Proceedings of The Institute of Acoustics

FO PERTURBATIONS AS A FUNCTION OF VOICING OF PREVOCALIC AND POSTVOCALIC STOPS AND FRICATIVES, AND OF SYLLABLE STRESS

Kim Silverman

MRC Applied Psychology Unit, Cambridge

INTRODUCTION

Fundamental frequency (henceforth FO) contours in human speech are known to be the result of segmental as well as suprasegmental structure. Vowels and consonants induce characteristic perturbations that seem to ride on top of the underlying intonation. Unfortunately, there are significant gaps and contradictions in the literature about perturbations around intervocalic consonants, and the general question of how they might interact with suprasegmental structure has been largely ignored.

The most established fact about segmental influences is that FO at the release of a prevocalic consonant is lower if the consonant is voiced than if it is voiceless. But there is no consensus in the literature about just what FO contours look like after release and in the following vowel. On the one hand, some investigators have reported a so-called "fall/rise dichotomy", meaning that the voicing status of a consonant determines the direction of subsequent FO movement. For example, Hombert [1] showed that FO falls from a high value in the first 100 milliseconds of a vowel following a voiceless stop, but that it rises from a low value if the vowel follows a voiced stop. Production data supporting this claim have been presented by, amongst others, House and Fairbanks [2], Lehiste and Peterson [3] and Mohr [4]. Perceptual data also supports this distinction. Experiments using synthetic speech have shown that the direction of movement of FO after stops can indeed act as a cue to discriminate between voiced and voiceless cognates (e.g. [5],[6]). Haggard, Summerfield and Roberts [7] have found that the direction of FO movement can be traded against, and even override, VOT as a cue to voicelessness. Lea [8] has even suggested the rise/fall dichotomy as a context-invariant phonetic cue to voicing.

Recently, however, in a series of experiments, Ohde [9] has consistently found that FO always falls during the transition from a stop to a vowel, regardless of the stop's voicing status. He did find that it falls from a greater height after voiceless stops, but nevertheless found no evidence of any rise in FO after voiced stops. A similar pattern was reported by Haycock and Haggard in 1970 [10]. This, then, is one of the contradictions in the literature: is there a fall/rise dichotomy after stop release?

Further disagreement can be found in studies of the effects of postvocalic stops on preceding vowels. Lea [13] suggests that both voiced and voiceless stops lower the FO of a preceding vowel. A similar pattern is reported in [11]. But other studies (e.g. [4]) indicate that postvocalic consonants either raise or lower FO in preceding vowels, depending on whether they are voiceless or voiced. Again, perceptual data using synthetic speech have shown that the direction of FO movement in preceding vowels can indeed act as a cue to voicing [12]. Yet Lehiste and Peterson [3] claim that postvocalic consonants in natural (as opposed to synthetic) speech have no effect at all on preceding vowels. The contradiction between these three different claims about postvocalic stops has not yet been resolved.

Although some of the above contradictions may be attributable to different articulatory gestures in different languages, there remains a clear need for more studies, with more controlled methodology. For example, none of the studies in the literature adequately controls or even describes the intonation used by speakers during experiments. Some use lists of isolated words, in which the choice of intonation is completely left to the speakers. Others use "frame" sentences that have the target words in nuclear position. Even if all sentences are spoken as isolated declaratives, the large pitch excursion associated with the nuclear accent may obscure or interfere with the segmental perturbations. In other studies (e.g. [3], [14]) continuous prose was used, with no control at all of the intonation.

As well as contradictions, there are several gaps in the published literature. There are relatively few studies of postvocalic stops, and even fewer studies of either pre- or postvocalic fricatives. There seems to be an assumption that since the airflow is disturbed less by fricatives than by stops, the associated segmental perturbations will be less extreme. Yet a recent study [15] has shown that, at least in Danish, fricatives have even larger perturbations than stops.

As mentioned above, the interplay between segmental influences and suprasegmental structure has been largely unaddressed. One possible interaction may be between consonantal perturbations and the degree of syllable stress. There are no studies that compare consonantal perturbations in stressed versus unstressed syllables. Yet effects of stress have been found in every other acoustic parameter measured, such as overall energy [16], duration [17], glottal waveshape [18], formant trajectories [19], and vowel intrinsic pitch [20],[21]. So it seems reasonable to look for similar differences in FO perturbations.

The investigation reported here addresses the above issues. Specifically it aimed to collect data on Southern British English, to investigate FO perturbations in vowels both before and after consonants, to compare stops with fricatives, and to look for an interaction with the degree of syllable stress.

METHOD

Speech Material

The material was chosen to include all of the singleton obstruents occurring in RP English, in a variety of stress contexts. Test words of the form:

ə C v C v C

(where C = Consonant, and v = vowel) with main (lexical) stress on the middle syllable, were embedded in the sentence:

The vowels in ______ are shortened.

FO PERTURBATIONS AS A FUNCTION OF VOICING, MANNER, AND STRESS

The consonants were:

	Stops	Fricatives
Voiced	b d g	v ð ʒ z
Voiceless	p t k	f θ ʃ s h

Affricates were not included in this study. The vowels in the test-words were either /i:/ or /a:/. In each test-word the vowels and consonants were the same (e.g. əpi:pi:p). Speakers were asked to reduce the "are" after the test-words to a schwa. This is the unmarked realisation of "are" in this context and all speakers found this quite natural. The result was that FO transitions could also be measured after the final consonant.

This design assumes that syllables can be classified into at least three degrees of syllable stress:
(i) stressed (i.e. carrying primary lexical stress and a rhythmic beat in the utterance) with full vowel quality: the middle syllable of the test words used in this investigation,
(ii) unstressed with full vowel quality: the last syllable of the test words, and
(iii) unstressed with vowel quality reduced to /ə/: the first syllable of the test words and the subsequent reduced "are".

The full set of 150 sentences (2 vowels x 15 consonants x 5 repetitions) was randomised and printed on index cards. Three adult speakers (two male and one female) of the South East England variety of RP English each recorded all of the sentences in one session in random order. All speakers were instructed to always use level intonation heads and a falling nucleus:

The vowels in ə C v C v C ə shortened.

Because the nuclear accent was located on the final word in the sentence, any extraneous FO movement attributable to "intonation" during the test-word was minimised.

Measurement Procedures

The 450 utterances were digitised at 10 kHz, after band-pass filtering from 50 Hz to 4.8 kHz with 90 dB/octave rolloff. With the aid of interactive computer graphics, FO was measured directly from displays of the speech waveforms. Rather than measuring every glottal period throughout each test-word, several points were selected that were comparable across utterances and were most likely to capture any segmental perturbations.

There were 14 points of measurement in all. These were:
- five glottal periods before each preconsonantal voice offset,
- at each preconsonantal voice offset (if voicing continued throughout closure then the point at attainment of maximum closure was measured),
- at each postconsonantal vowel onset,
- five glottal periods after each vowel onset, and
- the middle of the two full vowels.

At each point two successive glottal periods were measured and averaged. This averaging was necessary to eliminate the confounding effects of pitch jitter that often occurred after voiceless stops.

RESULTS AND DISCUSSION

Each speaker's utterances were divided into two groups, according to the voicing of the obstruents in the test words. At each of the 14 measured points, FO was averaged across all utterances within each group. The means are plotted for the two male speakers (IW and FJN) in figures 1a and 1b. At the time of writing, measurements of the female speaker's data are still being completed. The following discussion of the results is based on the mean FO values for IW and FJN, and so should be viewed as a progress report. After the trends reported here are subjected to appropriate statistical testing, this report will be revised and expanded for subsequent publication.

Effects of voicing and manner on FO

The pattern in the data is clear: obstruent voicing did not determine the direction of FO movement. Figure 1 shows that no rise/fall dichotomy was found for either speaker. Instead, FO fell after all obstruents in all syllables. But it did consistently fall from a higher level if the obstruent was voiceless, and this difference persevered into the vowel. Similarly, FO turned downwards before all obstruents in all syllables, reaching its lowest level if the following obstruent was voiced.

For both speakers, the difference in vowel FO height attributable to the voicing of surrounding consonants generally perseveres throughout the whole of the syllables containing full vowels. The structure of the test words does not make it possible to separately quantify the effects of a preceding and a following consonant on the FO contour through an intervening vowel. But it is clear that the height differences in the present data are not solely due to the prevocalic obstruents. There is also a backward contribution from those in the postvocalic position. Two patterns in the results support this claim. Firstly, a difference in FO height due to voicing is present in the initial syllable for both speakers. This difference must be due to the following consonant, because the preceding consonant in this syllable was always /n/ in every utterance. Secondly, if the differences were completely attributable to perseverence of the effect of prevocalic consonants then the FO contours in the vowels should converge, or at least remain parallel. However, the opposite pattern is evident in the present data. The FO contours preceding the consonants actually diverge according to the voicing of the following consonants in all positions for FJN and for two of the three positions for IW.

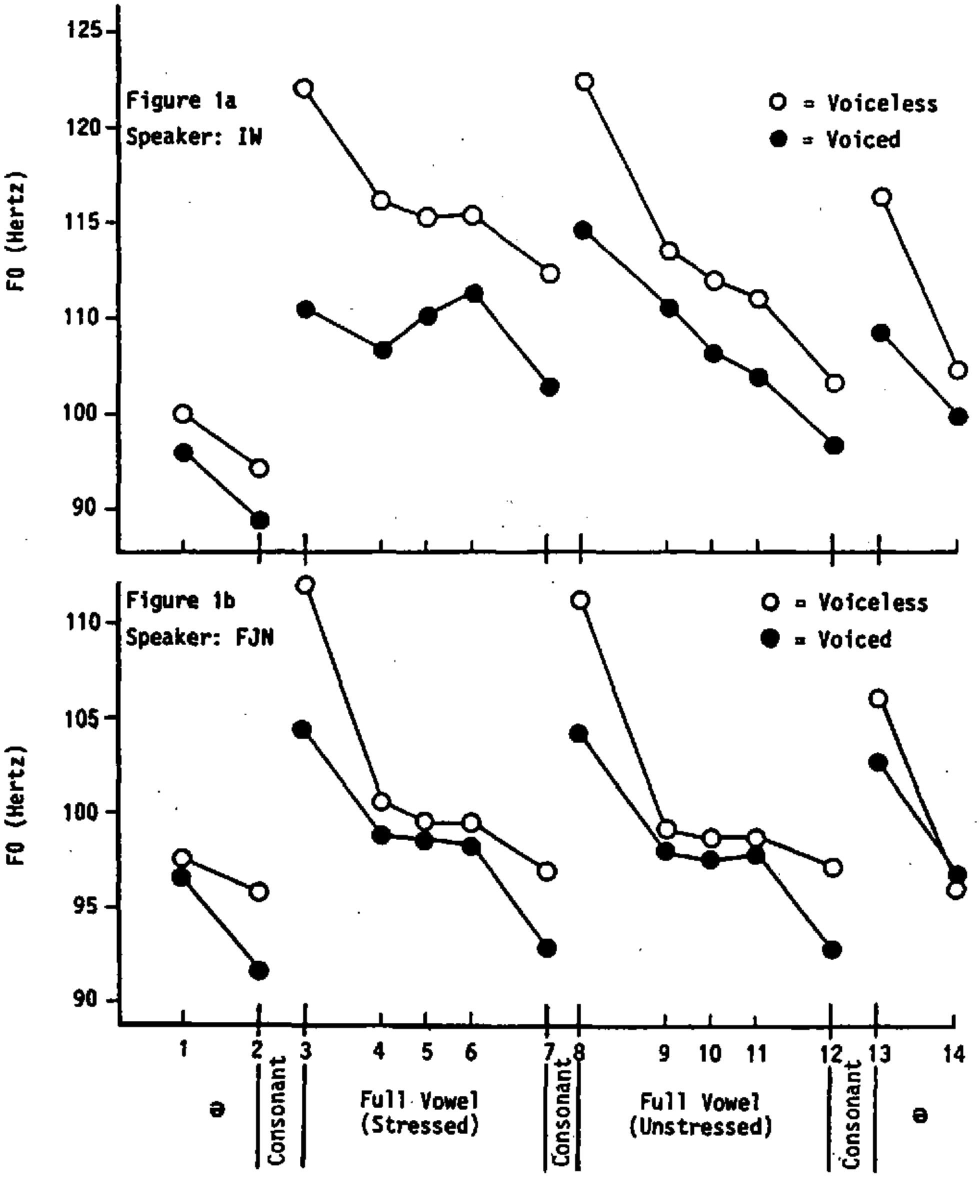

Figure 1. Time-normalised plots of mean FO at the 14 points of measurement for each speaker. Separate means are shown for voiced and voiceless consonants.

FO PERTURBATIONS AS A FUNCTION OF VOICING, MANNER, AND STRESS

A difference between stops and fricatives appeared for one speaker (IW). In the two syllables with full vowel quality the difference in FO height due to obstruent voicing was 61% greater for fricatives than for stops. Although this difference must be confirmed by statistical analysis, it should be pointed out here that it was consistent throughout both full syllables, and a similar trend can be seen in Reinholt Petersen's data [15].

Interactions with the degree of syllable stress

There are two ways in which the perturbations interact with syllable stress. Firstly, the difference in FO height due to voicing increases as stress increases, as can be seen in table I.

Table I. Differences between the average FO in voiced test words and the average FO in voiceless test words, according to syllable stress. (FO in the middle of the full vowels was excluded, to maintain comparability with the reduced syllables.)

degree of syllable stress	mean FO (voiceless - voiced)	
	IW	FJN
stressed, full vowel	7.4 Hz	3.7 Hz
unstressed, full vowel	4.6 Hz	3.5 Hz
unstressed, reduced vowel	3.6 Hz	1.9 Hz

Secondly, the overall perturbation of FO around obstruents (regardless of their voicing) increases with the stress of the syllable that the obstruents precede. This perturbation was calculated as the height of FO immediately before and after each consonant, relative to a reference level. This reference level for each consonant position was the FO level halfway between FO at the measurement point five glottal periods before voice offset and FO at the point five periods after voice onset. These relative heights are shown in table II.

Table II. Magnitude (in Hertz, pooling voiced and voiceless obstruents) of perturbation at voice offset and onset around consonants.

Consonant number in test word	Stress of following syllable	IW		FJN	
		offset	onset	offset	onset
1	stressed, full vowel	-7.3	+8.3	-4.6	+9.7
2	unstressed, full vowel	-2.9	+6.4	-3.8	+8.8
3	unstressed, reduced vowel	-2.6	+5.2	-2.2	+6.8

Functions of FO perturbations in Speech Communication

Does FO cue voiced/voiceless distinctions? Since the direction of FO movement was the same around both voiced and voiceless obstruents, in all positions, for both speakers, then one might be tempted to reject FO as such a cue. Certainly for English, the data reported here finally lay Lea's purported fall/rise dichotomy to rest. But that doesn't exclude FO from contributing to the distinction. After all, FO is always higher after obstruents if they are voiceless. But higher relative to what? To make use of this cue, listeners must have some reference for comparison. On the basis of the current results, I would suggest that the level of FO before consonant closure is crucial. The FO contour in a syllable, considered in isolation, means nothing. But the height of FO after release, relative to where a listener would expect it to be on the basis of the prior intonation, is reliably dependent on consonant voicing.

Other useful information about segmental structure is also available in FO. A sudden downward sweep of FO in a stream of speech can signal that a consonant is imminent. Furthermore, from table II, it is clear that for both speakers the extent of this downward movement before a consonant also predicts the degree of stress of the next syllable.

This latter relationship is particularly important, in the light of what is known about how humans process speech. In English, stressed syllables carry much more information about word identity than unstressed ones, and so it pays for listeners to focus their processing efforts on the stressed syllables in an utterance. Indeed, listeners are quicker to recognise phonemes in stressed syllables, and are able to predict the location of stressed syllables from the preceding suprasegmental structure [22]. The data reported here show that FO contains sufficient information to support such prediction.

Implications for speech synthesis

Compared to human speech, speech that has been synthesised by rule is less natural, less intelligible, takes longer to recognise, places more demands on working memory, and is less robust in a noisy environment. One of the underlying causes is inadequate FO synthesis. If the segmental perturbations presented in this paper were included in synthesis rules, then the quality of synthetic speech could only benefit. FO contours would not be so smooth (smooth and slowly-varying FO contours contribute to the characteristic nasal and mechanical sound of synthetic speech). Instead they would be more broken up -- more squiggly -- as are natural FO contours. Perhaps more importantly, there would be more redundant (and more natural) encoding of information about the location and voicing of consonants and the degree of syllable stress.

References

[1] Hombert, J-M. (1978), Consonant Types, Vowel Quality and Tone. In V. Fromkin (Ed), Tone: A Linguistic Survey. New York: Academic Press.

[2] House, A.S. and Fairbanks, G. (1953), The Influence of Consonantal Environment on the Secondary Acoustical Characteristics of Vowels. J.A.S.A., 25, 1, 105-135.

[3] Lehiste, I. and Peterson, G.E. (1961), Some Basic Considerations in the Analysis of Intonation. J.A.S.A., 33, 4, 419-425.

[4] Mohr, B. (1971), Intrinsic Variations in the Speech Signal. Phonetica, 23, 65-93.

[5] Chistovich, L. (1969), Variation of the Fundamental Voice Pitch as a Discriminatory Cue for Consonants. Soviet Physics: Acoustics, 14, 372-78.

[6] Haggard, M., Ambler, S. and Callow, M. (1970), Pitch as a Voicing Cue. J.A.S.A., 47, 613-617.

[7] Haggard, M., Summerfield, Q. and Roberts, M. (1981), Psychoacoustical and Cultural Determinants of Phoneme Boundaries: Evidence from Trading FO Cues in the Voiced-Voiceless Distinction. Journal of Phonetics, 9, 49-62.

[8] Lea, W.A. (1973), Segmental and Suprasegmental Influences on Fundamental Frequency Contours. In L.M. Hyman (Ed), Consonant Types and Tone. Southern California Occasional Papers in Linguistics, 1, 15-70.

[9] Ohde, R.N. (1984), Fundamental Frequency as an Acoustic Correlate of Stop Consonant Voicing. J.A.S.A., 75, 1, 224-230.

[10] Haycock, K. and Haggard, M. (1970), Pitch Skip in Stop Consonants. Speech Synthesis and Perception, 3, 23-26.

[11] Di Cristo, A. and Chafcouloff, M. (1977), Les Faits Microprosodiques du Francais: Voyelles, Consonnes, Coarticulation. VIIIèmes Journees d'Etude sur la Parole. Aix-en-Provence, 147-158.

[12] Gruenenfelden, T.M. and Pisoni, D.B. (1980), FO as a Cue to Postvocalic Consonant Voicing: Some Data from Speech Perception and Production. Perception and Psychophysics, 28, 6, 514-520.

[13] Lea, W. (1973), Influences of Phonetic Sequences and Stress on Fundamental Frequency Contours of Isolated Words. J.A.S.A, 53, 1, supplement, p346.

[14] Umeda, N. (1981), Influence of Segmental Factors on Fundamental Frequency in Fluent Speech. J.A.S.A., 70, 350-355.

[15] Reinholt Petersen, N. (1983), The effect of Consonant Type on Fundamental Frequency and Larynx Height in Danish. ARIPUC, 17, 55-86.

[16] Lehiste, I. and Peterson, G.E. (1959), Vowel Amplitude and Phonemic Stress in American English. J.A.S.A., 31, 4, 428-425.

[17] Klatt, D. (1976), Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence. J.A.S.A., 59, 5, 1208-1221.

[18] Fant, G. (1980), Voice Source Dynamics. STL-QPSR 2-3/1980.

[19] Stålhammar, U., Karlsonn, I. and Fant, G. (1973), Contextual Effects on Vowel Nuclei. STL-QPSR, 4/1973, 1-18.

[20] Reinholt Petersen, N. (1979), Variation in Inherent FO Level Differences between Vowels as a Function of Position in the Utterance and in the Stress Group. ARIPUC, 13, 27-59.

[21] Steele, S.A. and Godfrey, J.J. (1982), Vowel Intrinsic Pitch as a Function of Prosodic Context. Linguistic Society of America Summer Meeting, 1982.

[22] Cutler, A. and Darwin, C.J. (1981), Phoneme-monitoring Reaction Time and Preceding Prosody: Effects of Stop Closure Duration and of Fundamental Frequency. Perception and Psychophysics, 29, 3, 217-224.