*F*0 gives voicing information even with unambiguous voice onset times

D. H. Whalen
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511

Arthur S. Abramson
Haskins Laboratories, New Haven, Connecticut 06511 and The University of Connecticut, Storrs, Connecticut 06269

Leigh Lisker
Haskins Laboratories, New Haven, Connecticut 06511 and The University of Pennsylvania, Philadelphia, Pennsylvania 19104

Maria Mody
Haskins Laboratories, New Haven, Connecticut 06511 and the City University of New York, New York, New York 10036

(Received 19 May 1992; revised 1 October 1992; accepted 4 December 1992)

The voiced/voiceless distinction for English utterance-initial stop consonants is primarily realized as differences in the voice onset time (VOT), which is largely signaled by the time between the stop burst and the onset of voicing. The voicing of stops has also been shown to affect the vowel's *F*0 after release, with voiceless stops being associated with higher *F*0. When the VOT is ambiguous, these *F*0 "perturbations" have been shown to affect voicing judgments. This is to be expected of what can be considered a redundant feature, that is, that it should carry a distinction in cases where the primary feature is neutralized. However, when the voicing judgments were made as quickly as possible, an inappropriate *F*0 was found to slow response time even for unambiguous VOTs. This was true both of *F*0 contours and level *F*0 differences. These results reinforce the plausibility of tonogenesis, and they add further weight to the claim that listeners make full use of the signal given to them, even when overt labeling would seem to indicate otherwise.

PACS numbers: 43.71.Es, 43.71.An

INTRODUCTION

The voicing of utterance-initial stop consonants is perceptually determined primarily by the voice onset time (VOT), which is largely signaled by the time between release of the stop and the onset of voicing (Lisker and Abramson, 1964). This has been found to be true not only of languages such as Spanish that rely primarily on the presence or absence of voicing during closure, but also of languages such as English that, in some environments, do not voice the closure for the voiced stops and overlap the voicelessness with the release of the stop as aspiration. In addition, a falling fundamental frequency (*F*0) usually occurs after a voiceless stop, while a flat or rising *F*0 usually accompanies voiced stops (House and Fairbanks, 1953; Lehiste and Peterson, 1961; Ohde, 1984; Silverman, 1987). This has been called the *F*0 "perturbation" and has been found in a wide variety of languages (Hombert, 1975). In perception, perturbation effects have usually appeared only when the VOT was ambiguous, at least when the perturbation was of the same magnitude as found naturally (Fujimura, 1971; Abramson and Lisker, 1985; Whalen, Abramson, Lisker and Mody 1990). That is, an ambiguous VOT is more likely to be heard as voiceless when the *F*0 is falling after the onset of voicing than if it is flat or rising.

Studies of natural productions of stops have found that most measured VOT values fall within ranges that are unambiguously interpreted (Lisker and Abramson, 1964; Shimizu, 1989). Thus it may appear that the perceptual effects of *F*0 on voicing, though demonstrable, are unimportant in the actual use of language. Abramson and Lisker (1985, p. 32), for example, state that voicing judgments for certain VOT values "cannot be affected by *F*0." However, several studies have shown that acoustic differences that do not affect overt labeling can nonetheless affect speech processing, as shown by reaction times (Martin and Bunnell, 1981; Whalen, 1984; Whalen and Samuel, 1985; Tomiak, Mullennix and Sawusch, 1987). These studies focused on subcategorical mismatches created by splicing segments of natural speech from one (appropriate) environment to another (an inappropriate one). The mismatches have involved both vowel-to-vowel and fricative/vowel coarticulation. While the effects provide clear support for the idea that subjects are sensitive to all the linguistic information given to them, it could also be argued that the mismatch might have included an abrupt change in a resonance, which might be seen as a nonlinguistic source of the delay. This might hold true even in those cases where all the stimuli had been cross-spliced, at least from one token to another (Whalen and Samuel, 1985).

The present study was designed to extend the results on mismatched cues to a situation in which there is no possibility that the coherence of the signal has been violated. We chose the influence of the $F0$ perturbation on identifying VOT continua, as demonstrated in previous work (Fujimura, 1971; Abramson and Lisker, 1985; Whalen *et al.*, 1990), because in any sequence of voiceless stop followed by a voiced vowel, there must of course be a shift from a voiceless to a voiced source. Thus if there is any auditory "discontinuity" inherent in changing from a voiceless to voiced source, it is one that is normal for speech. The only manipulation we had to make was to vary the onset $F0$ value, a choice that should not, in itself, give rise to any auditory discontinuities. If responses are slower when the $F0$ information does not match that of the VOT, we can be even more confident that all the acoustic consequences of a speech gesture contribute to the perception of speech, even if the labeling fails to show it.

In addition, we wanted to assess the ability of listeners to detect these $F0$ differences when the VOT is unambiguous. It is the implication of certain language changes that the $F0$ perturbations are used perceptually: Many cases of the emergence or diversification of tone systems have been traced to the loss or realignment of a voicing distinction with a concomitant use of the perturbation's effect (Hombert, 1975; Hombert *et al.*, 1979; Maddieson, 1984). While the diachronic facts have not been questioned, it has remained an uneasy assumption that such small $F0$ differences could in fact be perceived in a natural context. The present experiments will show, at least, that these differences in $F0$ do affect perception, making the theory of tonogenesis that much more plausible.

I. EXPERIMENT 1

The first experiment uses reaction time to determine whether there is perceptual use of $F0$ for voicing judgments even when the labeling shows no effect. Such a subcategorical effect can be seen in the reaction times to stimuli with unambiguously labeled VOTs.

A. Method

1. Stimuli

Synthetic approximations to the English syllables /ba/ and /pa/ were created with the serial synthesizer designed by Ignatius G. Mattingly at Haskins Laboratories (as in Whalen *et al.*, 1990). The vowel steady-state formants were centered at 730, 1250, and 2440 Hz with bandwidths of 100, 100, and 125 Hz. (Since this was a serial synthesizer and the bandwidths were kept constant, there was no "$F1$ cutback.") The formant values at the beginning of the syllable were 450, 1080, and 2300 Hz for $F1$, $F2$, and $F3$, respectively, and they changed linearly to reach the steady-state values after 75 ms. Vowel amplitude was level until the last 30 ms of the syllable, at which point it decreased linearly to zero. Total duration of the syllables was 250 ms.

The VOT values were 5, 10, 15, 20, 25, 35, and 50 ms after the simulated release. These were obtained by turning off the voicing source (AV) and introducing aperiodic hiss (AH) for the appropriate number of synthesis frames. The $F0$ onset values were 98, 108, 114, 120, and 130 Hz. $F0$ changed linearly from these values to the steady-state $F0$ of 114 Hz over the first 50 ms of voicing. (Of course, with an onset of 114 Hz, the $F0$ was constant over the entire syllable.) These differences of onset values are similar in magnitude to those reported in the literature (e.g., Ohde, 1984). Each VOT was paired with each $F0$ onset, giving 35 unique stimuli.

2. Procedure

The stimuli were presented for identification as "b" or "p," with responses being made by pressing buttons labeled "b" (on the left-hand side) and "p" (on the right). All subjects participated in five conditions, three unspeeded and two speeded. The first was an unspeeded condition containing five repetitions of all 35 stimuli, while the other two, which differed only in the randomization, used a subset of these consisting of the 23 stimuli which appeared in either of the two speeded conditions. One speeded condition, the $F0$ condition, used all $F0$ onset values, but only the 5, 20, and 50 ms VOTs. The other speeded condition, the VOT condition, used all VOT values but only the extreme $F0$ values (98 and 130 Hz). Twenty repetitions of the stimuli were randomized for the speeded conditions.

The order of conditions was as follows. First came the unspeeded condition with all the stimuli. Then came the first speeded condition, which was the $F0$ condition for half of the subjects and the VOT condition for the other half. After the first speeded condition came an unspeeded condition with the selected stimuli. Next, the other speeded condition was given, so that all subjects had both conditions. Finally, the unspeeded task for the selected stimuli was given one more time.

In the unspeeded conditions, subjects were to press the button when they had made their decision, limited in time only by the 2.5 s between stimuli. If unsure, they were to guess. In the speeded conditions, they were to make their response as quickly as possible, using one finger of their right (dominant) hand to press the buttons. Between trials, they rested this finger on the keyboard.

3. Subjects

The subjects were 12 young adults from the Yale University community who had volunteered for listening experiments. All passed an audiometric screening for both ears. They were paid for their participation.

B. Results

1. Unspeeded conditions

The first unspeeded condition showed the pattern found in our earlier work (Whalen *et al.*, 1990): The three lowest $F0$ values elicited about the same percentage of "b" responses collapsed over VOT (48.6%, 51.0%, and 52.2% for 98, 108, and 114 Hz onsets, respectively), while the two higher values elicited fewer (43.6% for the 120-Hz onset and 36.4% for the 130). Figure 1 shows the effect on the judgments by giving the overall percentage of "p" responses for stimuli beginning with the stated $F0$ collapsed across all VOT values.

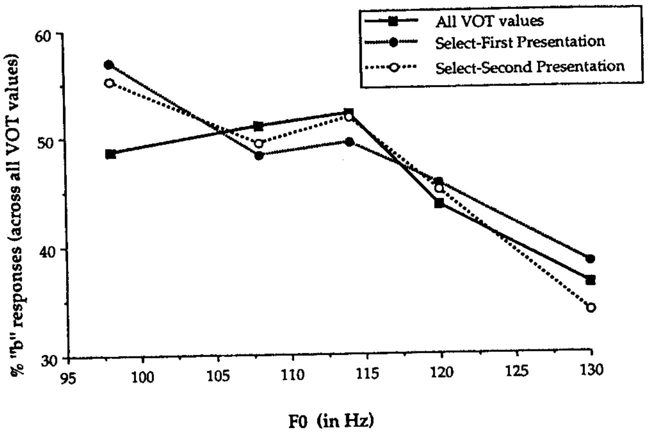

FIG. 1. Responses for the 12 subjects in the three unspeeded conditions of experiment 1, expressed as a percentage of "b" responses averaged across all VOT values.

As in our earlier study, $F0$ did not influence judgments for unambiguous VOT's, that is, those extreme values of VOT that received at least 80% judgments in one category (Fig. 2). Although the $F0$ functions do not converge except in the third panel for the two extreme VOT values, there is no consistency in the ordering of the $F0$ functions the way there is in the ambiguous region.

The functions in Fig. 1 are not monotonic, possibly due to the reduction in the number of stimuli presented. Recall that in the two unspeeded conditions with the selected stimuli, only three VOT values were used rather than seven of the $F0$ settings. This gives us somewhat less resolution in our measures, though it is acceptable for the replication of our earlier work (Whalen *et al.*, 1990).

There are two ways of looking at the selected conditions. The first is by whether the unspeeded test followed the $F0$ speeded condition or the VOT; the second is by the experimental order, i.e., whether the unspeeded test was the third or fifth in the experiment. The results will be considered in their experimental order, though in fact, it does not matter much since the two conditions are quite similar (Fig. 1). The general result is clear, and consistent with our earlier finding: After the initial experience with these stimuli, subjects begin using $F0$ for voicing information in a fairly gradient fashion from the lowest $F0$ to the highest. The functions are not monotonic, but they are clearly different from the first condition. The only stimuli that could show a large difference are the 98-Hz stimuli, and they do show one. An analysis of variance on the proportion of "b" with the factors $F0$ and Condition shows a significant main effect of $F0$ [$F(4,44)=39.32, p<0.001$]. Condition was not a significant main factor [$F(2,22)<1$, n.s.], but the interaction was [$F(8,88)=2.63, p<0.05$]. Further analyses (separate ANOVAs for the three pairings of the conditions) showed that the first condition differed from each of the selected ones, which did not differ from each other. (A significant interaction of $F0$ and condition appears with the first condition and each of the selected ones [$F(4,44)=3.32$ and $2.60, p<0.05$]; the selected conditions did not show an interaction [$F(4,44)=1.72$, n.s.].)

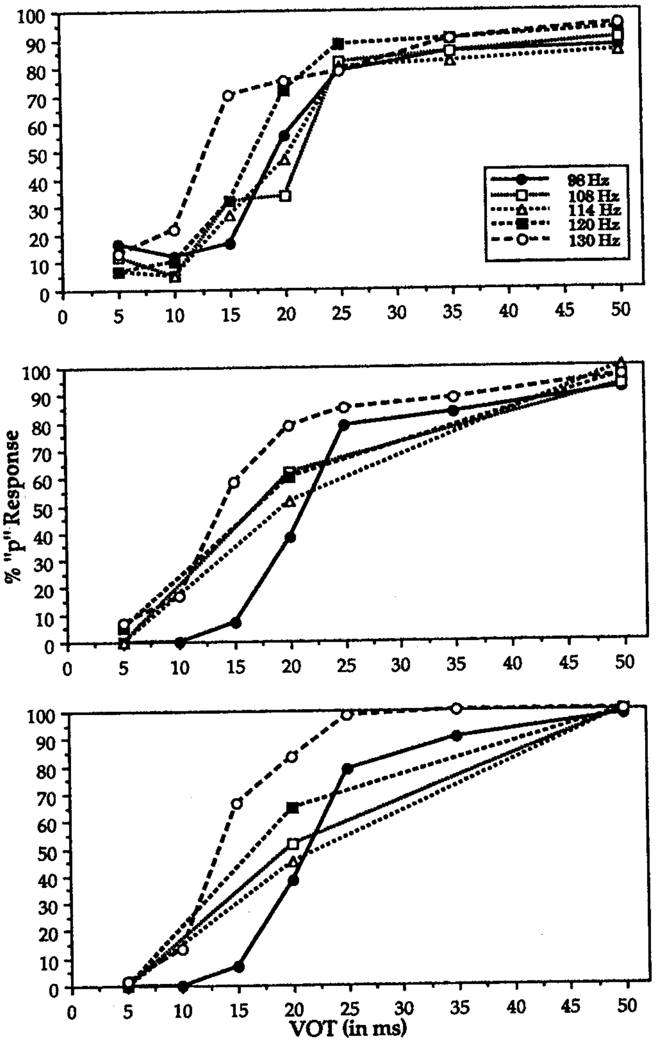

FIG. 2. Responses for the 12 subjects in the three unspeeded conditions of experiment 1, shown separately for each condition. The top panel is the first unspeeded condition. The middle panel is the first selected condition, and the bottom panel is the second selected condition.

After their initial exposure to these stimuli, the subjects make a more gradient use of $F0$ in their voicing judgments, as had been found in our earlier work (Whalen *et al.*, 1990).

2. Speeded condition: $F0$

The results for the $F0$ condition are shown in Fig. 3. Reaction times for all 12 subjects are averaged together. The "b" responses to the 50-ms VOT stimulus and "p" responses to the 5-ms VOT stimulus were excluded as being mistakes (based on the unspeeded results). These accounted for 1.6% of the responses. At the 20-ms VOT value, both responses were considered correct.

The most important result to be seen in this figure is that the extreme $F0$'s are associated with different reaction times depending on the category label applied. The extreme $F0$ values are given in the thicker lines. For "b" judgments, both at the unambiguous and the ambiguous VOTs, the 98-Hz onset gave faster times than the 130-Hz

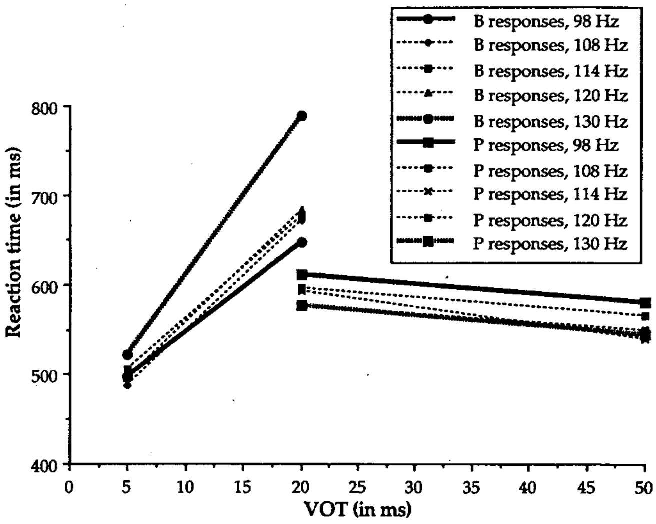

FIG. 3. Reaction times for the complete range of *F*0 onsets, experiment 1.

onset. Conversely, the 130-Hz onset gave the faster times for the "*p*" judgments. The other *F*0 values (shown with thinner lines) tend to range in between, but their arrangement is not monotonic. There is, perhaps, not enough resolution within this rather narrow range of reaction times for a monotonic pattern to emerge with this number of repetitions.

For statistical analysis, the unambiguous items (5, 10, 35, and 50 ms VOTs) were analyzed together, and then the responses to the ambiguous stimulus. The factors were response category and *F*0. Analysis of the means and standard deviations indicated the presence of an inhomogeneity of variance, which was minimized by using a speed transform, that is $1/RT$. All reported numbers are retransformed into times to make comparisons to other studies easier.

Response category was a significant main effect for the unambiguous VOTs [$F(1,11)=7.77$, $p<0.05$], since the "b" responses were faster by some 50 ms. *F*0 was not a significant main effect [$F(4,44)=2.14$, $p<0.10$], but the interaction of the two was [$F(4,44)=5.97$, $p<0.001$]. The interaction shows strong evidence of a differential effect of *F*0 depending on which category is selected as the response. Separate analyses for each response category shows *F*0 to be significant in itself [$F(4,44)=3.88$, $p<0.05$ for the "b" responses, $F(4,44)=4.29$, $p<0.01$ for the "p" responses]. We may therefore conclude that the appropriateness of the *F*0 affected decision time.

The magnitude of the mean reaction time difference to voiced and voiceless categories happens to be almost that of the difference in VOTs (50 ms in reaction time versus 45-ms difference in VOT), but this is likely to be due to factors other than the VOT itself. The 50 ms/VOT, though consistently heard as "p," is likely to be further from the prototypical value for /p/ than the 5 ms/VOT is for /b/ (Miller and Volaitis, 1989). If so, then we would expect reaction times to be longer (Pisoni and Tash, 1974; Whalen, 1991), since less prototypical tokens are harder to identify. If our function went further along the scale, the "b" and "p" times might become equivalent. Additionally, the synthesis parameters, notably the absence of a burst and the lack of *F*1 attenuation in the aspirated portion, may have been more detrimental to the "p" category than the "b."

TABLE I. Reaction times for the two response categories. Each group consists of nine subjects, six being common to both.

*F*0 (in Hz):	98	108	114	120	130
Means for "b"	657	686	712	683	721
Means for "p"	630	581	595	590	545

The mean values for the ambiguous items, shown in Fig. 3, display the effect that we would expect, with the *F*0 values affecting the two judgments differently. Unfortunately, half the subjects failed to find this VOT value ambiguous in this condition, with the consequence that their minority-category judgments are too few to give a reliable mean value. As it turns out, of the six subjects who did not find the stimuli ambiguous, three heard them primarily as "b" and three heard them mostly as "p." Therefore, two separate analyses were done, one for "b" and one for "p." The "b" analysis included the six who heard the stimuli ambiguously plus the three who made at least 80% "b" judgments. Two of the six had no "b" responses to the 130-Hz stimulus. Those two cells were filled with the value for the most similar stimulus, namely, the 120-Hz stimulus. The "p" analysis again included the six who heard the stimuli ambiguously plus the other three subjects, those who made at least 80% "p" judgments. The means for the two sets of nine subjects are shown in Table I. Despite the 64-ms difference, in the predicted direction, between the 98- and 130-Hz stimuli, the "b" analysis showed no effect of *F*0 [$F(4,32)<1$, n.s.]. The 85-ms difference, again in the predicted direction, for the "p" responses was a component of a significant effect [$F(4,32)=5.37, p<0.01$]. Thus, though not conclusive, the evidence shows that the *F*0 also has an effect on reactions times to ambiguous stimuli, as well as on the labeling.

3. Speeded condition: VOT

The results for the VOT condition are shown in Fig. 4. As with Fig. 3, the reaction time values are the means of the speeds for the 12 subjects reconverted into times. Even though the full VOT range was used, it was still the case that "b" responses above 20 ms and "p" responses below 20 ms were too few to analyze, so the two sets of functions overlap only at 20 ms. Here again, it can be seen that the 130-Hz onset slows decisions for "b" relative to the 98-Hz onset, while the reverse is true for "p" decisions. The effect appears larger at the ambiguous value for "b" but is absent for "p" at the ambiguous value.

For the statistical analysis of the unambiguous stimuli, only "b" responses to the short VOT stimuli and "p" responses to the long VOT stimuli were used. The 15, 20, and 25 ms VOTs were excluded from the analysis since they did not receive the 80% majority judgments needed to be

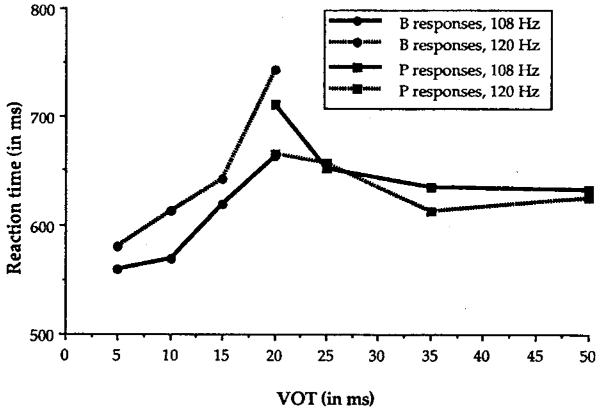

FIG. 4. Reaction times for the complete range of VOT values, experiment 1.

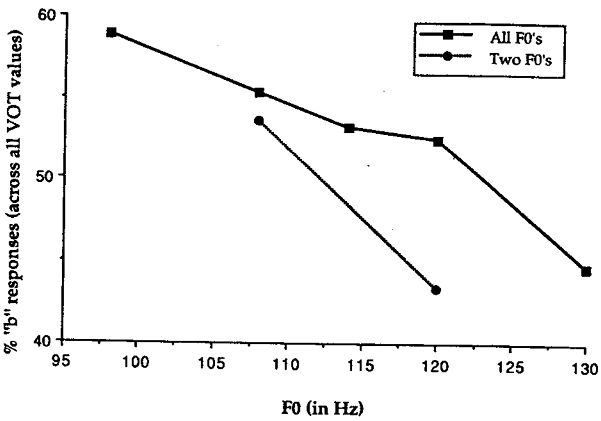

FIG. 5. Responses for the 11 subjects in the two unspeeded conditions of experiment 2.

called unambiguous. The factors were Response Category, $F0$, and VOT. The VOT factor represented 5 and 10 for the "b" responses and 35 and 50 for the "p" responses.

The interaction of $F0$ and category, the most important result for the present study, was significant [$F(1,11) = 6.75$, $p < 0.05$], indicating that the effect of the same $F0$ differed depending on which category was being assessed. Inappropriate $F0$ caused an average delay of 17 ms in the identification. Response category was also significant [$F(1,11) = 5.44$, $p < 0.05$], with "b" judgments being 52 ms faster. Neither $F0$ nor VOT was significant as a main effect [$F(1,11) < 1$, n.s., and $F(1,11) = 1.52$, n.s., respectively]. The interaction of response category and VOT was significant [$F(1,11) = 6.77$, $p < 0.05$]. Times were somewhat slower the less extreme the VOT, as we would expect given previous results with stimuli that approach the ambiguous region (Pisoni and Tash, 1974; Whalen, 1991). The three-way interaction was not significant [$F(1,11) = 2.34$, n.s.].

C. Discussion

The unspeeded tests confirmed our previous results, showing some increase in the use of the $F0$ information as the subjects became more familiar with the stimuli. The speeded conditions showed that even when the VOT was unambiguous, the $F0$ information is taken into account.

II. EXPERIMENT 2

The first experiment examined the dynamic perturbations at vowel onset. However, there may be differences throughout the vowel due to the stop voicing (e.g., Ohde, 1984; Whalen, 1990), though such differences are not found in every study (e.g., Löfqvist *et al.*, 1989). These level differences seem more relevant to tonogenesis than the initial perturbation since tonogenesis usually results in level tones, not contour tones. While the initial, dynamic portion of the $F0$ perturbation has been found in many studies, it seems that only Repp (1975) has examined level $F0$ differences. That study used dichotic presentation, where different syllables were presented to the two ears simultaneously and so is rather far removed from normal perception. The next experiment provides a simpler demonstration of the perceptual effectiveness of level $F0$ differences on voicing judgments.

A. Method

1. *Stimuli*

The stimuli were much like those in the first experiment, except that the $F0$ value at the onset was carried throughout the vowel. Thus the stimuli with an $F0$ of 98 Hz had 98 Hz throughout the vocalic segment, not just at the onset. As before, there were five $F0$ values (98, 108, 114, 120, and 130) and seven VOT values (5, 10, 15, 20, 25, 35, and 50).

2. *Procedure*

Two unspeeded conditions were run. In the first, five repetitions of all 35 stimuli were randomized together. In the second, only two values of $F0$ were used, namely, 108 and 120. These were the two values of a "b" $F0$ and a "p" $F0$ that differed by an amount closest to the 9 Hz difference found by Ohde (1984) in the midpoint of spoken vowels.

The final condition was a speeded one that used two values of $F0$ (108 and 120) and all seven VOT values. The instructions and equipment were the same as used in experiment 1.

3. *Subjects*

Twelve subjects from the same pool used for experiment 1 were run. Three had participated in experiment 1. A technical problem resulted in the loss of the data for one subject, so the results of the remaining 11 will be reported.

B. Results

1. *Unspeeded conditions*

Level $F0$s affected voicing judgments, just as $F0$ contours had. As can be seen in Fig. 5, the percentage of "b" responses declines as the $F0$ increases, as predicted. For the statistical analysis, the total number of "b" responses across all the VOTs for each $F0$ was subjected to an analysis of variance with the single factor of $F0$ level. The effect

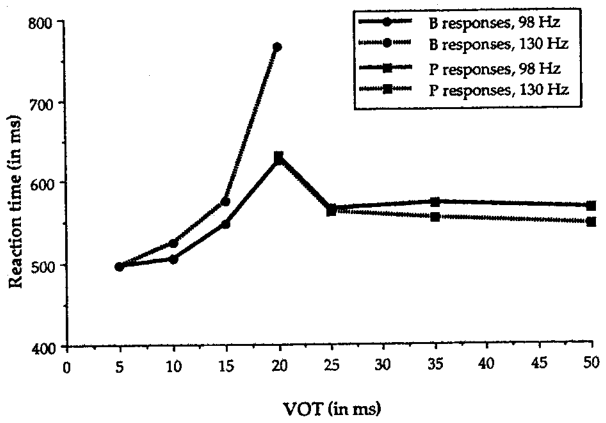

FIG. 6. Reaction times for experiment 2.

of $F0$ on the number of "b" judgments is highly significant [$F(4,40)=7.65$, $p<0.001$]. A Newman–Keuls *post-hoc* test reveals that responses to the highest $F0$ value are distinct from all the other $F0$ values, which do not differ among themselves. Still, it is clear that the $F0$ value of the syllables as a whole, not just the $F0$ perturbation, can affect the voicing judgment.

The second unspeeded condition was intended to ascertain whether subjects were able to treat the two $F0$ values as, in fact, separate. Since each syllable was presented as an isolated utterance, there was no immediate context, other than the experiment itself, by which to judge the relative height of the $F0$. So it was conceivable that the subjects would lose track of the "baseline" $F0$ and fail to show an effect.

As it turned out, there was a quite robust effect of the two $F0$ values on voicing identification, as seen in the dotted function of Fig. 5. Taking the "b" responses to all of the VOT values, we find that the lower $F0$ elicited 53.7% "b"s, while the higher $F0$ elicited 43.4%. This difference was highly significant by a t test [$t(10)=5.82$, $p<0.001$], showing that subjects were able to hear and make use of a 12-Hz difference in $F0$ across tokens. Arguably, subjects may have taken the 120-Hz value as being at the top of the range, since it had an effect similar to the 130-Hz value in the previous condition (see Fig. 5).

2. Speeded condition

Before examining the response times, we need to check whether the identification were consistent with those in the unspeeded task. If, for example, the time pressure reduced the effect of $F0$ on identification, we would not be able to interpret any reaction time difference. In fact, the percentages were almost the same as before, with 53.2% "b"s for the 108 Hz $F0$, and 44.5% for the 120 Hz. This difference was also significant by a t test [$t(10)=4.13$, $p<0.01$]. Clearly, $F0$ retains its cue value in the speeded test.

The reaction time results are presented in Fig. 6. As before, these are the means of the speeds of the responses retranslated into times. The individual boundaries for each subject varied much more than before, so that some subjects did not find the 20-ms stimulus ambiguous. Indeed, no matter where the ambiguous region was defined, there were subjects who did not, in fact, find that region ambiguous. So only the two least ambiguous VOT values of each category could be analyzed. The means of the "b" responses to the 5 and 10 ms VOTs were analyzed in one ANOVA with the factors VOT and $F0$, and the means of the "p" responses to the 35 and 50 ms VOTs were analyzed in another with the same factors. The "b" responses were significantly faster when the lower $F0$ was present [$F(1,10)=7.21$, $p<0.05$]. Subjects were also significantly faster on the 5 ms VOT than on the 10 [$F(1,10)=8.48$, $p<0.05$], as could be expected from previous work (Pisoni and Tash, 1974; Whalen, 1991). The interaction was not significant [$F(1,10)=1.15$]. For the longer VOTs, the $F0$ effect did not reach significance [$F(1,10)=1.00$], although the means are in the expected direction. Two subjects had large numbers of "b" responses throughout the continuum, which would contribute to making that end of the continuum less reliable than the short end. Neither the VOT effect [$F(1,10)<1$, n.s.] nor the interaction [$F(1,10)<1$, n.s.] was significant.

The differences in the boundary between voiced and voiceless varied so much that it was impossible to pick a single value of VOT at which all subjects had responses in both categories. In fact, the majority of the subjects (6 of the 11) did not have ambiguous cells for both keys at any one VOT. The statistical analysis of the ambiguous stimuli was therefore not attempted. Graphically, Fig. 6 shows us the expected pattern of longer times for responses to stimuli with inappropriate $F0$ values.

So, despite the lack of significance in the longer VOTs, the shorter ones clearly show that inappropriate $F0$ values slow the identification of unambiguous syllables.

C. Discussion

The unspeeded tests showed that different $F0$ levels, even when they are only anchored by the experimental context, can be interpreted as voicing information. The speeded conditions showed that these level $F0$s could also be interpreted when the VOT was unambiguous, and the inappropriate values delayed identification time.

III. GENERAL DISCUSSION AND CONCLUSION

The voicing feature in English is realized as an aspiration difference (positive VOT) in utterance-initial position. Another aspect of initial voicing is that after voiceless stops, fundamental frequency ($F0$) falls somewhat when the voicing of the vocalic segment begins, and remains somewhat higher throughout the vowel, in contrast with voiced stops. This $F0$ difference has been shown to affect labeling only when the VOT was ambiguous (e.g., Abramson and Lisker, 1985). In the present experiments, listeners identifying stimuli that varied in VOT and $F0$ made use of the $F0$ information for ambiguous VOTs. This is an expected response to redundant features. The listeners also took the $F0$ into account with unambiguous VOTs, however, indicating that the redundant features are *always* taken into account.

Since the present experiments used synthetic stimuli to explore the listener's behavior, there is, as always, the possibility that these results will not generalize to more natural stimuli and situations. The most clear case in the literature is the effect of lexical influence on voicing judgments (Ganong, 1980), which was found to disappear when the synthesis was improved (Burton *et al.*, 1989; McQueen, 1991). These concerns are greatly mitigated by the evidence of tonogenesis (see below) and by the small learning effect in the first experiment. Subjects were better able to use the $F0$ information within the "b" category after the initial exposure to the stimuli. It seems unlikely that an effect that depended on the strangeness of the stimuli would increase as familiarity with those stimuli increased. It would seem, rather, that subjects became more familiar with this "voice" and were able to accord it its full range of expression after the initial strangeness. This is also likely given that the two cues being dealt with in the present study are clearly phonetic, while the lexical effect mixes the phonetic with the extra-phonetic.

The mismatches inherent in these stimuli cannot be accounted for in psychophysical terms. If the stimuli included a mismatch that the ear could not resolve, we would expect there to be processing delays. However, for the vast majority of English utterance-initial stops, there must be a point at which the noise source gives way to a voiced source, whether this positive VOT is long or short and whether the category is voiced or voiceless (Lisker and Abramson, 1964). So the fact that such a change occurs in a stimulus can be neither more nor less psychophysically inappropriate for the matched $F0$s than for the mismatched ones. Similarly, even if a high or low $F0$ might be expected after a voiceless interval on purely physical terms, we need something additional to explain the $F0$ effect found here: high $F0$s slowed responses to short VOTs, but low $F0$s slowed responses to long VOTs. Along with the results of Whalen and Samuel (1985), we have solid evidence that the processing delays caused by these phonetic mismatches cannot be psychophysical in origin.

The results also make the theory of tonogenesis more plausible. In those cases in which the loss of a voicing distinction gives rise to new tonal categories, (e.g., Hombert *et al.*, 1979; Maddieson, 1984; Abramson and Erickson, 1992), we must make two assumptions. The first is that perturbations of $F0$ of the size found in natural productions must be perceptible. The present results, along with others, make this seem likely. Also, the learning that occurred in the present study indicates that even speakers who might not themselves depend on the $F0$ differences to make linguistic distinctions would be able to learn to appreciate those differences in other speakers. A second assumption is that the $F0$ effect of voicing must be enhanced before it can begin to be used distinctively. Otherwise, the loss of the voicing distinction would, of necessity, mean the loss of the $F0$ difference. This, of course, assumes that the configuration for the presence versus absence of voicing are directly responsible for the $F0$ perturbations. While such a view seems to hold true at present (Löfqvist *et al.*, 1989), the number of languages that has been examined to date is too small to reach any firm conclusions about whether the relationship is a necessary one and/or how widespread enhancement of the perturbations might be.

The present reaction time results clearly show that even when a phonologically primary feature is unambiguously specified, the perceptual system nonetheless takes a phonologically redundant feature (or purely phonetic effect) into account. This is perhaps to be expected for a system that can make use of those redundant features, but it has more often been proposed that these features are largely ignored unless the primary feature is impaired in some way. Instead, the perceptual system seems to be making use of all the information it has, even if it is phonologically redundant.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. HD-01994 to Haskins Laboratories. Some of the results were presented at the Annual Meeting of the Linguistic Society of America, January 1991, Philadelphia, PA. We would like to thank Bruno H. Repp, Carol A. Fowler, Terrance M. Nearey, and two anonymous reviewers for helpful comments.

Abramson, A. S., and Erickson, D. M. (1992). "Tone splits and voicing shifts in Thai: phonetic plausibility," in *Pan-Asiatic Linguistics: Proceedings of the Third International Symposium on Language and Linguistics* (Chulalongkorn University, Bangkok), Vol. 1, pp. 1–15.

Abramson, A. S., and Lisker, L. (1985). "Relative power of cues: $F0$ shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. A. Fromkin (Academic, New York), pp. 25–33.

Burton, M. W., Baum, S. R., and Blumstein, S. E. (1989). "Lexical effects on the phonetic categorization of speech: the role of acoustic structure." J. Exptl. Psych.: Human Percept. Perform. **15**, 567–575.

Fujimura, O. (1971). "Remarks on stop consonants: Synthesis experiments and acoustic cues," in *Form and substance: Phonetic and Linguistic Papers presented to Eli Fischer-Jørgensen*, edited by L. L. Hammerich, R. Jakobson, and E. Zwirner (Akademisk Forlag, Copenhagen), pp. 221–232.

Ganong, W. F. (1980). "Phonetic categorization in auditory word perception," J. Exptl. Psych: Human Percept. Perform. **6**, 110–125.

Hombert, J.-M. (1975). "Towards a theory of tonogenesis: An empirical, physiologically and perceptually-based account of the development of tonal contrasts in language, Ph.D. dissertation, Univ. California, Berkeley.

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

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

Lebhiste, I., and Peterson, G. E. (1961). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–423.

Lisker, L., and Abramson, A. S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 385–422.

Löfqvist, A., Baer, T., McGarr, N., and Story, R. S. (1989). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.

McQueen, J. M. (1991). "The influence of the lexicon on phonetic categorization stimulus quality in word-final ambiguity," J. Exptl. Psych.: Human Percept. Perform. **17**, 433–443.

Maddieson, I. (1984). "The effects on $F0$ of a voicing distinction in sonorants and their implications for a theory of tonogenesis," J. Phonet. **12**, 9–15.

Martin, J. G., and Bunnell, H. T. (**1981**). "Perception of anticipatory coarticulation effects in /stri,stru/ sequences," J. Acoust. Soc. Am. **69**, S92.

Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**, 505–512.

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

Pisoni, D. B., and Tash, J. (**1974**). "Reaction times to comparisons within and across phonetic categories," Percept. Psychophys. **15**, 285–290.

Repp, B. H. (**1975**). "Dichotic masking of consonants by vowels," J. Acoust. Soc. Am. **57**, 724–735.

Shimizu, K. (**1989**). "A cross-language study of voicing contrasts of stops," Stud. Phonolog. **23**, 1–12.

Silverman, K. (**1987**). "The structure and processing of fundamental frequency contours," Ph.D. dissertation, University of Cambridge.

Tomiak, G. R., Mullennix, J. W., and Sawusch, J. R. (**1987**). "Integral processing of phonemes: Evidence for a phonetic mode of perception," J. Acoust. Soc. Am. **81**, 755–764.

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

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

Whalen, D. H. (**1991**). "Categorical, prototypical and gradient theories of speech: Reaction time data," in *Proceedings of the 12th International Congress of Phonetic Sciences* (Universite de Provence, Aix-en-Provence), Vol. 3, pp. 90–93.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1990**). "Gradient effects of fundamental frequency on stop consonant voicing judgments," Phonet. **47**, 36–49.

Whalen, D. H., and Samuel, A. S. (**1985**). "Phonetic information is integrated across intervening nonlinguistic sounds," Percept. Psychophys. **37**, 579–587.