Journal of Speech and Hearing Research, Volume 39, 1182–1190, December 1996

Role of F_0 and Amplitude in the Perception of Intervocalic Glottal Stops

James M. Hillenbrand
Western Michigan University
Kalamazoo

Robert A. Houde
RIT Research Corporation
Rochester, NY

Glottal stops that occur in vowel-consonant-vowel context are often not realized as stops at all, but rather show voicing that is continuous throughout the glottal constriction gesture. Glottal articulations that are realized in this way are apparently marked by reductions in amplitude and fundamental frequency. In the present study measurements from naturally produced utterances containing the sequence /oʔo/ (i.e., a glottal stop separating two identical vowels) were used to create a set of synthetic stimuli that varied in their F_0 and amplitude contours. The utterances were resynthesized in six ways: (a) original pitch/original amplitude, (b) original pitch/flat amplitude, (c) flat pitch/original amplitude, (d) flat pitch/flat amplitude, (e) flat pitch/inverted amplitude, and (f) inverted pitch/flat amplitude. Results indicated that: (a) a dip in the pitch contour is nearly always sufficient to cue the presence of a glottal stop *in the absence of any drop in amplitude*, (b) a dip in the amplitude contour is usually sufficient to cue the presence of a glottal stop, and (c) signals with inverted contours were not heard as glottal stops, indicating that it is not merely an abrupt change that is needed to signal a glottal stop.

KEY WORDS: glottal stop, speech perception, speech synthesis, speech acoustics, fundamental frequency

It is a common observation that phonetic contrasts tend to be cued by a number of distinct acoustic properties. A good deal of attention has been paid to understanding the perceptual mechanisms that are involved in combining multiple cues to phonetic contrasts, and these phenomena have figured heavily in theoretical discussions of speech perception (see Repp, 1982, for a comprehensive review of this literature; see also Kingston & Diehl, 1995; Nearey, 1995). For example, it is known that stop consonant voicing distinctions can be cued by a large array of apparently diverse acoustic features, including voice onset time (VOT), first formant (F_1) onset frequency, aspiration amplitude, and fundamental frequency (F_0) at voicing onset (Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981; Lisker, 1975; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1979; Stevens & Klatt, 1974; Summerfield, 1982; Summerfield & Haggard, 1977). One of the most basic questions that can be asked about perceptual integration phenomena such as these is whether speech-specific mechanisms are required to explain listeners' behavior. As Repp (1982) points out, "the large number of trading relations . . . poses formidable problems for anyone who would like to explain speech perception in purely auditory terms. Why should cues as diverse as, say, VOT and F_1 onset, or silence and fricative noise duration, trade in the way they do?" (p. 92). Repp goes on to argue that two logical avenues are open to an investigator pursuing an auditory explanation of these phenomena: (a) the *auditory integration hypothesis*, and (b) the *auditory interaction hypothesis*. The auditory integration hypothesis essentially suggests that the cues that participate in a trading relationship are not as diverse as they may seem at first. For example, Diehl and his colleagues

0022-4685/96/3906-1182

have suggested cues such as voicing during or shortly after consonant constriction, low F_1 at voicing onset, and low F_0 at voicing onset participate in a trading relation because all three cues contribute to a single integrated auditory feature called the "low frequency property" (Diehl & Kingston, 1991; Diehl & Kleunder, 1989; Kingston, 1991; Kingston & Diehl, 1994, 1995; see also Nearey, 1995; Stevens & Blumstein, 1978, 1981). Repp argues that a weakness of the auditory integration hypothesis is that it applies most naturally when the cues that participate in a trading relation are relatively brief, spectral, and occur in close temporal proximity. For example, it is unclear how an auditory integration approach would explain phenomena such as the effect on perceived place of articulation of transitions both into and out of consonant closure (Dorman & Raphael, 1980; Repp, 1978; Repp & Mann, 1981), since the required integration interval would have to be quite long. Repp concluded that in most cases "... the cues that participate in a trading relation are simply too diverse or too widely spread out to make auditory integration seem plausible" (p. 93).

The auditory interaction approach is best explained with an example. Miller et al. (1976) and Pisoni (1977) offered psychophysical accounts of the perception of initial-stop voicing contrasts based on the discriminability of differences in the temporal order of acoustic events occurring near articulatory release. Pisoni argued that the psychophysical task in VOT perception can be viewed as judging the onset of high-frequency energy associated with articulatory release relative to the onset of low-frequency energy associated with voicing onset. However, it is known that the VOT boundary shifts to longer relative onset times as F_1 decreases in frequency (e.g., Lisker, 1975; Lisker et al., 1977). An auditory interaction hypothesis might be offered to explain this effect by suggesting that, for unknown reasons, judgments of temporal order are poorer for stimuli with low-frequency F_1 values or, more generally, when the high- and low-frequency stimulus components are more widely separated in frequency. If this is the case, it ought to be possible to demonstrate such an effect with nonspeech continua varying in relative onset time. Several studies have been conducted using "tone-onset time" (TOT) nonspeech analogs consisting of two pure tones varying in relative onset time. Results have been mixed: Although Parker (1988) found evidence that TOT boundaries shifted to longer relative onset times as the frequency of the lower tone was decreased, negative findings were reported by Hillenbrand (1984), Hirsh (1959), and Summerfield (1982).

Repp (1982) concluded that the great majority of cue-integration phenomena cannot be explained in purely auditory terms. Repp's view is by no means universally shared, but if one accepts his conclusion for the sake of argument, there would seem to be two quite different speech-specific accounts of these effects. The view favored by gesturalists is that these effects reflect the operation of a specialized recognition module that interprets the speech signal in terms of articulatory gestures: "... trading relations may occur because listeners perceive speech in terms of the underlying articulation and resolve inconsistencies in the acoustic information by perceiving the most plausible articulatory act" (Repp, 1982, p. 95). However, it is also possible that these effects simply reflect learned associations among acoustic features that tend to covary in natural speech, as Nearey has argued (Nearey, 1991, 1992, 1995). Differentiating between these two very different speech-specific views has not proven to be a simple matter.

The purpose of the present study was to gain additional information about perceptual integration phenomena by determining the relative contributions of amplitude and fundamental frequency changes in cueing the presence of intervocalic glottal stops. Although glottal stops do not have phonemic status in English, they occur with some regularity. For example, /t/ is often realized as a glottal stop in words such as "button" and "kitten," and in phrases such as "rat poison." The glottal stop is also an optional but very common marker of word juncture in phrases such as "known ocean" versus "no notion" and "she eats" versus "sheets" (e.g., Huber, 1988; Kiessling, Kompe, Niemann, Noth, & Batliner, 1993; Nakatani & Dukes, 1977). Glottal stops are also commonly used in "nonlexical intonation signals," as Luthy (1983) calls them, occurring in utterances such as /ˈˀʌˀʌ/ or /ˈˀmˀm/ ("no") and /ˈʌˀo/ ("oops!").

Although we will follow the convention of using the term "glottal stop," many investigators have observed that glottal stops are rarely produced with complete glottal closure (e.g., Fischer-Jorgensen, 1989; McCall, Dingwall, Ramkumar, & Thomas, 1993; Pierrehumbert and Talkin, 1992; Priestly, 1976). Glottal stops that are realized without complete closure appear to be marked by some combination of a reduction in amplitude, a drop in fundamental frequency and, in some cases, irregular glottal vibrations that are associated with glottal constriction (Fischer-Jorgensen, 1989; McCall et al., 1993; Pierrehumbert & Talkin, 1992). Also, although the feature has not been measured in naturally produced glottal stops, a change in spectral tilt also seems likely (e.g., Gobl, 1989; Kiessling et al., 1993). Figure 1 shows an example of a continuously voiced glottal stop. The glottal gesture appears to be marked by dips in the amplitude and F_0 contours.

In terms of the underlying physiology, the amplitude dip appears to be a direct result of the glottal constriction, which has the effect of reducing glottal airflow (Fischer-Jorgensen, 1989). The physiological mechanisms underlying the drop in F_0 are less clear. Fischer-Jorgensen's extensive investigation of the Danish stød was inconclusive on this point, but there was some evidence from stroboscopic videos and electromyographic recordings that the glottal constriction was associated with shorter and less tense vocal folds (see also Ludlow, Sedory, & Fujita, 1991; McCall et al., 1993; Priestly, 1976; Riber Peterson, 1973).

Very little is known about the acoustic cues that signal the presence of glottal stops, particularly those that are produced with continuous voicing. The purpose of the present study was to investigate the relative contributions of amplitude and F_0 changes in the perception of continuously voiced glottal stops in vowel-consonant-vowel context. This was done by asking listeners to identify synthetically generated utterances in which specific aspects of the F_0 and/or amplitude contour were manipulated.

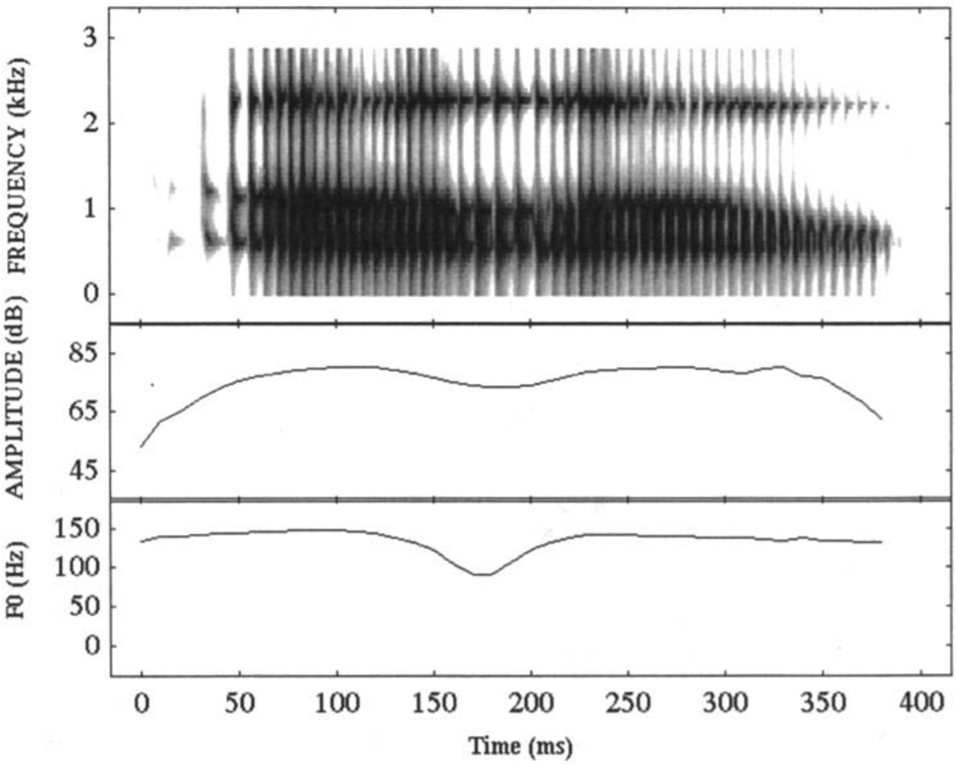

FIGURE 1. Oscillogram, F_0 contour, and amplitude contour for a naturally produced "glottal stop" that shows continuous voicing. The utterance is "oh-oh" (/oˀo/).

EXPERIMENT 1: PHONETIC LABELING

Methods

Synthesis

Utterances were synthesized in various ways based on acoustic measurements made from instances of the sequence /oˀo/ excised from the TI-DIGITS connected digits database, which consists of digit strings (e.g., "three-two-seven-oh-oh") spoken by a dialectically heterogeneous group of men, women, and children (Leonard, 1984). The utterances used in the present study were produced by 10 male talkers. All 10 utterances that were chosen were continuously voiced, and none showed evidence of irregular glottal vibrations. Measurements were made of rms intensity and F_0 every 10 ms using an autocorrelation pitch tracker (Hillenbrand, 1988). A custom interactive editor (Hillenbrand, Getty, Clark, & Wheeler, 1995) was used to hand edit the F_0 contours. The Klatt synthesizer (Klatt and Klatt, 1990) was used to generate six versions of each utterance: (a) original pitch/original amplitude, (b) original pitch/flat amplitude, (c) flat pitch/original amplitude, (d) flat pitch/flat amplitude, (e) flat pitch/inverted amplitude, and (f) inverted pitch/flat amplitude. Sample amplitude and F_0 contours for the various conditions are shown in Figure 2. In the original amplitude conditions, the AV (amplitude of voicing) parameter was driven by measured values of rms intensity. Similarly, in the original pitch conditions, the F_0 parameter was driven by measured values of fundamental frequency. In the flat amplitude conditions, the dip in the amplitude contour was removed by interpolating a straight line across the amplitude dip. The same procedure was used to remove dips in the F_0 contour. Table 1 shows the approximate extent of the F_0 and amplitude dips measured from the naturally produced "oh-oh" utterances. The values for extent of dip were calculated as the difference between the minimum value in the dip and average value at the "shoulders" of the dip. The values in Table 1 should be considered approximate since the start and end of the dip could not be determined very precisely given the rounded shoulders that were normally seen (see Figure 1). The inverted amplitude and pitch conditions were generated with "convex" contours, that is, with peaks in the contour of the same shape and magnitude as the original dips. The inverted conditions were used to determine whether glottal stops are heard whenever there is an abrupt change in the amplitude or pitch contour or, alternatively, whether it is specifically a dip that is required.

The vowel formants for all utterances were held constant at /ɑ/ (F_1 = 720, F_2 = 1240, F_3 = 2400, F_4 = 3300, F_5 = 3700). (The vowel formants were set to /ɑ/ instead of the original /o/ because we found it difficult to produce a convincing /o/ with stationary formants.) The synthesizer was run in cascade mode and all signals were synthesized with a 10 kHz sample rate.

Listening Test

Listeners consisted of 20 undergraduate and graduate students in the Speech Pathology and Audiology Depart-

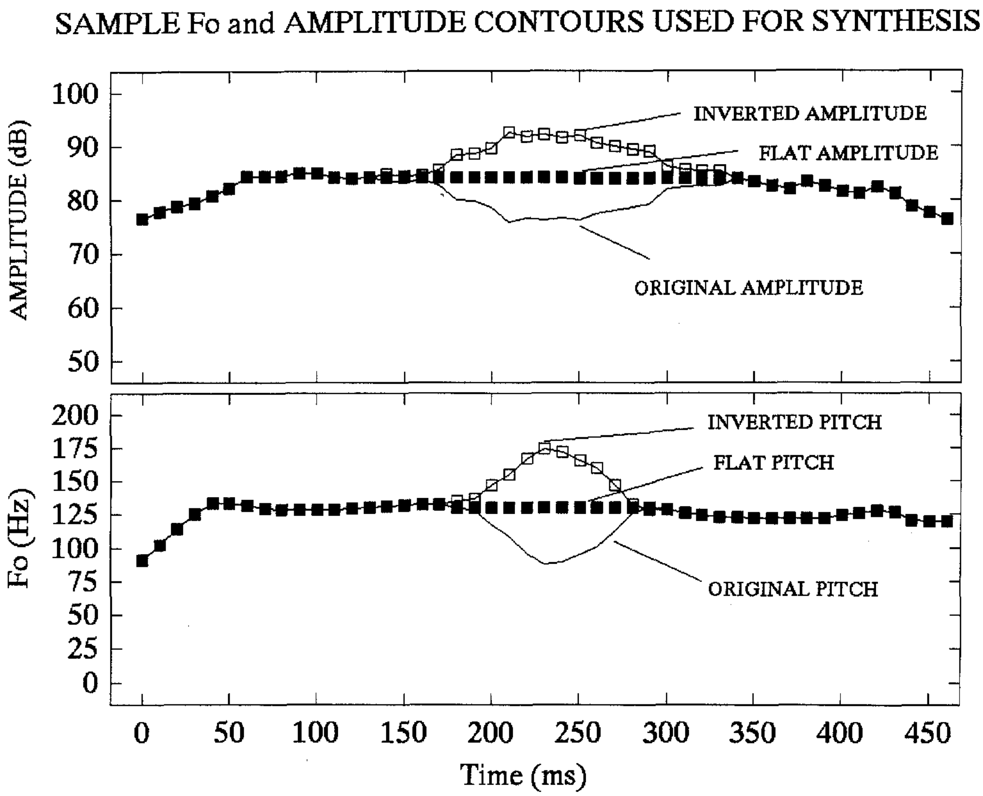

FIGURE 2. Sample F_0 and amplitude contours used to generate the synthetic signals used in Experiment 1.

ment at Western Michigan University. Subjects were audiometrically screened at 25 dB at 250, 500, 1000, 2000, and 4000 Hz and had no reported speech or voice disorder. The listeners for this and all subsequent experiments had taken an undergraduate course in phonetics and understood what was meant by the term "glottal stop." However, none of the subjects were experienced phoneticians. The signals were presented in a pseudorandom order consisting of five separate scramblings of the 60 signals (10 talkers × 6 synthesis conditions). The signals were lowpass filtered at 4.3 kHz, amplified, and delivered at approximately 75 dBA over a single loudspeaker positioned approximately 1 m from the listener's head (Boston Acoustics A60). Subjects were tested individually in a quiet room. The listeners were asked to press a key on a computer terminal to indicate whether or not each utterance contained a glottal stop. Listeners were told that all of the signals were created with a speech synthesizer.

TABLE 1. Approximate magnitude of the amplitude and F_0 dips, measured as the difference between the minimum value in the dip and the average value at the "shoulders" of the dip. F_0 values are given both in Hz and as a percentage of the average F_0 value at the shoulders of the dip.

Stimulus	Amplitude dip (dB)	F_0 Dip	
		Hz	Percent
1	17.9	65	56.3
2	6.1	62	42.6
3	10.0	46	28.0
4	13.4	86	58.9
5	16.8	95	60.1
6	7.0	43	32.3
7	10.1	41	46.0
8	12.6	85	53.1
9	6.3	51	50.2
10	5.5	17	17.8

Results and Discussion

Results of the listening test are shown in Figure 3. The very high glottal stop identification rate for the original pitch/original amplitude condition, combined with the very low glottal stop identification rate for the flat pitch/flat amplitude condition, indicates clearly that some combination of F_0 and amplitude is sufficient to cue the glottal stop. It can also be seen that the F_0 cue alone is nearly always (93%) sufficient to cue the glottal stop. The F_0 cue alone seems to be a more potent cue than the amplitude cue alone, which signaled a glottal stop on 66% of the trials. A one-way repeated measures analysis of variance on the labeling data showed a highly significant effect for stimulus type, $F(5, 95) = 448.8$, $p < 0.001$. Newman-Keuls post hoc tests, using an alpha level of 0.05, showed: (a) that the original pitch/original amplitude signals were identified as glottal stops more frequently than either of the single-cue signals, and (b) that the signals that were cued by F_0 alone were identified as glottal stops more frequently than the signals that were cued by amplitude alone.

The inverted pitch and inverted amplitude stimuli rarely

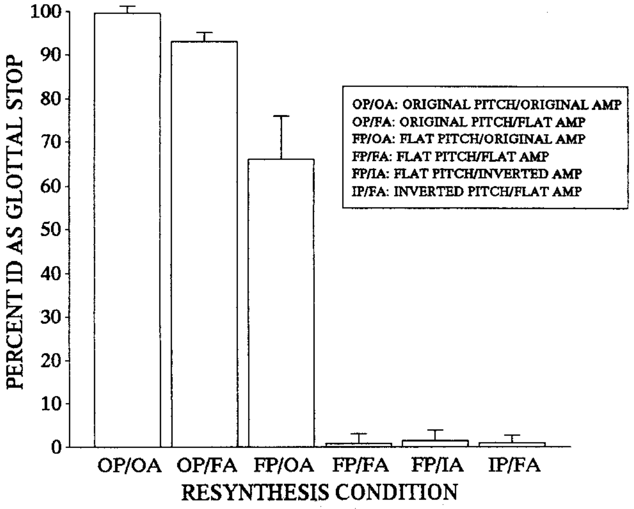

FIGURE 3. Percentage of tokens heard as glottal stops for the six synthesis conditions.

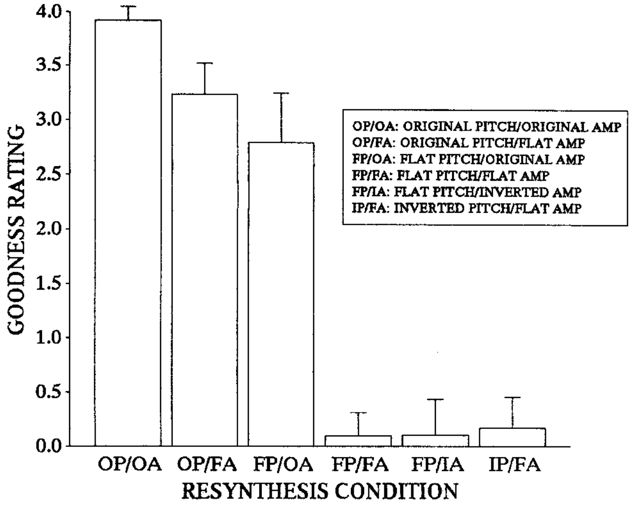

FIGURE 4. Average goodness ratings for the six synthesis conditions.

signaled a glottal stop. The inverted-cue findings indicate that it is not merely an abrupt change that is needed, but specifically a dip in F_0 or amplitude. The inverted-cue findings also indicate that subjects are not fooled into reporting a glottal stop for any utterance that shows two distinct beats since this impression is created by the inverted-cue signals as well.

EXPERIMENT 2: GOODNESS JUDGMENTS

Methods

The forced-choice labeling procedure used in Experiment 1 does not retain any graded information about the quality of the stimuli. It is possible, for example, that one or both of these signal manipulations results in signals that are accepted by listeners as glottal stops in a forced-choice task, but which might be judged to be poor examples of the category. In Experiment 2 a second group of listeners was asked to provide goodness ratings for the same set of stimuli. The subjects consisted of 15 phonetically trained listeners, 4 of whom had participated in Experiment 1. Listeners heard five randomly ordered presentations of the 60 signals using the instrumentation, subject selection criteria, and randomization scheme described previously. On each trial, subjects were asked to enter a number between zero and 5, with zero indicating no glottal stop, 1 indicating a "marginal" example of glottal stop, and 5 indicating a "very good" example of a glottal stop.

Results and Discussion

Average goodness ratings for the six types of stimuli are shown in Figure 4. The overall pattern of results is quite similar to the labeling data. The highest ratings tended to be given to the signals cued by both F_0 and amplitude, and the signals cued by fundamental frequency alone received somewhat higher ratings than the signals cued by amplitude alone. A one-way repeated measures analysis of variance on the goodness ratings showed a highly significant effect for stimulus type, $F(5, 70) = 543.8$, $p < 0.001$. Newman-Keuls post hoc tests showed: (a) significantly higher ratings for the original pitch/original amplitude signals than for either of the single-cue signals, and (b) significantly higher ratings for the signals that were cued by F_0 alone than for the signals that were cued by amplitude alone.

EXPERIMENT 3: PERCEPTION OF SYNTHETIC CONTINUA

The signals for Experiments 1 and 2 were generated using amplitude and F_0 contours that were measured from naturally produced utterances. The purpose of Experiment 3 was to measure phonetic labeling functions using synthetic continua varying systematically in the magnitude of amplitude and F_0 dips.

Methods

The signals for Experiment 3 consisted of one 16-step continuum varying in the magnitude of an amplitude dip and a second 16-step continuum varying in the magnitude of a fundamental frequency dip. The synthesis parameters for the two continua are shown in Figure 5. The amplitude dips began at 100 ms with a linear (in dB) drop in voicing amplitude (AV) from 100 to 130 ms. The AV parameter was held constant from 130 to 170 ms, then increased linearly from 170 to 200 ms. The magnitude of the amplitude dip was varied from 0 to 30 dB in 2 dB steps. Fundamental frequency for all signals on the amplitude continuum was held constant at 130 Hz from 0 to 230 ms, then decreased linearly to 100 Hz from 230 ms to 400 ms. An analogous set of signals varying in the magnitude of an F_0 dip was created. F_0 at the start and end of the dip was set to 130 Hz, and the magnitude of the F_0 dip varied from 0 to 30 Hz in 2 Hz steps. For all signals F_0 decreased linearly from 130 Hz to 100 Hz

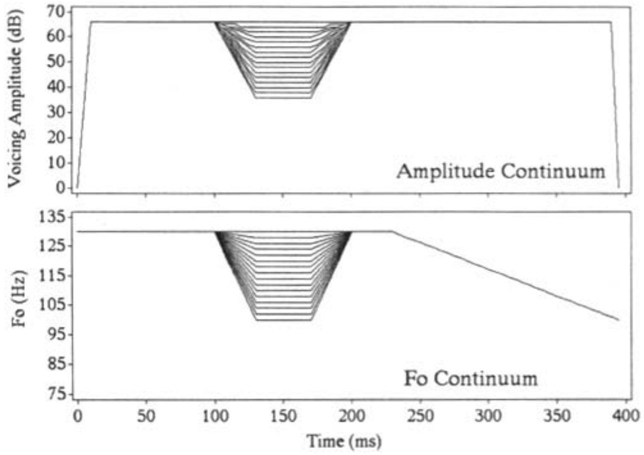

FIGURE 5. Amplitude and F_0 contours used in Experiment 3.

between 230 to 400 ms. As in Experiments 1 and 2, the vowel formants for all utterances were held constant at /ɑ/ (F_1 = 720, F_2 = 1240, F_3 = 2400, F_4 = 3300, F_5 = 3700), and the synthesizer was run in cascade mode with a 10 kHz sample rate.

The signals were presented for identification to a new group of 16 listeners with training in phonetic transcription. The signals were presented in a pseudorandom order consisting of 10 separate scramblings of the 32 stimuli (2 continua × 16 steps). Instrumentation and procedures were identical to Experiment 1. On each trial the listeners were asked to press a key on a computer terminal to indicate whether or not the utterance contained a glottal stop.

Results and Discussion

Group labeling results for Experiment 3 are shown in Figure 6. Panel a shows the labeling function for the F_0 continuum and panel b shows the labeling function for the amplitude continuum. The phonetic boundaries, calculated by linear interpolation of the 50% point on the group identification functions, are 9.4 Hz for the F_0 continuum and 8.0 dB for the amplitude continuum. The simplicity of these group functions belies a good deal of intersubject variability. Labeling functions for the individual subjects are shown in panels c and d of Figure 6. The intersubject variability is consistent with our own subjective impressions of these stimuli. Our informal impression is that these continua do not show abrupt changes. For example, as F_0 or amplitude dips become increasingly shallow, our impression is that the glottal stop does not suddenly disappear, but rather sounds as if the glottal gesture is being produced with less force. We believe that the subject variability reflects subject-to-subject differences in decision criteria about when the percept is so weak that a "no glottal stop" judgment is warranted.

EXPERIMENT 4: EFFECTS OF COMBINING F_0 AND AMPLITUDE CUES

Pilot results suggested that the effects of amplitude and F_0 dips were additive; for example, an amplitude dip of a certain magnitude was more likely to be perceived as a glottal stop if combined with even a very small F_0 dip. The purpose of Experiment 4 was to determine the effects of

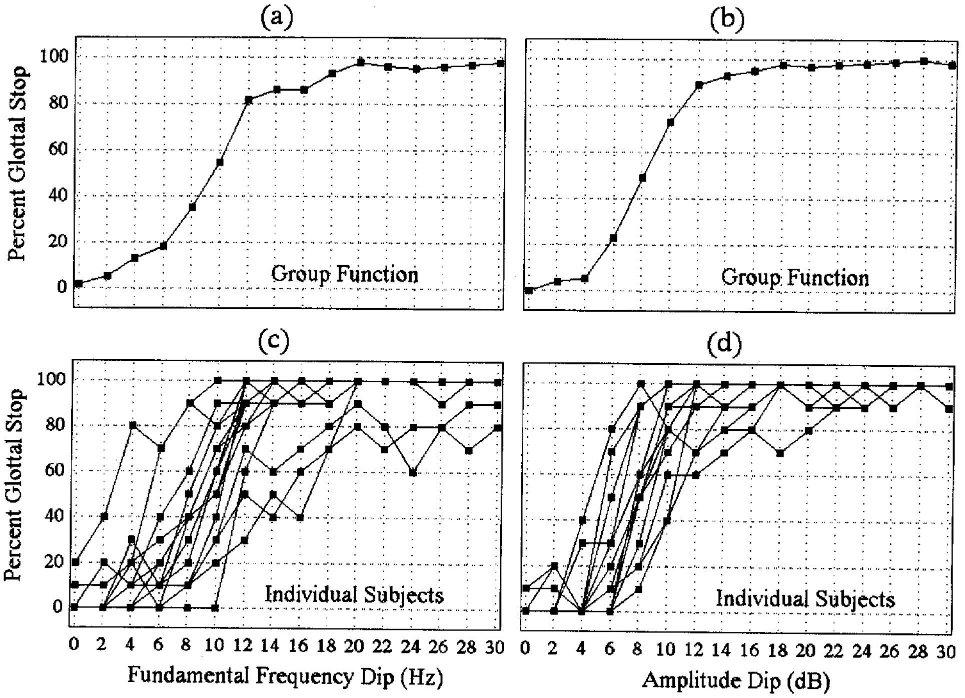

FIGURE 6. Phonetic labeling functions for Experiment 3: (a) group identification functions for a continuum varying in the magnitude of a fundamental frequency dip, (b) group identification functions for a continuum varying in the magnitude of an amplitude dip, (c) identification functions for 16 individual subjects for the fundamental frequency continuum, and (d) identification functions for 16 individual subjects for the amplitude continuum.

combining F_0 and amplitude cues on the perception of intervocalic glottal stops.

Methods

Four 11-step continua were generated varying in the magnitude of an amplitude dip from 0 to 30 dB in 3 dB steps. The four continua were generated with F_0 dips of 0, 1, 2, or 3 Hz, following the parameter-contour shapes shown in Figure 5. Four additional 11-step continua were generated varying in the magnitude of the F_0 dip from 0 to 30 Hz in 3 Hz steps. The four continua were generated with amplitude dips of 0, 1, 2, or 3 dB, again using the contour shapes shown in Figure 5. The signals were presented for identification to the same 16 listeners who participated in Experiment 3. The signals were presented in a pseudorandom order consisting of 10 separate scramblings of the 88 stimuli (8 continua × 11 steps). Procedures and instrumentation were identical to those of Experiment 3.

Results

Figure 7 shows the labeling results for the four F_0 continua (top panel) and the four amplitude continua (bottom panel). It can be seen that the F_0 functions shift in the direction of shallower F_0 dips when increasingly large amplitude dips are combined with the F_0 cues, indicating that the two cues combine in some way to increase the probability of a glottal stop percept. Similarly, the amplitude functions shift to the left when increasingly large F_0 dips are combined with the amplitude cues. Phonetic boundaries were calculated by linear interpolation of the 50% point on the labeling functions. The phonetic boundary shifts are significant for both continua, F_0 Continua: $F(3, 45) = 56.2$, $p < 0.001$; Amplitude Continua: $F(3, 45) = 6.0$, $p < 0.01$.

Notice that with the exception of the difference in step size, the 0 dB F_0 continuum from Experiment 4 is identical to the F_0 continuum that was used in Experiment 3. Similarly, the 0 Hz amplitude continuum is identical to the amplitude continuum that was used in Experiment 3, except that a slightly larger step size was used. Comparing Figures 6 and 7, it can be seen that the labeling functions for the two single-cue F_0 continua are, in fact, quite similar (compare panel a of Figure 6 with the filled squares from panel a of Figure 7). However, the labeling functions from the two single-cue amplitude continua are quite different (compare panel b of Figure 6 with the filled squares from panel b of Figure 7). In Experiment 4, which included many signals with dips in both F_0 and amplitude, the labeling function for the 0 Hz amplitude continuum is considerably shallower than the corresponding function from Experiment 3 and remains well below 100% even for the largest amplitude dips. Listeners apparently shifted their decision criteria and were less willing to identify glottal stops cued by amplitude alone when the signals were presented in a listening session that included stimuli that were cued by both amplitude and F_0. It is not clear why a similar shift did not occur for the F_0 series.

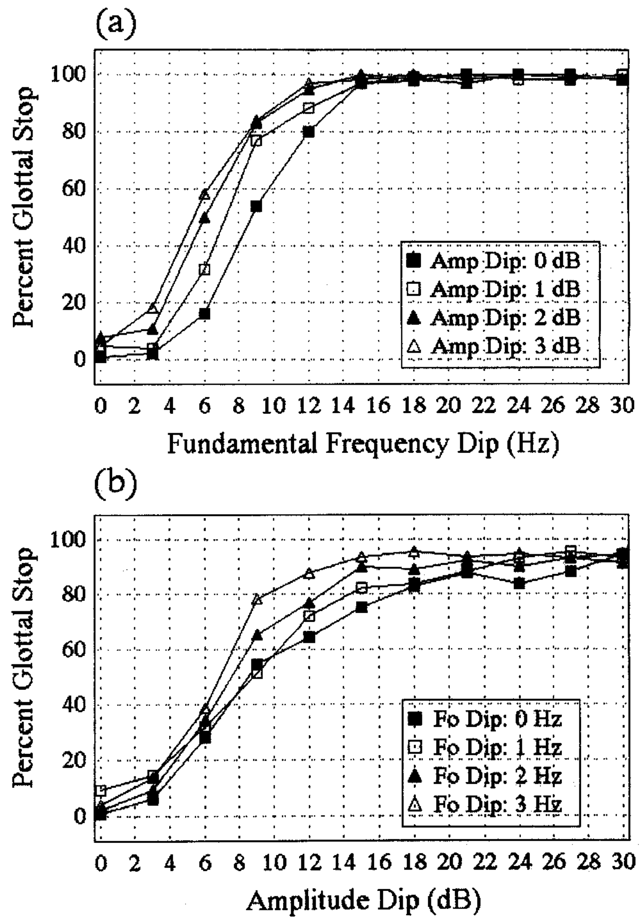

FIGURE 7. Phonetic labeling functions for Experiment 4. Panel a shows group identification functions for four continua varying in the magnitude of a fundamental frequency dip; the parameter is the magnitude of an amplitude dip. Panel b shows group identification functions for four continua varying in the magnitude of an amplitude dip; the parameter is the magnitude of an F_0 dip.

General Discussion

To summarize briefly, the experiments with utterances resynthesized from natural speech indicated that: (a) a dip in the F_0 contour is nearly always sufficient to cue the presence of a glottal stop *in the absence of any drop in amplitude*, (b) a dip in the amplitude contour is usually sufficient to cue the presence of a glottal stop, although the amplitude cue is less potent than the F_0 cue, and (c) it is not merely an abrupt change that is needed to signal a glottal stop, but specifically a dip in F_0 or amplitude. Studies with synthetically generated continua indicated: (a) the labeling boundary for a continuum varying in the magnitude of an amplitude dip occurred at approximately 8 dB, (b) the labeling boundary for a continuum varying in the magnitude of an F_0 dip occurred at approximately 9 Hz, and (c) the effects of amplitude and F_0 dips appear to be additive.

Our results are largely consistent with a very recent synthesis study by Pierrehumbert and Frisch (in press), which showed that a sharp drop in F_0 was easily the most important cue in signaling glottalization in utterances such as, "heavy oak," "steady awning," and "twenty ears." However, informal listening suggested that the most natural-sounding utterances resulted from a synthesis method that faithfully modeled the drop in F_0, the drop in amplitude, and a flattening of the source spectrum. One point of disagreement is that Pierrehumbert and Frisch concluded, again on the basis of informal listening, that a convincing glottal stop could not be signaled by an amplitude drop alone. The discrepancy may be related to differences in temporal factors, which were not examined in either study. It is very likely that variables such as the slope or duration of F_0 or amplitude changes and the duration of the dips play a role, and it would not be surprising to find that these temporal features differed between the two studies, especially in light of the differences in speech material. In fact, it is virtually certain that temporal features will play some role in the perception of glottalization since F_0 or amplitude changes that occur over relatively long intervals will clearly be assigned by the listener to intonation and/or stress rather than a glottal gesture. It remains to be determined exactly what temporal cues cause listeners to assign F_0 and amplitude changes to a segmental event versus intonation and rhythm.

One of the main findings of this study is that dips in either amplitude or F_0 were effective in cueing a glottal stop percept, although not equally effective. The role played by F_0 and amplitude in cueing glottal stops has something in common with trading relations phenomena that have been widely reported in the literature (e.g., Kingston & Diehl, 1995; Nearey, 1995; Repp, 1982), although there are important differences as well. In the present case, each of two quite different acoustic cues is sufficient by itself to cue a glottal stop, and their effects are additive. However, we do not mean to suggest that F_0- and amplitude-cued glottal stops sound identical to one another. The significant differences that were observed in both labeling and goodness ratings between the F_0-cued signals and the amplitude-cued signals indicate that these two kinds of signals are not, in fact, identical, which agrees with our own informal impressions. This clearly differs from trading relations such as "say-stay" in which listeners have difficulty discriminating signals that are cued by different combinations of closure duration versus formant transition onsets (Fitch, Halwes, Erickson, & Liberman, 1980).

A significant challenge in interpreting these results is to explain the effectiveness of either F_0 or amplitude in cueing the presence of a glottal stop. Although it cannot be absolutely ruled out, a low-level auditory explanation for this relationship does not immediately present itself since it seems implausible that amplitude and F_0 dips are equivalent to one another at an early stage of auditory analysis. If the effectiveness of either F_0 and amplitude in cueing the presence of a glottal stop is not explained at a low auditory level, then what does account for this phenomenon? Motor Theory advocates might argue that both the drop in amplitude and the drop in F_0 serve as interchangeable cues indicating a glottal constriction. Similarly, it might be argued that the perceived similarity between true and continuously voiced glottal stops is accounted for by the similarity in the underlying articulatory gesture. Stroboscopic videos that we made of three talkers producing both true and continuously voiced glottal stops suggest that the gestures are, in fact, quite similar to one another. Based strictly on observation of the videos, the two gestures both appear to involve glottal constriction, with the continuously voiced version appearing to be a less forceful version of the true glottal stop (see also Fischer-Jorgensen, 1989).

It might also be argued that the effectiveness of either F_0 and amplitude in cueing a glottal stop is acquired as a result of a learned association between two cues that are highly correlated in natural speech. It is possible, for example, that listeners are initially sensitive only to the drop in amplitude that is common to all stop consonants and only later acquire sensitivity to the drop in F_0 that is produced as a side effect of the glottal constriction gesture. This kind of learning lies at the heart of Nearey's "double-weak" theory of speech perception. Developmental evidence would be especially interesting since a key aspect of this model is that there must be some age at which listeners do not show sensitivity to this kind of cue variation.

It is obvious that there are many aspects of the perception of this little-studied sound that are not fully understood. A systematic measurement study would be useful in clarifying the degree to which the acoustic cues to glottal stops covary and some attention to the role of temporal factors is warranted. Further, as Fischer-Jorgensen's (1989) study makes clear, there is a great deal about both the physiology and acoustics of glottal constrictions that remains poorly understood.

Acknowledgments

This work was supported by a research grant from the National Institutes of Health (NIDCD 1-R01-DC01661). We are grateful to Michael Clark and Robert Erickson for comments on previous drafts and for many hours of critical listening, and to Terry Nearey for thoughtful comments on the interpretation of the findings. We are also grateful to Bruce Smith and three anonymous reviewers for helpful comments on a previous draft. An earlier version of this paper was presented at the June, 1994, meeting of the Acoustical Society of America, Cambridge, Massachusetts.

References

Diehl, R. L., & Kingston, J. (1991). Phonetic covariation as auditory enhancement: The case of the [+voice]/ [−voice] distinction. In O. Engstrand & C. Kylander (Eds.), *Current phonetic research paradigms: Implications for speech motor control* (pp. 139–143). Stockholm: University of Stockholm.

Diehl, R. L., & Kleunder, K. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.

Dorman, M. F., & Raphael, L. J. (1980). Distribution of acoustic cues to stop consonant place of articulation in VCV syllables. *Journal of the Acoustical Society of America, 67,* 1333–1335.

Fischer-Jorgensen, E. (1989). Phonetic analysis of the stød in Standard Danish. *Phonetica, 46,* 1–59.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics, 27,* 343–350.

Gobl, C. (1989). A preliminary study of acoustic voice quality correlates. *Speech Transmission Laboratory Quarterly Status and Progress Report, 4,* 9–27.

Haggard, M. P., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47,* 613–617.

Haggard, M. P., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced-voiceless distinction. *Journal of Phonetics, 9,* 49–62.

Hillenbrand, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America, 75,* 231–240.

Hillenbrand, J. (1988). MPITCH: An autocorrelation fundamental-frequency tracker. [Computer software]. Kalamazoo, MI: Western Michigan University.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31,* 759–767.

Huber, D. (1988). *Aspects of the communicative function of voice in text intonation.* Unpublished doctoral dissertation, Chalmers University, Goteborg/Lund.

Kiessling, A., Kompe, R., Niemann, H., Noth, E., & Batliner, A. (1993). Voice source state as a source of information in speech recognition: Detection of laryngealizations. In A. J. Rubio Ayuso & J. M. Lopez Soler (Eds.), *Proceedings of the NATO ASI Conference: New Advances in Speech Recognition and Coding* (pp. 329–332). Berlin: Springer-Verlag.

Kingston, J. (1991). Integrating articulations in the perception of vowel height. *Phonetica, 48,* 149–179.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language, 70,* 419–454.

Kingston, J., & Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. In R. Connell & A. Avanti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 7–27). Cambridge, England: Cambridge University Press.

Klatt, D. H., & Klatt L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87,* 820–857.

Leonard, R. (1984). A database for speaker-independent digit recognition. *IEEE-ICASSP,* 42.11.1–42.11.5.

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America, 57,* 1547–1551.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech, 20,* 209–216.

Ludlow, C. L., Sedory, S. E., & Fujita, M. (1991). Neurophysiological control of vocal fold adduction and abduction for phonation onset and offset during speech. In J. Gauffin & R. Hammarberg (Eds.), *Vocal fold physiology* (pp. 197–205). San Diego: Singular Press.

Luthy, M. J. (1983). Nonnative speakers' perceptions of English 'nonlexical' intonation signals. *Language Learning, 33,* 19–36.

McCall, G. N., Dingwall. W. O., Ramkumar, U., & Thomas, B. A. (1993). *Mechanisms of voicing control revealed in vocal signals.* Paper presented at the American Speech-Language-Hearing Association Annual Convention.

Miller, J. D., Weir, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times. *Journal of the Acoustical Society of America, 60,* 410–417.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62,* 714–719.

Nearey, T. M. (1991). Perception: Automatic and cognitive processes. *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence, 1,* 40–49.

Nearey, T. M. (1992). Context effects in a double-weak theory of speech perception. *Language and Speech, 35,* 153–172.

Nearey, T. M. (1995). A double-weak view of trading relations: Comments on Kingston and Diehl. In R. Connell & A. Avanti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 28–40). Cambridge, England: Cambridge University Press.

Parker, E. M. (1988). Auditory constraints on the perception of stop voicing: The influence of the lower-tone frequency on judgments of tone-onset simultaneity. *Journal of the Acoustical Society of America, 83,* 1597–1607.

Pierrehumbert, J., & Frisch, S. (in press). Synthesizing allophonic glottalization. In J. van Santon, R. Sproat, J. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis.* New York: Springer-Verlag.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty & R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–119). Cambridge, England: Cambridge University Press.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America, 61,* 1352–1361.

Priestly, T. M. S. (1976). A note on the glottal stop. *Phonetica, 33,* 268–274.

Repp, B. H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics, 24,* 471–485.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America, 69,* 1154–1163.

Riber Peterson, P. (1973). An instrumental investigation of the Danish stød. *Annual Report of the Institute of Phonetics, University of Copenhagen, 7,* 195–234.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64,* 1358–1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America, 55,* 653–659.

Summerfield, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America, 72,* 51–61.

Summerfield, Q., & Haggard, M. P. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 72, 62,* 435–448.

Received January 18, 1996
Accepted July 16, 1996

Contact Author: James M. Hillenbrand, Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, MI 49008. Email: james.hillenbrand@wmich.edu