THE ROLE OF F_1 TRANSITIONS IN THE PERCEPTION OF VOICING IN INITIAL PLOSIVES

Peter HOWELL, Stuart ROSEN, Harriet LANG and Stevie SACKIN

SPEECH, HEARING AND LANGUAGE: WORK IN PROGRESS
U.C.L. No 6 (1992)

THE ROLE OF F_1 TRANSITIONS IN THE PERCEPTION OF VOICING IN INITIAL PLOSIVES

Peter HOWELL*, Stuart ROSEN, Harriet LANG and Stevie SACKIN*

Abstract

There is a long-standing controversy regarding the extent to which the perception of voicing in initial plosives depends upon voice onset time (VOT) or aspects of the first formant (F_1). Two studies bearing on this issue are reported. First, in an experiment using synthetic speech presented to adults, we have shown that changes in F_1 transition duration do not affect judgments of voicing. A previous study which shows such an effect (Stevens & Klatt, 1974) appears to do so on the basis of correlated changes in F_2 transitions. Second, using both natural and edited stimuli presented to children (with a median age of about 4 years) and adults, we have found that VOT is on average the dominant cue, with the developmental trend towards increasing reliance on it (contrary to the findings of Simon & Fourcin, 1978, using synthetic stimuli). We conclude that the role of F_1 is weak in comparison to VOT, at least for adults.

Introduction

Differences in the F_1 transition between initial voiced and voiceless plosives result from differences in the timing of changes in excitation. Aperiodic aspiration noise has its energy weighted to the high frequencies in comparison with voiced excitation. Thus, when the interval from plosive release to onset of periodicity (voice onset time or VOT) is long this transition takes place while excited by aspiration only, and is therefore relatively weak. When voicing begins immediately after the release of the stop, however, F_1 is strongly excited by the low frequency energy in laryngeal vibration, resulting in a clearly defined transition. Similarly, because the F_1 transition takes place over a duration similar to the VOT in voiceless plosives, the F_1 at voicing onset is at higher frequencies for voiceless plosives than for voiced ones (Summerfield and Haggard, 1977).

Stevens and Klatt (1974) claimed that listeners are strongly influenced by the amount of voiced transition that occurs in a plosive when judging whether it is voiced or voiceless. In their study, listeners were presented with synthetic continua varying in F_1 and F_2 transition rate as well as VOT. Although there were important individual differences, overall the presence of a significant transition in F_1 after voicing onset led to increased judgments of voicing.

Stevens and Klatt (1974) went on to point out that a mechanism based on F_1 transition detection could account for why the voiced/voiceless phoneme boundary

* Members of the Psychology Department, University College London.

differs across place of articulation (the place effect). If it is accepted that speakers are trying to signal the voiced/voiceless distinction via the presence or absence of an F_1 transition, then produced VOTs must lengthen as formant transitions do in order that the transitions can be completed before voicing starts. Stevens and Klatt (1974) note that transition rate is fastest for the lips, next fastest for the tongue tip and slowest for the tongue body. Thus, the boundaries of velar stimuli should occur at longer values of VOT than those of alveolar or bilabial stimuli.

The importance of an F_1 transition as a cue to voicing has been the subject of some dispute. Lisker (1975), for example, defended the argument for VOT as the dominant cue to voicing in adult listeners. He presented evidence that a voiced F_1 transition is neither necessary nor sufficient for the perception of a voiced plosive, nor is its absence necessary for the perception of a voiceless plosive, providing the VOT values are appropriately set. In particular, a long enough VOT is always labelled voiceless, regardless of F_1.

However, another explanation for the variation of the voiced/voiceless boundary for different F_1 transitions is possible for Stevens and Klatt's experiment. In their stimuli, both F_1 and F_2 transition rate were varied together, so it may well be that the changes in phoneme boundary were caused by changes in F_2.

Studies on the development of speech perception also lead to differing views about the importance of F_1 transitions (or associated features) for the perception of voicing in plosives. Miller and Eimas (1983) tested the discrimination of four-month-old infants on two "ta"-"da" series in one of which the formant transitions were 25 ms and in the other 85 ms. In this study, the F_1 onset frequency co-varied with VOT, as is typically the case with synthetic VOT continua. For the 25-ms transition series, the infants were able to distinguish between VOTs of 5 and 30 ms, but not of 40 and 55 ms, whereas for the 85-ms transition stimuli, the reverse occurred. One interpretation of this is that the infants are showing a change in the region of heightened discrimination on VOT continua as a function of the onset frequency of F_1, analogous to the way in which the adult phoneme boundary shifts.

Simon and Fourcin (1978), on the other hand, claimed that sensitivity to F_1 is learnt during childhood and is dependent on the specific linguistic environment. They asked French and English children to label a set of synthetic speech sounds which varied both in VOT and in the presence or absence of an F_1 transition. For VOTs less than 30 ms, F_1 was either rising (as in natural speech) or level. For VOTs above 30 ms, F_1 was always level. Only the English children learnt to make use of the F_1 transition as a cue to voicing, with most eight-year olds only consistently giving stimuli a voiced label if a rising F_1 transition was present (described as a typically adult performance). However, closer study of the English results reveals that at four years old, the children's responses were more similar to those of teenagers, in that the presence of an F_1 transition had a marked effect.

Given the disagreement about the relative importance of VOT and F_1 transitions in distinguishing voiced from voiceless plosives, two new studies were performed.

The role of F_1 transition rates in the perception of voicing in adults

As a starting point, we decided to replicate Stevens and Klatt's (1974) experiment (with some minor differences) on the role of formant transition rates in determining voicing judgments (also performed by Lisker, 1975). Three VOT continua were created, all of which varied from 10 to 50 ms in 10 ms steps, but whose F_1 and F_2 transition durations differed across the three continua (40, 50 and 60 ms).

In presenting their data, Stevens and Klatt plotted the number of /d/ and /t/ responses against the voice onset time relative to the end of the formant transition (hereinafter referred to as VOT[re:FTE] - VOT re Formant Transition End) rather than the more typical VOT (time from stimulus onset to the onset of voicing). VOT[re:FTE] captures something about the saliency of the formant transitions. Insofar as VOT[re:FTE] is the governing factor in the voicing decision, the voiced/voiceless boundary for stimuli with different formant transition rates should occur at the same point on the VOT[re:FTE] axis. Similarly, if VOT (as ordinarily defined) determines the voiced/voiceless boundary, then plots on a VOT axis would show all the boundaries at the same place. In fact, both in Stevens and Klatt's data, and in our replication, neither method of plotting the data leads to coincident boundaries for the variations in transition duration. This leads to the conclusion that both VOT and VOT[re:FTE] are influential in determining voicing.

However, as already pointed out, it may be that the effect of the voiced transition duration is through a change in F_2 transitions. A further experiment was conducted to establish whether F_1 transition, F_2 transition, or both, determine boundary location. The three previous continua were modified by making F_2 constant at a value of 2200 Hz, based on our estimate of the average locus frequency of the transitions in Stevens and Klatt's experiment. All other aspects of stimulus construction were as earlier. Six adults identified each sound 10 times. When the data are plotted against VOT[re:FTE] and VOT (see Figure 1), the picture is quite clear - when F_2 is constant, the plots against VOT[re:FTE] are separated by the amount of reduction in F_1 transition, but all curves superimpose when plotted against VOT. Therefore, it would appear that the variation in F_2 gave the boundary shifts with VOT[re:FTE] noted in Stevens and Klatt's data and our replication. In short, VOT completely controls whether the sound is classed as voiced or voiceless, independent of the duration of the F_1 transition, at least in these continua.

This finding may be thought somewhat surprising given the role of F_1 onset frequency found by Summerfield and Haggard (1977) and Summerfield (1982). However, changes in F_1 onset are relatively small here as the frequency of F_1 at 0 ms VOT would be identical for the three continua varying in F_1 transition duration. Although we plan further investigations of the role of F_1 onset, for the moment we have focused our attention on developmental issues, to gain a somewhat different perspective on the problem.

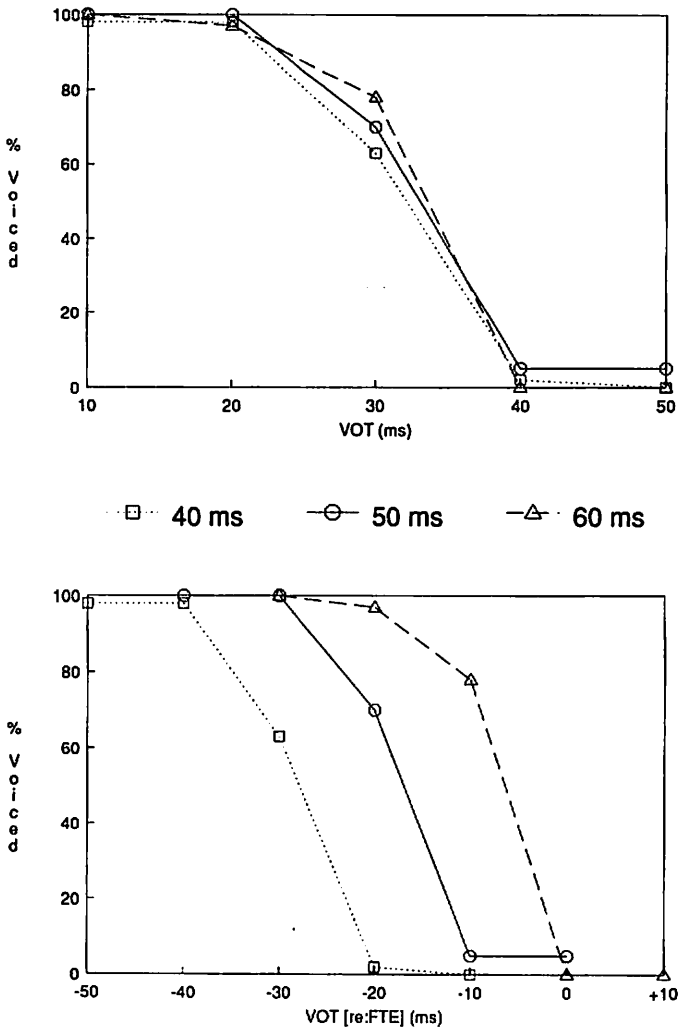

Figure 1. *The percentage of times stimuli were labelled as "d" by 6 adults for three synthetic VOT continua varying in F_1 transition duration. The data in the two plots is identical, but plotted against VOT measured in two different ways. Results are based on 60 responses per stimulus (6 listeners x 10 presentations).*

Perception of voicing in children and adults

In our experiments reported above, we focused on the extent to which F_1 transition duration (and implicitly F_1 onset frequency) affected voicing judgments. However, all the short VOT stimuli had an F_1 transition of some degree, which prevented us from answering a related question - to what extent does the presence or absence of an F_1 transition affect voicing judgments? Simon and Fourcin's (1978) study is particularly relevant to this question, because their findings differ from those of other related studies, in two main ways. Firstly, the effect of the presence or absence of an F_1 transition is rather more marked for their stimuli for teenagers and adults than is typically found (e.g., by Lisker, 1975). Secondly, their claim that young children are insensitive to the presence or absence of an F_1 transition, while not strictly contradictory to the finding that changes in F_1 transitions can affect infant behaviour (e.g., Miller and Eimas, 1983), is at least surprising. Therefore, we compared the performances of adults to young children of the ages Simon and Fourcin claim show no sensitivity to F_1.

The adult and child listeners were asked to identify various recorded versions (both natural and manipulated through editing) of the words "bat", "pat", "bees" and "peas" spoken by a young woman. Children responded by pointing to one of four pictures. Tests of live spoken and natural recorded phrases ("Point to Pat") showed that most children with English as their mother tongue could respond appropriately (10 of 11 children tested).

The edited stimuli were obtained by cross splicing the release burst and aspiration from the "p" member of the minimal pair onto the "b" vowel and final consonant (long VOT with an F_1 transition), and vice versa (short VOT with no F_1 transition). It might be expected that absence of an F_1 transition in the presence of a short VOT should have a greater effect on perception with an /æ/ vowel than an /i/ vowel, owing to its higher F_1 and hence expected greater transition. (Note, though, that these vowel onsets can differ in a variety of ways, and not only in the presence or absence of an F_1 transition.)

For the tests with isolated recorded words, a child's responses to edited stimuli were only tallied once the appropriate natural pair were responded to correctly. Twelve adults were also asked to label the same stimuli, a total of 20 observations per stimulus per listener.

Table I compares the averaged results of the adults to the averaged results of the five children who reached criterion on bat/pat as well as bees/peas (median age 3 years;7 months with a range from 2;10 - 4;8). The results labelled as "short" or "long" VOT represent the release burst and aspiration from the originally spoken "b" or "p" respectively. Thus since the percentage of "p" judgments is given in the table, short VOTs tend to have low scores and long VOTs high scores. The columns are labelled for children and adults as appropriate and the rows for the sounds with the vowels from the "b" or "p".

Table I. *The percentage of times stimuli were labelled as "p" by children and adults for natural and edited versions of two minimal pairs of words. Results in bold were obtained from natural (unedited) stimuli. Results for the adults are based on 240 responses per stimulus (obtained in two test sessions) whereas those for the children are based on about 50 responses per stimulus (obtained over a number of short test sessions).*

BEES-PEAS

	VOT			
	Short		Long	
	Children	Adults	Children	Adults
"b" vowel	**2.0**	**0.8**	88.5	99.6
"p" vowel	14.8	23.8	**98.0**	**100**

BAT-PAT

	VOT			
	Short		Long	
	Children	Adults	Children	Adults
"b" vowel	**0.0**	**0.8**	97.8	100
"p" vowel	41.9	9.6	**100**	**100**

For the long VOT values, the presence or absence of F_1 had no effect on the perception of adults (as found by Lisker, 1975), but some discernable effect on children's responses, at least for the bees/peas pair. For the short VOT pairs, both adults and children were affected by the following vowel, although by differing amounts. The size of the effect was considerably greater in the children for bat/pat, and somewhat greater in the adults for bees/peas. (Averaged data from all 10 children who reached criterion for bees/peas was similar to the children's bees/peas results included in the Table above.) As the measured differences in F_1 transition were much more marked between the vowels in bat/pat, this data suggests that there may be other cues in the vowel stem that the adults are attending to, but that the children are not. However, no differences in fundamental frequency contour were found, and differences in vocal fold open quotient (the percentage of each period that the vocal folds were open, and hence a measure of subglottal coupling) were also much more marked at vowel onset in the bat/pat pair. It is therefore difficult to know why the adults seemed to be more sensitive to the exchange of vowels in the bees/peas pair, although there may be yet other cues that influence voicing judgements (e.g., the presence or absence of a voiced F_2 transition).

In short, children between the ages of 2 and 4 are sensitive to spectral features at vowel onset when labelling initial plosive consonants that differ in voicing, contrary to Simon and Fourcin's (1978) claims. There are several reasons why this discrepancy might have occurred, not least because Simon and Fourcin tested the velar plosive contrast and used a synthetic continuum of stimuli, thus lacking other factors in natural stimuli to which the children might be sensitive. Yet when using a whole continuum any effects of F_1 should have been more apparent because it tests at values of VOT closer to the phoneme boundary, with corresponding greater ambiguity. It may also be that the F_1 transition in the synthetic stimuli was unnaturally salient, leading to larger effects in adults than we found, yet unnatural enough to be ignored by the younger children. Unavoidable variability in children's responses and an averaging of all the results may also have obscured effects in some of the children. Furthermore, it should be remembered that the results of the four-year-olds did seem to show strong sensitivity to differences in F_1.

It is interesting to note that our data appear to indicate that children are generally more sensitive to the spectral aspects of the following vowel in plosive voicing contrasts than are adults, both in the average data, and in the individual results: 1) Adults are never affected by the exchange of vowels when a long VOT is present, whereas the children sometimes are. 2) All the children who were able to respond to the natural bat/pat pair were affected to some degree by the exchange of vowels when the VOT was short (6/6), whereas only 4 of 12 adults were. By the same token, the average size of the effect was much greater for the children (see Table I). 3) Although the average size of the effect of vowel exchange for the bees/peas pair with a short VOT was a little greater for the adults than for the children (see Table I), the proportion of listeners affected was similar in adults and children (58.3% of 12 adults and 60% of 10 children, including all children who were able to label the bees/peas pair, not only those included in Table I). Therefore, it appears that the developmental trend is away from reliance on factors other than VOT, a finding consistent with the results of Experiment 1, in which it appeared that adult judgements of voicing did not depend on changes in F_1 transition duration.

Acknowledgements

Grateful thanks to Valerie Hazan and Bo Shi for their help in the design and execution of the study on children. Valerie Hazan also made many useful criticisms of an earlier version of the manuscript. The study on children and adults using natural and edited stimuli was done as part of a final year project in Speech Communication by Harriet Lang. This work was supported by the Medical Research Council of the UK.

References

Lisker, L. (1975) In (qualified) defense of VOT. *Haskins laboratories: Status Report on Speech Research* SR-44, 109-117.

Miller, J. & Eimas, P.D. (1983) Studies on the categorization of speech by infants. *Cognition*, 135-165.

Simon C. & Fourcin, A. (1978) A cross-language study of speech pattern learning. *J. Acoust. Soc. Am.*, 63, 925-935.

Stevens, K.N. & Klatt, D.H. (1974) Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.*, 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. *J. Acoust. Soc. Am.*, 72, 51-61.

Summerfield, Q. & Haggard, M. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *J. Acoust. Soc. Am.*, 62, 453-448.