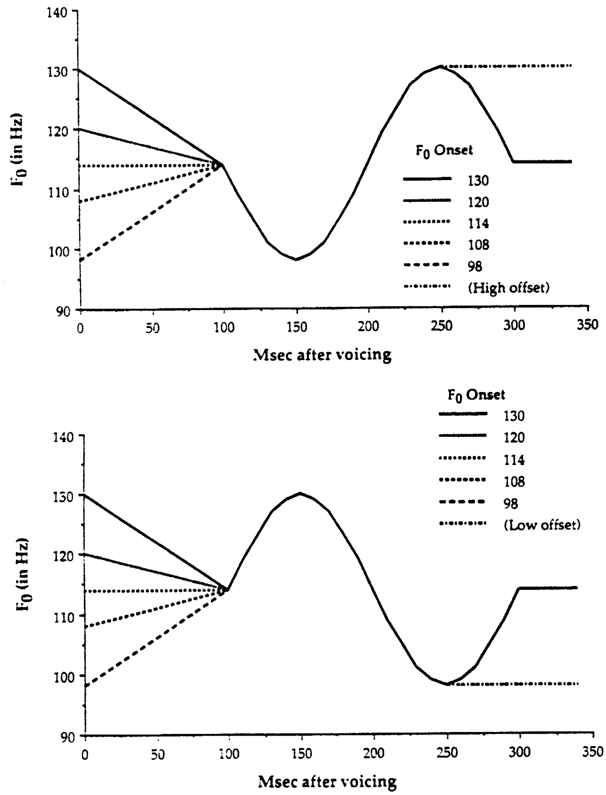

Fig. 4. Intonation patterns for the stimuli of Experiment 3. The upper panel represents two of these patterns, one which ends at the mid-point (114 Hz) and one which remains at the high level (High offset). The lower panel represents the other two patterns, one which ends at the mid-point and one which remains at the low point (Low offset).

tions in figure 4), or the extreme was maintained (see the dot-dash lines labeled 'high offset' and 'low offset'). These patterns will be called Down/Mid (solid function, upper panel), Down/High (dot-dash, upper panel), Up/Mid (solid function, lower panel), and Up/Low (dot-dash, lower panel).

These patterns do not correspond to any tonal sequence in English, since there are two pitch accents in one syllable. Thus the intonational pattern that they determine is determined by the regularity of the change, not by the rules of intonation. The contours did not, however, sound unnatural, rather more as if the speaker was indecisive. Since there has been some concern in the literature about the minimum time in which pitch changes can be accomplished [Ohala and Ewan, 1973; Lyberg, 1981], we validated the parameters used by synthesizing them via the vocal tract of the first author (a speaker not normally noted for large changes in F_0). Within the course of an hour or two, he was able to produce two 30-Hz changes within a 370-ms syllable. Although his F_0 is higher than that used here (and thus the percentage of the changes was not as large as in the experimental stimuli), it appears that the F_0 changes used in the stimuli are in fact humanly possible. In contrast, the F_0 perturbation portion of the pattern occurs within the time frame that subjects in an intonational imitation task could not mimic [Pierrehumbert and Steele, 1989]. Together, these considerations lead us to suppose that the perceived F_0 pattern would resemble the sinusoidal function of the post-perturbation F_0, and that the perturbation itself would be perceived as a perturbation from the projected continuation of the sinusoidal pattern.

Note that the cue value of these patterns differs depending on whether the intonation or the mean level is most relevant. If intonation contour, that is, the contour implied by the continuation of the F_0 pattern through the vocalic segment, is the most relevant variable, all five onsets should be distinct in the Up contours, and the offset should not matter, whereas all five onset values should be equivalent in the Down contours. In the F_0 level version (Low, Mid, High), Low should decrease 'p' judgments and High should increase them.

Procedure

The test consisted of 6 repetitions of each stimulus randomized and then recorded on audio tape as in the first experiment. The first repetition was considered to be a familiarization set, and was not scored (a procedure also used in Whalen and Beddor [1989]). Subjects listened as before and wrote their responses as 'b' or 'p'. Since a limited range of VOTs was used, the subjects were urged to use both categories even if they did not at first feel the need to.

Subjects

Data were obtained from 12 colleagues at Haskins Laboratories with no knowledge of the purpose of the research. All were native speakers of English, with no reported hearing difficulties, and with different degrees of experience with phonetics. (One subject reported that she heard nothing but 'b's and had adopted the 'artificial' strategy of relying on the intonation. Even though she in fact chose 'p' every time there was a high F_0 near the end of the syllable, she was replaced with another Haskins colleague, so that 12 subjects could be analyzed.)

Results and Discussion

An analysis of variance was performed on the number of 'p' responses per stimulus, with the factors VOT, F_0 Onset, Middle offset (which contrasts the 'High' and 'Low' contours with the two that end up at 114 Hz), and Up/Down (contrasting the two intonational patterns at the beginning of the syllable). VOT, though representing a step of only 5 ms, was highly significant as a main effect ($F(1,11) = 23.20$, $p < 0.001$). Since VOT did not enter into any interactions, it will be averaged across in subsequent displays.

Table 2 shows the responses for each of the 5 onset F_0s comparing the 4 contours. While F_0 Onset had a reliable effect ($F(4,44) = 62.43$, $p < 0.001$), neither Middle Offset ($F(1,11) < 1.0$) nor Up/Down ($F(1,11) = 1.80$, n.s.) did. The only significant interaction was between F_0 Onset and Middle Offset ($F(4,44) = 2.98$, $p < 0.05$). Although 'p' responses increased with increasing F_0 onset for each series, there was a large jump between 114 and 120 for the middle-offset stimuli, and a large jump between 120 and 130 for the high or low offset (fig. 5). We have no explanation for this interaction.

Two further analyses were performed to examine directly the two different sets of parameters within the test. The first looked at the Up and Down stimuli only. These differed in the direction of the initial F_0 pattern but not in the average F_0 of the vocalic segment after the perturbation. F_0 onset was again significant ($F(4,44) = 50.77$, $p < 0.001$), and the lower 3 onsets did not differ significantly from each other (despite the obvious pattern in figure 5). This would be in accord with Silverman's [1986] predictions if the contour was functionally flat. If, however, the intonation was an extension of the sinusoidal pattern, then these results do not fit his prediction. There was no effect whatsoever of the direction of the initial change ($F(1,11) = 0.94$, n.s.). None of the interactions was significant either. So although the results for the onsets by themselves seemed to fit the intonational predic-

Table 2. Average number of 'p' responses (out of a possible 5) across 12 subjects and two VOTs for the different F_0 combinations of Experiment 3

	F_0 onset, Hz					
	98	108	114	120	130	Average
Intonation						
Up/Mid	0.88	1.50	1.71	2.71	3.54	2.07
Down/Mid	1.25	1.96	2.04	3.04	3.96	2.45
Up/Low	1.08	1.75	1.75	2.08	3.54	2.04
Down/High	1.29	2.08	2.71	2.75	4.29	2.63

See text for an explanation of the intonation types and the pattern expected from them.

tions, the intonation type itself did not have any effect.

The complementary analysis focused on the high- and low-offset stimuli. As mentioned before, the direction of the contour and the overall F_0 level would lead toward different predictions for the two stimuli. The departure from the contour would predict more 'p' responses for the low offset, which had an initial rise in the F_0. However, the overall F_0 would favor 'p' responses for the high offset over the low offset. Here again, in this analysis, the F_0 Onset was quite significant ($F(94,44) = 36.67$, $p < 0.001$). The High/Low factor was at best marginally significant ($F(1,11) = 2.75$, $p < 0.15$). However, the trend that was there favored the overall F_0 predictions, since the high offset had more 'p' responses than the low (2.63 vs. 2.04).

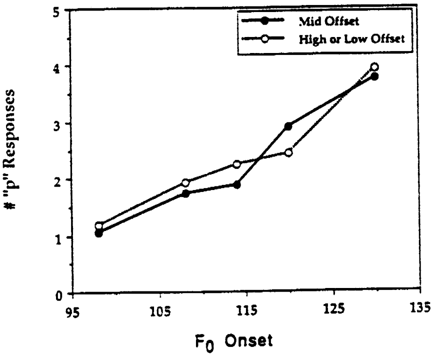

Fig. 5. Average number of 'p' responses (out of a possible 5) for the 2 contours which end at the mid-level (solid function) versus the 2 which end either at the high value or the low value (dotted function). Experiment 3.

The results of this experiment must be interpreted with caution. both because of the weakness of the statistical power and because of the unnaturalness of the intonation contours. In particular, subjects may have been inferring complex tonal sequences which would push the underlying F_0 pattern in directions different from those specified by the sinusoidal pattern. However, it seems apparent that even with unnatural contours, subjects can ascribe part of the F_0 pattern to the voicing percept in a way which is marginally gradient. Further, the gradience is reliable for the two falling patterns, and only suggestive for the two rising patterns. This seems to be the case regardless of the intonation contour of the remainder of the stimulus.

Conclusion

F_0 can influence the perception of voicing in syllable-initial stops of ambiguous VOT. Although this has been found before [Haggard et al., 1970; Fujimura, 1971; Haggard et al., 1981; Abramson and Lisker, 1985], the present results show more clearly that onset F_0 results in a gradient effect, though there is a difference between the lower and higher onsets. That is, the higher the F_0 onset, the more voiceless responses. For the lower onsets, there is a statistically marginal tendency for a similar gradient effect, with more voiced responses being elicited by lower F_0s.

The F_0 of the vocalic segment also affects the voicing judgment, though it is not clear at present whether this is due to the relationship of the perturbation to the underlying intonation (as proposed by Silverman [1986]) or to the mean level of the F_0 (a relationship reflected in measurements of natural productions; Ohde [1984]; Whalen [1990]). A more detailed prediction of Silverman's, that all values of F_0 below the intonation contour have equivalent effects on the voicing judgment, should be modified in light of the present results. Several statistical properties suggest that the difference between the higher values and the lower ones is real. It takes more judgments for the lower values to begin to separate (comparing the present results with both Abramson and Lisker [1985] and Silverman [1986], who used fewer judgments), and the post-hoc tests in the present results tend to show clear separation in the higher region and not in the lower. However, in the third experiment here, where direction of postulated intonation was dissociated from average F_0, only the absolute direction of the F_0 onset was relevant. Different intonational patterns gave the same results in this case. Given the success that the intonational account has had in explaining some of the production results, there is no need at present to reject it. However, for the stimuli used here, it seems likely that the effect of intonation, if present, is weaker than the effect of mean F_0.

As speech research moves from more obvious acoustic properties to less, it becomes necessary to use more sensitive measures [Whalen, 1984]. When only the most relatively powerful cues are examined, a few judgments on several tokens are sufficient to establish interesting facts. As the cues become relatively less powerful, more detail is needed. In this case, a simple increase in the number of judgments on (rather unnatural) synthetic syllables revealed a more complex pattern than had previously been found. Other conclusions may require even more sensitive measures, such as reaction time. We are currently pursuing such measures to see in even greater detail the effect of F_0 on voicing judgments.

Acknowledgments

This research was supported by NICHD Grant HD-01995 to Haskins Laboratories. Portions of this research were presented at the 116th Meeting of the Acoustical Society of America, Honolulu, November 1988. We thank Andrea G. Levitt, Carol A. Fowler, Kim Silverman, Randy Diehl, and an anonymous reviewer for extremely helpful comments.

References

Abramson, A. S.: Laryngeal timing in consonant distinctions. Phonetica *34:* 295–303 (1977).

Abramson, A. S.; Lisker, L.: Relative power of cues: F_0 shift versus voice timing: in Fromkin, Phonetic linguistics: Essays in honor of Peter Ladefoged, pp. 25–33 (Academic Press, New York 1985).

Fujimura, O.: Remarks on stop consonants – synthesis experiments and acoustic cues; in Hammerich, Jakobson, Zwirner, Form and substance: phonetic and linguistic papers presented to Eli Fischer-Jørgensen, pp. 221–232 (Akademisk Forlag, København 1971).

Haggard, M.; Ambler, S.; Callow, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47:* 613–617 (1970).

Haggard, M. P.; Summerfield, Q.: Roberts, M.: Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading f_0 cues in the voiced-voiceless distinction. J. Phonet. *9:* 49–62 (1981).

Hombert, J. M.: Towards a theory of tonogenesis: an empirical, physiologically and perceptually, based account of the development of tonal contrasts in language; thesis Berkeley (1975).

Hombert, J. M.: Ohala, J.: Ewan. W.: Phonetic explanation for the development of tones. Language *55:* 37–58 (1979).

House, A. S.; Fairbanks, G.: The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. acoust. Soc. Am. *25:* 105–113 (1953).

Lehiste, I.: Peterson, G. E.: Some basic considerations in the analysis of intonation. J. acoust. Soc. Am. *33:* 419–423 (1961).

Löfqvist, A.: Intrinsic and extrinsic F_0 variations in Swedish tonal accents. Phonetica *31:* 228–247 (1982).

Löfqvist, A.; Baer, T.; McGarr, N.; Story R.: The cricothyroid muscle in voicing control. J. acoust. Soc. Am. *85:* 1314–1321 (1989).

Lyberg, B.: Some consequences of a model for segment duration based on F_0-dependence. J. Phonet. *9:* 97–103 (1981).

Moore, B. C. J.: Frequency difference limens for short-duration tones. J. acoust. Soc. Am. *54:* 610–619 (1973).

Ohala, J. J.; Ewan, W. G.: Speed of pitch change. J. acoust. Soc. Am. *53:* 345(A) (1973).

Ohde, R. N.: Fundamental frequency as an acoustic correlate of stop consonant voicing. J. acoust. Soc. Am. *75:* 224–230 (1984).

Pierrehumbert, J. B.; Steele, S. A.: Categories of tonal alignment in English. Phonetica *46:* 181–196 (1989).

Silverman, K.: F_0 segmental cues depend on intonation: the case of the rise after stops. Phonetica *43:* 76–91 (1986).

Silverman, K.: The structure and processing of fundamental frequency contours; thesis Cambridge (1987).

Whalen, D. H.: Subcategorical phonetic mismatches slow phonetic judgments. Percept. Psychophys. *35:* 49–64 (1984).

Whalen, D. H.: Coarticulation is largely planned. J. Phonet. *18:* 3–35 (1990).

Whalen, D. H.; Beddor, P. S.: Connections between nasality and vowel duration and height: elucidation of the Eastern Algonquian intrusive nasal. Language *65:* 457–486 (1989).

Received: November 10. 1989
Accepted: August 8, 1990

Dr. D. H. Whalen
Haskins Laboratories
270 Crown Street
New Haven. CT 06511 (USA)