Segmental influences on F0: Automatic or controlled?[1]

John Kingston

> "Aus der Kriegsschule des Lebens. Was mich nicht umbringt, macht mich stärker".
> Sprüche and Pfeile 8,
> *Götzen-Dämmerung oder Wie man mit dem Hammer philosophirt* (1889)
> Friedrich Wilhelm Nietzsche

Abstract

Results of two experiments are reported in which prosodic context was varied to investigate the automaticity of vowel height's and obstruent voicing's influence on F_0. The results of the first experiment, in which the author was the sole speaker, showed that F_0differed between high, mid, and low vowels in accented but not unaccented syllables, suggesting that F_0 differences between vowels contrasting in height are controlled rather than automatic. This result replicates earlier findings reported by Ladd and Silverman (1984) and Steele (1986). F_0 differed next to voiced and voiceless obstruents in unaccented as well as accented syllables, suggesting that the F_0 differences next to obstruents contrasting in voicing are instead automatic. The results of the second experiment, in which four naive speakers produced the materials, were quite different: F_0 differences between vowels contrasting in height were found in unaccented as well as accented syllables, and F_0 differences next to obstruents contrasting in voicing were small, inconsistent across speakers, and not reliably influenced by prosodic context. The results of the second experiment indicate that a more sensitive manipulation is needed of information content, its implementation by pitch accents, and their effect on F_0 differences related to vowel height and obstruent voicing contrasts to determine whether these differences are automatic or controlled.

1. Introduction

In higher vowels and in vowels next to voiceless obstruents, the vocal folds vibrate faster than they do in lower vowels and in vowels next to voiced obstruents. These differences are widely thought to arise as unintended and automatic side effects of intended differences in tongue height or vocal fold vibration (Lehiste 1970; Hombert 1978; Hombert, Ohala, and Ewan 1979; Whalen and Levitt 1995; Connell 2002). I here present the results of two experiments designed to test the contrary hypothesis: that these F_0 differences might themselves also be produced by intended or controlled articulations (Kingston 1991, 1992; Kingston and Diehl 1994).

The design of the experiments rests on the assumption that the speaker cannot control automatic differences in F_0. If they are automatic, then they should be observed whenever the speaker produces the intended articulations that automatically cause F_0 to differ. This is what it means to be automatic. This prediction is tested by varying the prosodic context in which the target segments occur, such that they occur in unaccented as well accented words and in words bearing low (L) as well as high (H) tone. Automatic differences should appear in all these contexts. If, however, these F_0 differences are controlled, then it's possible, perhaps even likely that they would appear in some of these contexts but not others. Specifically, they are expected to occur in prosodically prominent contexts, i.e. in accented words, and in words bearing a L* as well as H* pitch accents.

Why should this be so? What purpose is served by varying F_0 with either vowel height or obstruent voicing and how is that purpose served by only doing so in prominent contexts? Varying F_0 adds another acoustic difference to vowels contrasting in height and obstruents contrasting in voicing and enhances these contrasts. Adding these differences in prominent contexts enhances these contrasts at sites where information content is high.

If phonetic properties such as these F_0 differences are controlled rather than automatic, then we must revise our notions of how distinctive features are realized phonetically and indeed of what the distinctive features are that represent these contrasts. Instead of speakers producing just one or a few phonetic differences intentionally, e.g. differences in tongue height or voicing, and getting many additional phonetic differences for free because they depend mechanically on the intended differences, speakers must control many more of the differences, perhaps all of them. If many or all these differences are controlled, it is at best arbitrary to treat the contrast as one of tongue height or voicing rather than of any of the other phonetic differences. The contrast is described better corporately in terms of the entire array of phonetic differences that might realize it. Because the contrast's realization also varies across contexts, no phonetic property is essential to its realization.

If no phonetic property is essential to realizing a contrast, then one must explain why speakers produce the particular combinations of phonetic differences that they do and why a particular combination is used in one context and not another. In other work (Kingston and Diehl 1994; Kingston and Macmillan 1995; Kingston, Macmillan, Dickey, Thorburn, and Bartels 1997; Kingston, Diehl, Kirk, and Castleman, submitted), I've argued that speakers produce these combinations and not others because the members of the observed combinations integrate in such a way that they exaggerate one another's perceptual effects. On this account, it's also unsurprising that the observed F_0 differences would be added to height and voicing differences preferentially in prominent contexts, as exaggeration there would ensure that the segment's high information context is reliably conveyed.

2. Controlled F_0 differences between different vowel heights?

This section reviews the evidence that leads me to think that F_0 differences between vowels contrasting in height are controlled. That evidence also inspired the method chosen to look for evidence of control, manipulating the prosodic context in which the sounds of interest occur. No comparable evidence indicates that F_0 differences next to obstruents contrasting for voicing are also controlled; however, none of the attempts to explain them as automatic side effects of other articulations in voiced and voiceless obstruents have been successful (Kingston and Diehl 1994). These failures suggest that it would be worthwhile to test whether they, too, are instead controlled.

My first reason for doubting that the F_0 differences are automatic side effects of other, intended articulations is that their automatic nature has been explained in so many different ways. Vowel height has been argued to affect F_0 acoustically, aerodynamically, and mechanically, and a similar variety of incompatible explanations have been offered to explain how obstruent voicing affects F_0 (Ohala 1973; Ohala and Eukel 1987; DiCristo et al. 1979; Shadle 1985; Steele 1986; Silverman 1987; Sapir 1989; Fischer-Jørgensen 1990; Honda and Fujimura 1991; Halle and Stevens 1971; Riordan 1980; Hombert et al. 1979; Löfqvist, Baer, McGarr, and Seider Story 1989). The variety and incompatibility of these explanations and the lack of consensus are themselves reasons to consider alternatives. Before turning to the empirical reasons for treating these F_0 differences as controlled, it is nonetheless useful to consider at least the more plausible explanations of how differences in vowel height or obstruent voicing might automatically cause F_0 to differ – these effects will be referred to henceforth as 'VF_0' and 'CF_0'.[2]

For VF_0, raising the tongue body to produce a high vowel could pull mechanically on the vocal folds via the hyoid bone and increase their tension and rate of vibration (Ohala and Eukel 1987). For CF_0, the folds may be slackened because the larynx is lowered to expand the oral cavity and slow the rise in intraoral air pressure. The larynx does lower otherwise when a low F_0 is produced (Collier 1975; Ewan 1976; Erickson, Honda, Hirai, and Beckman 1995; Honda, Hirai, Masaki, and Shimada, 1999), but it's presumably the aerodynamic benefit and not F_0 lowering that's its purpose in voiced obstruents. These accounts are excellent examples of attempts to explain F_0 differences as automatic. In each case, the speaker produces some articulation deliberately, tongue body raising or larynx lowering, in order to achieve a particular goal, a close constriction or voicing, and F_0 is automatically and inadvertently altered by that intended articulation. There are a host of problems with both explanations, but because space is limited I'll only mention the most important ones here. (Kingston 1991, 1992; Kingston and Diehl 1994.)

The first problem for explanations of both VF_0 and CF_0 differences is that they are apparently determined by the vowel's or obstruent's phonological specification and not its tongue height or voicing. With respect to VF_0, Fischer-Jørgensen (1990) reports that F_0 is higher in high than mid vowels in German, but F_0 doesn't differ between tense and lax vowels of the same phonological height, despite the tongue being considerably higher in the tense vowels than their lax counterparts. With respect to CF_0, evidence reviewed by Kingston and Diehl (1994) shows that F_0 is raised next to voiceless unaspirated stops in languages such as French where this is the pronunciation of [–voice] stops, but lowered next to such stops in languages such as English where this is the pronunciation of [+voice] stops. Both of these findings are hard to reconcile with the claim that some other intended articulation automatically influences F_0, because the F_0 differences don't appear to depend in any obvious way on the other phonetic properties of the vowel or obstruent. However, both are expected if speakers deliberately produce the F_0 differences to enhance the height or [voice] contrast.

Another important difficulty for the tongue pull explanation of VF_0 differences is that a number of studies have reported higher levels of activity in the cricothyroid muscle in higher vowels (Vilkman, Aaltonen, Raimo, Arajävi, and Oksanen 1989; Dyhr 1990). As contracting this muscle is perhaps the principal means of deliberately raising F_0 (Ohala 1970; Atkinson 1978; Collier 1975), these findings suggest that it might be contracted to raise F_0 deliberately in higher vowels, too. In this connection, it's important to acknowledge that Whalen, Gick, Kumada and Honda (1998) recently failed to find higher levels of cricothyroid activity in higher vowels, so the data on this point are not entirely consistent.

In the rest of this section, I will present other empirical reasons to doubt that tongue pull is responsible for F_0 differences between high and low vowels. These findings also underpin the experimental design used to collect the new data on CF_0 as well as VF_0 differences reported in this paper.

This design was inspired by the insights and results of Ladd and Silverman's (1984) study of the effects of intonational prominence on the production of VF_0 differences in German and the similar study using English data carried out by Steele (1986). In both studies, the intonationally prominent syllable bore a high pitch accent (H*), which was realized as an F_0 peak on its vowel. In the German data, syllables which were not intonationally prominent were only in the domain of a L– phrase accent, but in the English data, could be in the domain of a H– or a L– phrase accent. Both studies obtained the same results: VF_0 differences shrank or disappeared completely in syllables that weren't intonationally prominent, and in the English data, this shrinkage and disappearance occurred as much in H– as L– domains. If VF_0 differences are a mechanical side effect of tongue height differences, their shrinkage and disappearance in unaccented syllables is utterly unexpected, as the vowels in the non-prominent syllables remained lexically stressed and therefore didn't reduce in either language.

Even though the unaccented vowels didn't reduce in either German or English, they may still have been hypo-articulated compared to accented ones. Neither Ladd and Silverman nor Steele assessed differences in the pronunciation of accented vs. unaccented vowels, so it's not possible to tell from their data whether VF_0 differences shrank and disappeared in concert with the shrinkage and disappearance of other differences in the vowels they compared. If hypo-articulation shrinks differences in tongue height, then the tongue pull hypothesis predicts that covarying F_0 differences should shrink as well, in proportion to the shrinkage in tongue height differences. It is certainly possible that hypo-articulation is responsible for the some of the shrinkage of VF_0 differences in Ladd and Silverman's and Steele's data, but they shrink too much – in some instances all the way to 0 – for all or even most of the shrinkage to be attributed to hypo-articulation.

The dependence of VF_0 differences on tone in tone languages suggests that their occurrence and size is probably not after all an automatic side effect of how hyper- or hypo-articulated the tongue height differences are. In two tone languages, Yoruba (Hombert 1977) and Taiwanese (Zee 1980), significant VF_0 differences are found in high (H) but not low (L) tone syllables. Similarly, twice as many speakers produced significant VF_0 differences in syllables with higher tones as did in those with lower tones in the data from four West African tone languages, Ibibio, Kunama, Mambila, and Dschang, reported by Connell (2002). None of these studies report any differences in the size of tongue height differences between vowels in H vs. L tone syllables, and I'm not aware of any other studies that do, so it's unlikely that the disappearance and shrinkage of VF_0 differences in L and lower tone vowels is due to any hypo-articulation of vowels bearing such tones. Taken together with the German and English results, these findings from tone languages suggest that VF_0 differs between high and low vowels when they bear a H tone, although in non-tone languages, that H must be prominent and not merely high in F_0, i.e. a pitch accent and not merely a phrase accent.

Standard Copenhagen Danish shows that it's prominence that matters and not the height of the tone. In that language,VF_0 differs significantly between high and low vowels in lexically prominent syllables, which bear a L tone, but not in immediately following syllables, whose tone is H instead (Reinholt Petersen1978).[3] The only way to incorporate this last result into a general account of the effects of prosody on VF_0 is to distinguish between languages where (some) tones occur exclusively on prominent syllables and others appear elsewhere and languages where particular tones aren't associated with prominence. In languages of the former kind like German, English, or Danish, VF_0 differs significantly only in prominent syllables, regardless of whether a H or a L tone marks that prominence, while in languages of the latter kind like Yoruba or Taiwanese, VF_0 differs significantly in syllables bearing H but not L tone.

Because VF_0 differences appear in some prosodic contexts but not others and the tongue height differences between vowels differ little or not at all between these prosodic contexts, it's difficult or impossible to treat the VF_0 differences where they do occur as an automatic side effect of another intended articulatory difference such as the height of the tongue. They might instead be plausibly treated as the product of controlled articulations in their own right.

Why would a speaker control F_0 in this way? It's easy to answer this question for languages such German, English, or Danish where the VF_0 differences appear in prosodically prominent contexts. Such contexts are sites where the local information content is high. In the German and English materials analyzed by Ladd and Silverman (1984) and Steele (1986), a vowel is in a prominent context because it occurs in a word bearing a pitch accent that arises in the intonation, while in the Danish materials analyzed by Reinholt Petersen (1978), the vowel occurs in the lexically prominent syllable in the word and the pitch accent is specified in the lexicon. The difference between intonational and lexical pitch accents is simply one of scale: prominence and information content are high for the entire word in the German and English materials, but they are high only for one syllable in the Danish materials. Even so, the pitch accent is aligned with the lexically prominent syllable in the German and English words, so the interval during which prominence and information content are high may actually be no more extensive in words from these languages than in the Danish words. VF_0 differences are present or much larger in these prominent syllables because exaggerating the phonetic differences enhances the contrast between the vowel that occurs there and another vowel of a minimally different height that could occur there.

It is much less obvious why a speaker should control VF_0 in tone languages such Yoruba or Taiwanese, where a H tone conveys no more prominence than a L tone. However, H tones differ from L tones in another way that motivates permitting VF_0 differences at the top of the speaker's range, while limiting them at the bottom. Speakers are far freer to vary F_0 at the top when pronouncing a H tone (Liberman and Pierrehumbert 1984), in particular they can raise F_0 more without pushing that tone's F_0 target into the range of another tone. But when pronouncing a L tone at the bottom, speakers run up against a rather hard floor that prevents further F_0 lowering, and they cannot raise F_0 much without raising that tone's F_0 target into the range of a higher tone's target.

This explanation is functional, like that for the restriction of VF_0 differences to prominent syllables in languages where some tones mark syllables as prominent. Both explanations refer directly to the information the speaker is trying to convey to the listener in choosing to produce VF_0 differences in some contexts but not others. The two explanations nonetheless differ. VF_0 differences are restricted to prominent syllables in languages where pitch accents confer phrasal or lexical prominence because they are one of the means of exaggerating contrasts in sites that the speaker estimates to have high information content. VF_0 differences are restricted to H tone syllables in languages where tones convey contrasts between one word and another because they won't lead to confusion of one tone with another at the top of the speaker's range.

Before continuing it is worthwhile to consider a competing explanation for the finding that VF_0 differences shrink or disappear in non-prominent syllables in German, English, and Danish. Perhaps, raising the tongue automatically pulls on the vocal folds and causes them to vibrate faster in higher vowels, but this pull is suppressed in unaccented syllables.[4] These differences might be suppressed in such syllables to ensure that they are sufficiently hypo-articulated to be perceived as non-prominent. This proposal resembles the suggestion above that VF_0 differences are suppressed in syllables with L tones in Yoruba and Taiwanese to prevent confusion with higher tones.

However, if VF_0 differences can be suppressed when the speaker wishes to hypo-articulate the vowel, it is difficult to see how they can be an automatic consequence of the tongue pulling on the vocal folds. They must instead be controlled, albeit to reduce prominence by not enhancing the vowel height contrast.

There is another, subtler objection to this explanation. Most syllables in an utterance are not prominent, and the ones that are prominent are typically pronounced with a more extreme F_0 excursion, longer duration, and/or greater intensity to convey their prominence. Recent work has shown that articulations which realize features of individual segments are also more extreme in prominent contexts (Fougeron and Keating 1997; Keating, Cho, Fougeron and Hsu 1999; Cho and Keating 2001). These more extreme articulations presumably enhance the contrast between these segments and minimally different ones. In light of these findings that speakers otherwise only produce something extra in the relatively small number of prominent syllables in an utterance, it would be quite surprising if putatively automatic F_0 differences were only produced in the few prominent syllables and not the many non-prominent ones.

In this paper, I try to extend the finding that segmentally conditioned F_0 differences depend on prosody in two ways. The first extension is measuring the occurrence and size of VF_0 differences in English syllables bearing L* as well as H* pitch accents and in H– and L– domains. This extension tests the generality of the claim that VF_0 differences will only be found in prominent syllables in a language of this kind, and that these differences will be found regardless of whether the prominent syllable bears a L or a H tone. English differs from Danish in that lexically prominent syllables are only pronounced with a pitch accent when the word containing them is intonationally prominent, and F_0 only differs between high and low vowels when they bear an accent (Steele 1986). The second extension measures the occurrence and size of CF_0 differences in the same prosodic contexts. So far as I know, the effects of manipulating prosodic context on these differences have never been measured, so these results, whatever they are, will be entirely novel.

Two sets of results are reported here, pilot data collected from one speaker (the author) and a larger and more comprehensive set of data collected from four other speakers. To anticipate the results, the pilot data showed that VF_0 differences vary with prosodic contexts but CF_0 differences do not. On this evidence, VF_0 differences appear to be controlled, while CF_0 differences are not. These results replicate the findings of Ladd and Silverman (1984) and Steele (1986) for VF_0 and show that they do not extend to CF_0. The investigation of the four additional speakers turned quite differently. For all four speakers, VF_0 differed consistently across prosodic contexts, and neither its presence nor size depended on prosody. CF_0 differences were small and inconsistent across speakers and contexts. These results indicate that VF_0 may instead be automatic, and leaves the question open for CF_0. In the discussion, I suggest that the speakers may not have treated the manipulation of pitch accents as signalling differences in local information content, so they may not have manipulated either VF_0 or CF_0 in ways that reflect variation in this content.

3. Experiment 1: Pilot data

3.1 Methods

The pilot data were collected from a single speaker, the author, who grew up in Detroit, Toronto, western Pennsylvania, northern Indiana, and Chicago. His speech generally reflects characteristics of the Northern Cities dialects without closely resembling any one of them.

He was recorded producing sentences containing the names, *O'Neill*, *O'Nail*, *O'Nall*, *O'Tail*, and *O'Dale*. F_0 was measured in the middle of the high, mid, and low vowels [i], [e], and [æ] in the names *O'Neill*, *O'Nail*, and *O'Nall* to assess the effects of vowel height. The middles of the vowels were estimated by eye from displays of the waveforms and spectrograms of the vowels. Following [n], vowel onset was the point where the second and higher formant frequencies abruptly shifted in frequency and increased in amplitude. Following [t^h] and [d], it was the first full period of voicing after the stop release.

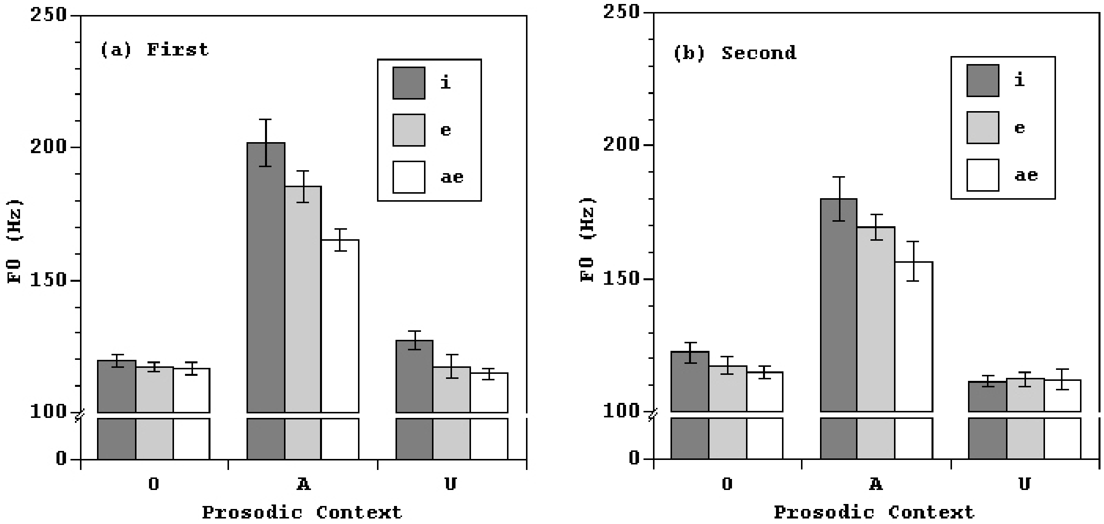
VF_0, when it differs, is expected to rank the vowels [i] > [e] > [æ]. Because F_0 is expected to have a particular value for the mid vowel [e], even if that value is intermediate between [i]'s high value and [æ]'s low one, [e] isn't really a control for assessing vowel height's effects on F_0. F_0 was also measured at the beginning and middle of the vowels following the consonants [n], [t^h], and [d] in the names *O'Nail*, *O'Tail*, and *O'Dale* to assess the effects of consonant voicing. The sonorant [n] is not expected to raise or lower F_0, so it is a genuine control consonant, with which the F_0 values after [t^h] or [d] may be compared. Praat's autocorrelation algorithm was used with the standard parameter settings to extract these values (Boersma and Weenink 1992–2005; Boersma 2001).

The names were produced in sentences of the form, *Mr. Name {sent for, has just now sent for, has just now sent impatiently for} Mr. Name by the very fastest means*. The number of words between the two names was manipulated for the purposes of another study, and the measurements reported here are collapsed across that manipulation. The sentences were pronounced with a contrastive H* pitch accent on one, the other, or neither name; when neither name was contrastively accented, the nuclear accent fell on the word 'fastest'. All possible combinations of names were produced except for those in which the name was the same in both positions. Twelve tokens of each name were obtained for each prosodic context (accented, other name accented, neither name accented). To determine whether F_0 also differs between vowels of different heights bearing a L* pitch accent, the names *O'Neill*, *O'Nail*, and *O'Nall* were also pronounced in the yes/no questions corresponding to these sentences, namely *{Did, Has, Has} Mr. Name {send for, just now sent for, just now sent impatiently for} Mr. Name by the very fastest means?* The L* occurred uniformly on the name preceding the verb. Again, twelve tokens were obtained of each name.

3.2 Vowel height

Mean F_0 values with 95% confidence intervals for measurements taken from the middle of the vowels in the names *O'Neill*, *O'Nail*, and *O'Nall* are plotted in Figure 1. The middle set of bars represents the measurements taken from accented syllables (A), the right set those in which the other name was accented (U), and the left set those in which neither name was accented (0). The top panel is for the name preceding the verb, the bottom for that following the verb.

Figure 1. Mean F0 (95% confidence intervals) for [i] (dark grey), [e] (light gray), and [æ] (white) in contexts where neither name was accented (0), the measured name bore a H* pitch accented (A), and the other name was accented (U). (a) Names preceding the verb, (b) names following the verb.

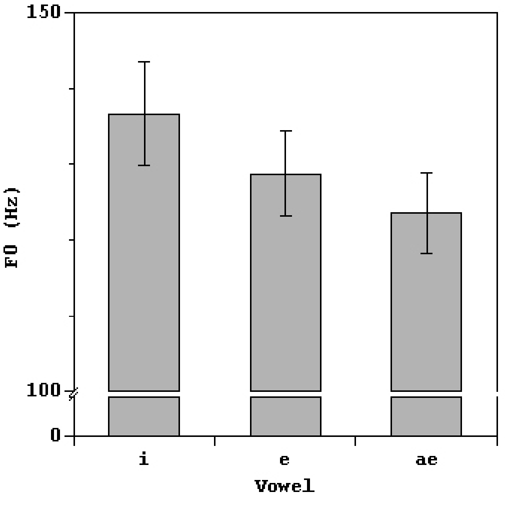


The figure shows that the means for all three vowels are reliably separated from one another when the name is accented – the mean for each vowel is outside the 95% confidence interval of the adjacent vowel(s). However, when the names are unaccented, the means for all three vowels no longer differ reliably. F_0 is still higher for [i] than the other two vowels in an unaccented name before the verb when the other name is accented and in an unaccented name after the verb when neither name is accented. F_0 doesn't differ between unaccented [e] and [æ] under any condition.

These F_0 values were used as the dependent variable in a repeated measures ANOVA in which Vowel ([i] vs. [e] vs. [æ]), Accent (accented, other accented, neither accented), and Position (before vs. after the verb) were within-subject independent variables. All main effects and interactions were significant, an outcome which shows that F_0 values depend simultaneously on accent, position, and vowel, as can be seen in the figure. The interaction of greatest interest here is that between Vowel and Accent, where $F(4,44) = 25.183$, $p < 0.001$. In pairwise comparisons of the vowels in the different prosodic contexts, F_0 is significantly higher in [i] than [e] in accented syllables ($F(1,11) = 21.116$, $p = 0.001$ and it is also significantly higher in [e] than [æ] in such syllables ($F(1,11) = 59.118$, $p < 0.001$). When neither name was accented, F_0 is also significantly higher in [i] than [e] ($F(1,11) = 6.707$, $p < 0.05$), but only marginally higher in [e] than [æ] ($F(1,11) = 3.470$, $p = 0.089$). None of these differences depended on the position of the name in which the measurement was taken, before vs. after the verb. When the other name was accented, F_0 was only significantly higher in [i] than [e] in names that preceded the verb ([i] vs. [e] x Position: $F(1,11) = 20.551$, $p = 0.001$), and F_0 didn't differ significantly between [e] and [æ] in either position ($F(1,11) = 2.424$, $p > 0.10$).

Mean F_0 values (with 95% CIs) are plotted in Figure 2 for the vowels in these names when they occurred bearing a L* pitch accent in yes/no questions.

Figure 2. Mean F0 (95% confidence intervals) for [i], [e], and [æ] in names bearing L* pitch accent.



The means for [i] and [e] are reliably outside each other's confidence intervals, while the means for [e] and [æ] are just barely within each other's confidence intervals. The F_0 values were used as the dependent variable in a repeated measures ANOVA with Vowel ([i] vs. [e] vs. [æ]) as the sole within-subjects independent variable. The main effect of vowel was significant ($F(2,22) = 8.442$, $p = 0.002$). Contrary to what is expected from the confidence intervals, planned contrasts surprisingly showed that the difference between [i] and [e] is only marginally significant ($F(1,11)$

= 3.987, p = .071), while that between [e] and [æ] is significant ($F(1,11) = 5.428$, $p = 0.04$). F_0 differs less between high, mid, and low vowels when they bear L* rather than H* pitch accents, but the L* still makes the vowels prominent enough that the three heights' F_0 values nearly separate reliably from one another. They certainly do so more reliably than when the vowels bear no pitch accent at all.

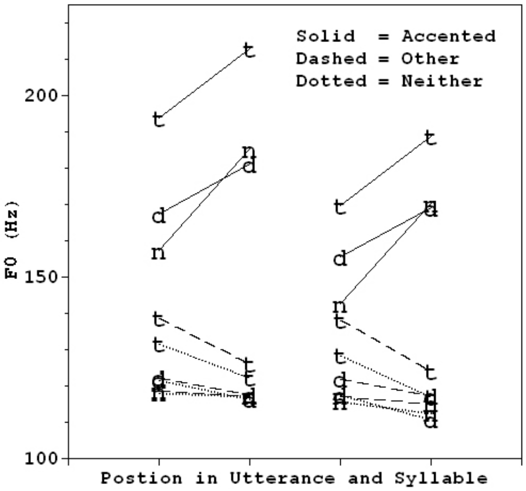
3.3 Discussion

F_0 differed significantly between high, mid, and low vowels when they bore a H* pitch accent, and nearly did so when they bore a L* pitch accent. However, when the vowel was unaccented, F_0 differences between vowels of different heights shrank or disappeared entirely. The shrinkage and disappearance of F_0 differences between vowels of different heights in unaccented syllables compared to H*- accented syllables replicates Ladd and Silverman's (1984) and Steele's (1986) findings. The finding that F_0 also differs in L* accented syllables extends those findings and shows that VF_0 differences are produced in all intonationally prominent contexts in English. If VF_0 differences were only found in H* contexts in languages such as English or only in H tone contexts in tone languages such as Yoruba, the tongue pull hypothesis would not be in jeopardy, because the larynx is pulled down when low F_0 targets are produced (Ohala 1970; Collier 1975; Ewan 1976; Erickson et al. 1995; Honda et al. 1999). Larynx lowering could counteract any upward pull on the larynx by tongue raising. However, VF_0 does differ significantly in L* contexts in English (as well as Danish). VF_0 also doesn't differ whenever the tone is H, particularly not in the context of a H– phrase accent (Steele 1986), so the pull of the tongue doesn't always affect F_0 even when larynx lowering doesn't work against it. That F_0 differs between vowels contrasting for height only when they're prominent suggests that speakers produce these differences deliberately, to exaggerate segmental contrasts in sites where information content is high, rather than the tongue pulling automatically on the vocal folds.

3.4 Consonant voicing

The results were quite different for consonant voicing's effect an F_0. Figure 3 displays mean F_0 values at the beginning and middle of vowels following the consonants [t^h], [d], and [n] in the names *O'Tail*, *O'Dale*, and *O'Nail* (confidence intervals have been omitted because they would have cluttered the display too much). The two measurements for each vowel are connected by a line whose quality identifies the initial consonant, solid = [t^h], dashed = [n], and dotted = [d]. The quality of the line identifies the prosodic context. The pairs of connected points on the left are for the name preceding the verb, those on the right for those following it.

Figure 3. Mean F0 at vowel onset (left points at the end of each line) and the middle of the vowel (right points) following [t], [d], and [n] in names preceding the verb (left) and following it (right) when the name bears a H* pitch accent (solid lines), the other name is accented (dashed lines), and neither name is accented (dotted lines).



F_0 is substantially higher at both the beginning and middle of the vowel following [t^h] than [d] or [n] when that vowel is accented, and it is also higher at the beginning of the vowel following [t^h] when the vowel is unaccented, when either the other name is accented or neither name is. F_0 didn't differ between [d] and the ostensible control consonant [n] except at the beginning of accented vowels.

These F_0 measurements were used as the dependent variable in a repeated measures ANOVA in which the within-subjects variables were Consonant ([t^h] vs. [d] vs. [n]), Accent (name accented vs. other name accented vs. neither name accented), Position (before vs. after the verb), and Location of measurement (beginning vs. middle of the vowel). All main effects were highly significant, as were the interactions Location x Accent ($F(2,22) = 944.842$, $p < .001$), Consonant x Accent ($F(4,44) = 15.978$, $p < .001$), Consonant x Location ($F(2,22) = 37.616$, $p < .001$), Consonant x Accent x Position ($F(4,44) = 5.388$, $p = .001$), and Consonant x Accent x Location ($F(4,44) = 10.921$, $p < .001$).

The Location by Accent interaction reflects the rise in F_0 between the beginning and middle of the vowel when the vowel is accented vs. its fall between the two measurement locations when the vowel is instead unaccented. The other two-way interactions, between Consonant and Accent or Location, are superseded by the significant three-way interactions in which these variables participate, so only the latter will be interpreted. Consonant by Accent by Position is significant because the top-to-bottom order of F_0 values for the name beginning with [t^h] but not the names beginning with [d] or [n] is accented > neither name > other name when it precedes the verb, but accented > other name > neither name when it instead follows the verb. Consonant by

Accent by Location is significant because in accented pronunciations F_0 is higher following $[t^h]$ than [d] or [n] at the middle as well as the beginning of the vowel, but in unaccented pronunciations, F_0 is only higher at the beginning of the vowel following $[t^h]$.

3.4 Discussion

Even though the F_0 differences following the three consonants are smaller in unaccented than accented words, F_0 does remain reliably higher following $[t^h]$ than [d, n]. This outcome is quite different from that observed in the analysis of the effects of vowel height, where F_0 differences between vowels of different heights effectively disappeared in unaccented syllables. Because the effects of consonant voicing are more impervious to the manipulations of prosody, they are more likely to be automatic than the effects of vowel height.

3.5 General discussion

These results suggest that VF_0 differences are controlled, while CF_0 differences are automatic. Their value is limited, however, because they were obtained from just a single speaker, and the author at that. My speech may have been influenced by knowledge of the hypothesis being tested. The materials were also limited in at least two important ways. First, the names in which the consonants differed weren't recorded in L* nor H– contexts. Second, [n] was a bad choice to represent sonorants. Coarticulatory nasalization extended far enough into the following vowel to make estimates of F1 and thus vowel height unreliable. These estimates are needed to determine whether the vowels are hypo-articulated for height in non-prominent contexts and thus whether hypo-articulation is the likely cause of any shrinkage of VF_0 differences. The results of Experiment 2 aren't limited in any of these ways. Four speakers who were unaware of the purpose of the study were recorded, prosodic contexts were manipulated in a more comprehensive way, and [n] was replaced by [l].

4. Experiment 2

The principal purpose of the second experiment was to test the generality of the findings obtained in the pilot experiment by collecting similar data from more speakers. The full crossing of prosodic contexts and segments is described in the methods section that immediately follows.

4.1 Methods

4.1.1 Participants

Two female and two male adult native speakers of English were recorded for this study. One of the female speakers grew up in Ohio (F1), the other in Toronto (F2). One of the male speakers grew up in Wisconsin (M1), the other in Northern California and Washington State (M2).

4.1.2 Materials

The effects of consonants were assessed by measuring F_0 in vowels following [t^h], [l], and [d] in the names *Taylor*, *Layland*, and *Daylor*, while the effects of vowel height were assessed by measuring F_0 in the middle of the vowels [i], [e], and [æ] in the first syllables of the names *Leland*, *Layland*, and *Lalland*. These names were produced in the place of 'Name' in the following dialogues:

(1) Q: What about Minnie NAME? Who will she marry in the new year?
H*
A: Minnie Name will marry Norman NAME in the new year.
L– H*

(2) Q: What about Norman NAME? Who will marry him in the new year?
H*
A: Minnie NAME will marry Norman Name in the new year.
H* L–

(3) Q: Will Minnie NAME marry Norman Name in the new year?
L* H–
A: No, Minnie NAME will marry Norman Name in the new year.
H* L–

(4) Q: Will MARIE Name marry Norman Name in the new year?
L* R H–
A: No, RENEE Name will marry Norman Name in the new year.
H* F L–

(5) Q: Will Minnie Name marry LAMAR Name in the new year?
L– L* R
A: No, Minnie Name will marry RENE Name in the new year.[5]
L– H* F

All upper case identifies instances where the name was supposed to be pronounced in contrastive focus, i.e. with a pitch accent. Other instances of the names were supposed to be pronounced without any pitch accent aligned with the stressed syllable of the name. The intended pitch accents are given below each name, along with the expected phrase accents that specify F_0 targets in unaccented pronunciations of the names. ToBI conventions are used for identifying tones (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert, and Hirschberg 1992; Pitrelli, Beckman, and Hirschberg 1994; Beckman and Ayers Elam 1997). In dialogues 4 and 5, the first syllable of the target name immediately follows the accented syllable of the first name. As a result, the syllable where the measurements were taken occurs during the rapid F_0 rise (R) from the L* pitch accent to the ensuing H– phrase accent in the questions or during the rapid F_0 fall (F) from the H* to the ensuing L– in the answers.

The five names were rotated through the positions in these dialogues so that all possible combinations were produced except those in which the same name occurs twice in the same clause. There are 20 combinations in dialogues 1, 2, 4, and 5, and 15 in dialogue 3. All the combinations in each dialogue were read together in a fixed order because pilot work showed that people couldn't read these sentences in the desired way if they were randomized. All four speakers produced the dialogues twice, reading them first from 1 through 5 and then reversing the order on the second reading.

Dialogues 1 and 2 contrast in whether the accented name in the answer is early (dialogue 2) vs. late (dialogue 1). Dialogues 4 and 5 contrast similarly in whether the post-accentual rises and falls occur early (dialogue 4) or late (dialogue 5). Dialogue 3 contrasts with dialogues 1 and 2 in that the accented name in the question in dialogue 3 is expected to bear a L* pitch accent rather than the H* pitch accent it bears in the questions in dialogues 1 and 2. Finally, dialogues 4 and 5 contrast with dialogues 1–3 in having one of the names pronounced in the context of a post-accentual rise or fall, i.e. between tonal targets rather than in the context of a tonal target.

Although all four speakers pronounced these utterances with the intended accents on the intended names, they differed from one another in how they rendered the names that were supposed to be unaccented in dialogues 1 and 5. These differences are indicated in the shaded boxes in Table 1, which lists the tonal contexts in which the names were pronounced in the five dialogues are listed for each speaker:

Table 1. Tonal contexts in which the names appeared in the five dialogues in each speaker's pronunciation. The letters 'a' and 'u' indicate intended accented vs. unaccented pronunciations, and the letter "p" indicates that the name occurs immediately after an accent ('p' for post-accentual). Transcriptions of tones follow the ToBI conventions, except that 'R' and 'F' indicate contexts during which F_0 rises or falls between flanking tonal targets. Shaded cells pick out contexts where speakers differed from one another in their choice of tones.

Dialogue		M1		M2		F1		F2	
1	Q	—	a=H*	—	a=H*	—	a=H*	—	a=H*
	A	u=!H*	a=H*	u=H*	a=H*	u=H*	a=H*	u=L–	a=H*
2	Q	—	a=H*	—	a=H*	—	a=H*	—	a=H*
	A	a=H*	u=L–	a=H*	u=L–	a=H*	u=L–	a=H*	u=L–
3	Q	a=L*	u=H–	a=L*	u=H–	a=L*	u=H–	a=L*	u=H–
	A	a=H*	u=L–	a=H*	u=L–	a=H*	u=L–	a=H*	u=L–
4	Q	p=R	u=H–	p=R	u=H–	p=R	u=H–	p=R	u=H–
	A	p=F	u=L–	p=F	u=L–	p=F	u=L–	p=F	u=L–
5	Q	u=H*	p=R	u=H*	p=R	u=H*	p=R	u=H–	p=R
	A	u=H*	p=F	u=L–	p=F	u=H–	p=F	u=L–	p=F

4.1.3 Measurements

To assess the effects of consonants, F_0 was measured at the beginning of each stressed vowel. F_0 measurements used the autocorrelation algorithm in Praat (Boersma and Weenink 2005; Boersma 2001) with the standard settings for all parameters but voicing threshold, which was reduced from 0.45 to 0.2 to capture F_0 values closer to voiceless-voiced transitions. For vowels following [l], the measurement point was 5 ms after F2 began to rise steeply from its low value during the [l]; for those following [t^h] and [d], it was the first full glottal pulse following the stop release. The stressed vowels in all the names were followed by [l], so the end of the vowel was identified as the moment when F2 reached its minimum value in the [l]. F2 fell noticeably in all cases because the target vowels were all front. To assess the effects of vowels, F_0 was measured across the middle 30 ms of each stressed vowel in the names beginning with [l]. The values from this 30 ms interval were then averaged, and the average was used in all further analyses of the vowels' effects.

Preliminary inspection of the data showed that all speakers often produced the first 5–10 periods of the vowel with breathy voice following [t^h] (see also Ní Chasaide and Gobl 1993a,b). Because F_0 is often lower in breathy than modal voice, this consonant might not raise F_0 after all.

Accordingly, the extent to which the voice quality was breathy was included as a possible predictor of F_0 values in the analyses reported here. The breathiness measure was the difference in dB between the first and second harmonic (H1–H2), which is larger for breathier voices (Ní Chasaide & Gobl, 1997). Breathiness was measured in the same intervals that F_0 was.

Finally, LPC was used with 10 coefficients to measure F1 at the middle of each vowel. The F1 measurements used in the analyses reported here are the mean calculated across the same 30 ms in the middle of the vowel as the F_0 measurements. F1 was measured to assess the extent to which vowel height differences vary as a function of prosodic context. That is, are vowel height differences larger in accented than unaccented syllables, and if so, does the size of F_0 differences between vowels of different heights correlate positively with the size of the vowel height differences as registered by F1. If F1 differences are larger and F_0 differences correlate with them, then the tongue pull hypothesis would be supported.

4.2 Results

4.2.1 Analysis

So that the data from all four speakers could be combined in a single analysis, the F_0 values were normalized by converting them to *z*-scores. The mean (μ) and standard deviation (σ) of all the F_0 values from a speaker were calculated and then used to obtain the *z*-score corresponding to each individual value:

The values of the breathiness measure (H1–H2) and F1 were converted to *z*-scores in the same way. This transformation replaces the original units of the measure, either Hz or dB, with units of standard deviation. The converted F_0 values were then used as the dependent variable in linear multiple regression model in which the categorical independent variables (predictors) and their values were:

Table 2. Categorical predictors and their values used in the multiple regression models.

Predictor	Coding			
Consonant	[t^h] = 1	[l] = 0	[d] = –1	
Vowel	[i] = 1	[e] = 0	[æ] = –1	
Accent	Accented = 1	Unaccented = 0		
Level	High = 1	Rising = 0.5	Falling = –0.5	Low = –1
Position	Early = 1	Late = –1		

Also included in some models were the interactions of the prosodic predictors with one another and with Consonant or Vowel. The *z*-transformed H1–H2 or F1 values served as a continuous predictors or dependent measures, depending on the analysis.

The values assigned to the predictors reflect expectations about how they should affect F_0

values. Thus, [t^h] and [i] are expected to elevate F_0 compared to [l] and [e], respectively, and are coded 1 compared to their 0 coding in [l] and [e], while [d] and [æ] are expected to lower F_0 compared to these intermediate sounds, and are accordingly coded –1. Because F_0 is expected to be more extreme in accented syllables, they are coded 1 while measurements taken from unaccented syllables are coded 0. This coding reflected the actual pronunciations of these words by each speaker, not the intended pronunciations, and thus takes in account any differences between speakers in their choice of accents or tones. As Table 1 above shows, all speakers pronounced words that were supposed to be accented with the intended accents, but a number of them also pronounced some words that were supposed to be unaccented with accents. Measurements taken during rising and falling intervals are also treated as unaccented for the purposes of this coding. The coding of Level as 1, 0.5, –0.5, and –1 reflects the expectation that F_0 will get progressively lower from syllables in H tone contexts, down through those in Rising and Falling contexts, and ultimately bottom out in L tone contexts. Finally, declination will have lowered F_0 little early in an utterance, so an early measurement is coded 1, while a later measurement, in a syllable that has likely undergone more declination, is coded –1.

Consonant, Vowel, Accent, etc. are called ‘predictors’ because it’s expected that the measured F_0, H1–H2, or F1 values can be predicted from what the consonant or vowel is, whether the syllable is accented, etc. Regression is a statistical technique that determines how strongly a measured value depends on a predictor and how much that measured value is predicted to change for each change in the value of the predictor. The strength of the dependence between the measured value and the predictor is represented by the proportion of the variation in the measured value that is accounted for by the predictor (henceforth the ‘proportion of variance’). The size and direction of the predicted change in the measured value is represented by a coefficient. For example, a regression analysis with the Vowel predictor accounts for 0.026 (2.6%) of the variance in F_0. This proportion seems rather small, but it’s significant ($F(1,1729) = 46.776$, $p < 0.001$). The value of the Vowel coefficient is 0.201, which indicates that F_0 is 0.201 standard deviation units higher in [i] than [e] and likewise 0.201 standard deviation units lower in [æ] than [e]. This coefficient is significantly different from 0 – its 95% confidence interval spans 0.143 to 0.258.

Simple regression models could be built of the data, in which the effect of each predictor is considered alone, but the question of interest here is whether the effects of segmental properties on F_0 depend on the prosodic context in which the segments are pronounced. For that reason, multiple regression models must be built instead. These models not only include the individual segmental and prosodic predictors, but also interactions between them. To assess the strength of the interaction, the first multiple regression model built includes the segmental and prosodic predictors, but not their interaction. Such a model is obtained by adding the Accent predictor to the simple regression model described in the preceding paragraph where Vowel was the only predictor. The proportion of variance accounted for once Accent is added to the model increases to 0.106, a significant improvement on the 0.026 accounted for by the Vowel predictor alone (partial $F(1,1728) = 154.181$, $p < 0.001$). This result is unsurprising, as we expect Accent to influence F_0 substantially. The coefficient for the Accent predictor is 0.576, which indicates that F_0 increases by 0.576 standard deviation units when a syllable is accented over when it’s not. The interaction between these two predictors, Vowel x Accent, can now be added to the model. The proportion of variance accounted for increases to 0.109, or only 0.003 more than that accounted for before this interaction was added to the model. This appears to be a very small increment indeed, but it’s again significant (partial $F(1,1727) = 4.929$, $p = 0.027$). The way the accent’s presence affects the vowel’s influence on F_0 is revealed by the coefficient for this interaction, which is 0.126. F_0 is

0.126 standard deviation units higher when vowel is [i] and accented than when it's [e] and accented, and 0.126 standard deviation units lower when the vowel is [æ] and accented. The values of the coefficients for the Vowel and Accent predictors differ somewhat in this model from their values before the Vowel x Accent interaction was added: Vowel = 0.145 and Accent = 0.575. The shrinkage in the value of the Vowel coefficient shows that some of the effect of Vowel in the model without the Vowel x Accent interaction was due to the influence of the accent on the vowel and not to the vowel alone. These coefficients continue to represent the effects of vowel and accent that are independent of their interaction with one another. The F_0 value predicted by this model for a particular vowel in a particular prosodic context can be obtained from:

(1) $$F_0 = C + 0.145\{1,0,-1\} + 0.575\{1.0\} + 0.126\{1,0,-1\}$$

where C is a constant, equal to –0.307 in this model, and the appropriate values are chosen from between the curly brackets for that vowel in that prosodic context. For example, if the vowel is [æ] and the prosodic context is accented, then the outcome is:

(2) $$-0.003 = -.307 + 0.145 * -1 + 0.575 * 1 + 0.126 * -1$$

A final point is of some importance. Models are built hierarchically such that each successive model include all the predictors in the previous model. This is the only way that the significance of increases in the proportion of variance accounted for can be evaluated.

In the first analyses, a hierarchy of five models was evaluated. The hierarchy started with a segmental predictor, Consonant or Vowel, because F_0 differences resulting from segmental contrasts are the cynosure of this paper. The breathiness measure was added next because it, too, represents a segmental property that might influence F_0. The next two models added the individual prosodic predictors, Accent, Level, and Position, and then their interactions, Accent by Level, Accent by Position, Level by Position, and Accent by Level by Position, because differences in prosodic context are expected to have very large effects on F_0 that are independent of whatever effects segmental properties have on F_0. (This expectation has already been confirmed by the example analysis described above.) These independent effects must be accounted for before assessing the extent to which segmental effects depend on the prosodic context, which was done at the fifth level of the hierarchy by adding the interactions between the segmental predictor, Consonant or Vowel, and the prosodic predictors, Accent, Level, and Position.

Because it will frequently be necessary to refer to the effects of interactions among prosodic predictors in determining which manipulations significantly affected F_0, the values of the predictors representing these interactions are listed here. The values in each cell are the products of the values of the individual predictors. For example, the upper left cell in the Accent by Level column represents the case where the values of both the Accent and the Level predictors are 1, and thus their product is 1. A look back at Table 2 shows that this is the case when a syllable bears a H* pitch accent. Similarly, the lower left cell in this column represents the case where the value of the Accent predictor is 1 but the value of the Level predictor is –1, and their product is thus –1. This is the coding for syllables bearing a L* pitch accent. The products in the second and third rows in the left hand column are in parentheses because there were no cases where a syllable was accented (Accent = 1) and F_0 was rising or falling (Level = 0.5 or –0.5). All the values in the right hand side of this column are 0 because the value of the Accent predictor is 0 for all unaccented syllables, and thus its product with any value for the Level predictor is 0. Values in the other cells in this table are to be interpreted in the same way.

Table 3. Values of predictors representing the interactions among prosodic variables. The values 0.5 and –0.5 are listed in parentheses for the Accent by Level and Accent by Level by Position interactions, because Level only takes on these values in words pronounced during a rapid rise or fall in F_0 between targets and these words are all unaccented in these materials. Acc = accented, Un = unaccented.

Accent x Level		Accent x Position		Level x Position		Acc x Lev x Pos Acc (Un = 0)	
Acc	Un	Acc	Un	Early	Late	Early	Late
1		1		1	–1	1	–1
(0.5)		–1	0	0.5	–0.5	(0.5)	(–0.5)
(–0.5)	0			–0.5	0.5	(–0.5)	(0.5)
–1				–1	1	–1	1

The values of any interactions between the segmental and prosodic predictors, including the interactions between prosodic predictors listed in Table 3, will be the same as the values of the prosodic predictor when the Consonant or Vowel predictor equals 1 (when the consonant is $[t^h]$ or the vowel is [i]); they will be 0 when the Consonant or Vowel predictors have that value (when the consonant is [l] or the vowel is [e]), and their signs will be reversed when the Consonant or Vowel predictors equal –1 (when the consonant is [d] or the vowel is [æ]).

Tables IV and VI show the proportion of variance accounted for at each stage in the hierarchy for the analyses of the consonant and vowel data, respectively. Cross-cutting the hierarchy of models in these tables, as well as those showing the coefficients for each predictor (Tables V and VII), is a series of models that differ in which cases are included in the analysis. The hierarchy in the first column of each table includes all cases, while that in the second column excludes cases whose exclusion from the model would change the absolute value of the standardized difference in fit by more than $2*(p/n)^{1/2}$, where the number of parameters (p) = 16 and n = the number of cases. As n is somewhat greater than 1700, this criterion is somewhat greater than 0.19 (see the tables for precise values). These cases are outliers and excluding them naturally improves the fit of the models.[6] The remaining columns in these tables show the effects of removing all the data provided each speaker in turn. Jackknifing the analyses like this shows whether the proportions of variance and coefficient values change substantially when one or another speaker is left out. If they do, then the speakers differ from one another.

4.2.2 Consonant voicing

Table 4 shows that Consonant alone accounts for very little of the variance in F_0 values, from a proportion less than .01 to one just over .02. The proportion of variance accounted for by Consonant is about twice as large when one of the female speakers' data is left out than when one of male speakers' data is. This indicates the effects of consonant are greater in the male than the female speakers' data. Adding the breathiness measure, H1–H2, changes the proportion of variance accounted for very little in any analysis except that where speaker M2 is left out, where the proportion accounted for increases nearly sixfold compared to the model in which H1–H2 isn't included. F_0 is apparently far more dependent on breathiness for the other three speakers than M2. The proportion of variance accounted for jumps by an order of magnitude or more once the prosodic predictors are included. This isn't at all surprising as the tones introduced by intonation are expected to influence F_0 to a much greater extent than the segments are. Adding the interactions among the prosodic predictors also boosts the proportion of variance accounted for by a good bit, adding from just under 0.12 to over 0.14. By comparison, the increment achieved by adding all the interactions between Consonant and the prosodic predictors is tiny, ranging from less than 0.01 to just 0.015.

Table 4. Proportion of variance accounted for (R^2) by successive models of CF_0: (1) constant and Consonant alone, (2) model 1 plus the breathiness measure (H1–H2), (3) model 2 plus the individual prosodic predictors, Accent, Level, and Position, (4) model 3 plus the interactions between prosodic predictors, and (5) model 4 plus the interactions between Consonant and the prosodic predictors. The first column includes all cases. In the second and subsequent columns, cases are excluded whose exclusion changes the absolute value of the standardized difference in fit by more than 0.19346 (= $2*(p/n)^{1/2}$, where the number of parameters p = 16 and n=1710). In the third and subsequent columns, one speaker in turn is excluded from the model. The numbers in parentheses at the top of each column are the number of cases in the models.

Model		all (n=1710)	|dfit| < .19346 (1607)	– M1 (1200)	– M2 (1201)	– F1 (1220)	– F2 (1200)
1	constant, Consonant (C)	.008	.013	.007	.010	.021	.018
2	1+H1–H2	.009	.013	.013	.059	.021	.026
3	2 + Accent (A), Level (L), Position (P)	.504	.582	.563	.602	.605	.590
4	3 + AxL, AxP, LxP, AxLxP	.640	.717	.713	.724	.724	.733
5	4 + CxA, CxL, CxP, CxAxL, CxAxP, CxLxP, CxAxLxP	.649	.728	.723	.736	.733	.748

In Table 5 are listed the values of the coefficients (βs) for each predictor obtained in model 5. Values in shaded cells aren't significant ($p > 0.10$), parenthesized values are at best marginally significant ($0.10 > p > 0.05$), and all others are significant ($p < 0.05$). No entries are given for two interactions, Accent x Level x Position and Consonant x Accent x Position for the model in which speaker F1 is left out because the values of these predictors are not independent of those of other predictors when this speaker's data is left out. The values of the coefficients for the prosodic predictors and their interactions will be discussed first because they are more consistent across models.

The effects of the prosodic variables themselves are rather complex, but because they are not the principal concern of this paper, they'll be described only briefly. The coefficients for both Level and Position are consistently positive, which indicates that F_0 is higher when the F_0 target is higher and early in the utterance. Unexpectedly, the coefficients for Accent are consistently negative because the L* pitch accent and H– phrase accent in the yes-no questions in dialogues 3–5 differ more than the H* pitch accent and L– phrase accent do elsewhere. The negative sign of the coefficients for the Level by Position interaction reflects the lowering of higher H and R F_0 targets and the raising of the lower F and L targets early in the utterance compared to late. The coefficients for the Accent by Level interaction are also consistently negative, because F_0 is lower in syllables bearing both H* and L* pitch accents compared to those bearing H– and L– phrase accents. In three out of the four jackknifed models, the coefficient for the interaction of Accent with Position is also negative, which shows that F_0 is lowered in early accented syllables but raised in the late ones compared to unaccented syllables in those positions. Finally, the coefficient for the Accent by Level by Position interaction is consistently positive; this is evidence that that H* and L* targets differ more from one another when they occur early than late.

Compared to these prosodic predictors, Consonant and H1–H2 have small effects. For Consonant, the effects are moreover only significant in one model, that from which speaker M2's data are excluded. In that model, the coefficient for Consonant is negative, which indicates that F_0 is lower following [t^h] compared to [l] and higher following [d] compared to [l]. These consonants were expected to affect F_0 in the opposite directions. The coefficient for H1–H2 is positive in all models, which again apparently reverses the expected effect, i.e that F_0 is lower in breathier vowels. However, if voice quality is breathier following [t^h] than the other consonants, then F_0 should be higher following this consonant, as expected. In other words, [t^h] doesn't raise F_0 relative to [l] by itself, but only to the extent that it makes the voice quality of the following vowel breathy.

This hypothesis is confirmed by two further analyses not reported in detail here. In the first, the roles of F_0 and H1–H2 were reversed, with F_0 used as a continuous predictor of H1–H2. (Only models from which outliers were excluded by the criterion described above are considered.) In all models, the coefficients for Consonant are significantly positive: 0.926 (all speakers), 0.882 (– M1), 1.007 (– M2), 0.894 (– F1), and 0.940 (– F2), which indicates that compared to after [l], voice quality is markedly more breathy following [t^h] and markedly less so following [d]. The coefficients for F_0 in these models were also consistently positive, although not always significantly so: 0.103 (all speakers), 0.000 *ns* (– M1), 0.265 (– M2), 0.116 (– F1), and 0.032 *ns* (– F2). This finding merely confirms the positive correlation between F_0 and H1–H2 observed in the models where the roles of these measures were reversed. Larger H1–H2 is more reliable at the beginning of a vowel following [t^h] for these speakers than higher F_0, but because F_0 is raised more when the vowel onset is breathier, higher F_0 is only somewhat less reliable.[7]

In the second analysis, H1–H2 was simply left out as a predictor of F_0 values. The expectation was that the coefficients representing the effect of Consonant would turn consistently positive, as the effect that's represented by H1–H2 in the models presented in detail here is taken up by this predictor. This expectation was met, but the coefficient for this predictor was significant in only one model and marginally significant in just one more: 0.037 *marginal* (all speakers), 0.029 *ns* (– M1), 0.012 *ns* (– M2), 0.075 (– F1), and 0.031 *ns* (– F2). Categorical differences between the three consonants don't entirely capture the detailed differences in their pronunciations represented by the continuously varying H1–H2 values.

The only significant interaction between Consonant and any prosodic predictor is that with Level, whose coefficients are consistently negative in all models. Given the values of this predictor (Table 3), the negative coefficient shows that F_0 values in higher H and R contexts converge with those in lower F and L contexts following [t^h] but diverge in these two contexts following [d].

Table 5. Model coefficients for CF_0: The first column includes all cases, the second and subsequent columns including only those for which absolute value of the standardized difference in fit value <= 0.19346 (= $2*(p/n)^{½}$, where the number of parameters (p) = 16 and n=1710), and the third and subsequent columns exclude cases for each speaker in turn (in each of these models cases are excluded whose standardized difference in fit value > 0.19346). The numbers in parentheses at the top of each column are the number of cases in the models. Cells containing non-significant coefficients are shaded, and coefficients which are marginally significant (0.05 < p < 0.10) are in parentheses.

Predictor	all (*n*=1710)	|dfit| < .19346 (1607)	– M1 (1200)	– M2 (1201)	– F1 (1220)	– F2 (1200)
constant	.173	.168	.150	.130	.196	.208
Consonant	–.033	.009	.030	–.084	(.046)	.022
H1–H2	.071	.030	.000	.095	.032	.010
Accent	–.153	(–.215)	–.243	–.178	–.790	–.218
Level	.776	.800	.784	.727	.807	.911
Position	.107	.093	.052	.110	.127	.086
AxL	–.529	–.557	–.446	–.429	–.027	–.785
AxP	–.282	–.267	–.168	–.263	.241	–.305
LxP	–.424	–.401	–.425	–.424	–.373	–.335
AxLxP	.553	.538	.450	.573		.516
CxA	.084	.090	.095	.124	(.107)	.061
CxL	–.122	–.120	–.121	–.102	–.091	–.150
CxP	–.017	–.002	.019	–.010	–.015	–.009
CxAxL	.144	.184	.146	.133	.122	.286
CxLxP	.033	.037	.029	.073	.040	.013
CxAxP	.038	–.022	–.100	–.012		.011
CxAxLxP	–.036	–.026	.036	–.062	–.003	–.063

4.2.3 Vowel height

Table 6 shows that the proportion of variance accounted for by the model in which Vowel is the only predictor (other than the constant) is not large, but a look back at Table 4 shows that this proportion is 1.5 to 2 times as large as it was for Consonant. H1–H2 was added as a predictor in the next model only to make the model hierarchy as parallel as possible to that for consonant voicing. There was no expectation that breathiness would vary with vowel height nor that it might predict F_0 values in the middle of the vowel. Nonetheless, the proportions of variance accounted grew substantially in all the models once H1–H2 was added. The proportion of variance accounted for grows further between each of the remaining steps in the hierarchy in much the same way as it did in the hierarchy of models for consonant voicing effects, except that the growth between models 3 and 4, when the interactions among prosodic predictors are added is only about one half to two thirds as large for the vowel height as the consonant voicing hierarchy: increments of just 0.06–0.11 compared 0.12–0.14. The growth in the proportion of variance accounted for by adding the interactions between Vowel and the prosodic predictors is also smaller: 0.001–0.004 vs. 0.01–0.015. Although this is a difference by an order of magnitude, the increments are tiny in both cases.

Table 6. Proportion of variance accounted for (R^2) by successive models of VF_0: (1) constant and Vowel alone, (2) model 1 plus the breathiness measure (H1–H2), (3) model 2 plus the individual prosodic predictors, Accent, Level, and Position, (4) model 3 plus the interactions between prosodic predictors, and (5) model 4 plus the interactions between Vowel and the prosodic predictors. The first column includes all cases; in the second and subsequent columns, cases are excluded in which the absolute value of the standardized difference in fit > 0.19206 (= $2^*(p/n)^{½}$, where the number of parameters $p = 16$ and n=1735); and in the third and subsequent columns one speaker in turn is excluded from the model. The numbers in parentheses at the top of each column are the number of cases in the models.

Model		all (*n*=1735)	|dfit| < .19206 (1631)	– M1 (1218)	– M2 (1223)	– F1 (1228)	– F2 (1224)
1	constant, Vowel (V)	.026	.025	.018	.024	.029	.027
2	1+H1–H2	.089	.097	.034	.387	.080	.042
3	2 + Accent (A), Level (L), Position (P)	.649	.704	.674	.762	.728	.694
4	3 + AxL, AxP, LxP, AxLxP	.728	.794	.777	.821	.806	.805
5	4 + VxA, VxL, VxP, VxAxL, VxAxP, VxLxP, VxAxLxP	.731	.796	.781	.822	.808	.808

Even a quick glance at Table 7 shows that coefficient values are more consistent across models for vowel height than they were for consonant voicing. First of all, the coefficient for Vowel is consistently significantly positive, as is that for H1–H2: F_0 is higher for higher and breathier vowels. The sign and significance of the coefficients for Level and Position show that F_0 is also reliably higher for vowels pronounced in higher tonal contexts and as well as being higher early in the utterance than late.

In all models, two of the interactions between prosodic predictors are significantly negative, those between Level and Accent or Position, while the coefficient for the three way interaction between these predictors is significantly positive, in all but the model from which F2's data were excluded. The signs of these coefficients match those obtained in the analysis of the consonant voicing effects and reflect the same influences of the prosodic predictors. The negative sign of the coefficient for the Level x Accent predictor shows that F_0 is lower in accented syllables with H* targets and higher in those with L* targets. The negative sign of the Level x Position predictor's coefficient indicates that higher H and R targets are closer to lower F and L targets early in the utterance but farther apart late. Finally, the positive sign of the coefficient for the Level x Accent x Position predictor indicates that an accent exerts the reverse effect, pushing higher targets away from one another early in the utterance and pulling them together late.

Given the truly tiny increments in proportion of variance accounted for in going from model 4 to model 5, when the interactions between Vowel and the prosodic predictors were added, it should come as no surprise that with just one exception in one model, none of the coefficients for these interactions are significant in any model. Whatever vowel height differences do to F_0 in these data, they do so independently of prosody.

Table 7. Model coefficients for VF$_0$: The first column includes all cases, the second and subsequent columns including only those for which absolute value of the standardized difference in fit value <= 0.19206 (= $2*(p/n)^{½}$, where the number of parameters (p) = 16 and n=1735), and the third and subsequent columns exclude cases for each speaker in turn (in each of these models cases are excluded whose standardized difference in fit value > 0.19206). The numbers in parentheses at the top of each column are the number of cases in the models.

Predictor	all	|dfit| < .19206	– M1	– M2	– F1	– F2
	(n=1735)	(1631)	(1200)	(1223)	(1228)	(1224)
constant	–.164	–.158	–.165	–.220	–.146	–.118
Vowel	.179	.187	.171	.210	.187	.171
H1–H2	.119	.119	.071	.255	.108	.085
Accent	–.038	–.077	–.065	–.059	–.101	–.090
Level	.964	1.001	.949	.852	1.000	1.215
Position	.043	.042	.016	.061	.064	.031
AxL	–.385	–.430	–.346	–.272	–.370	–.752
AxP	–.102	–.103	–.084	–.073	–.097	(–.112)
LxP	–.342	–.326	–.362	–.296	–.334	–.205
AxLxP	.254	.219	.214	.173	.245	.111
VxA	.053	.020	.029	–.009	.048	.022
VxL	.028	.021	.042	.004	.020	–.007
VxP	–.010	.011	.013	.030	.010	.002
VxAxL	..005	.009	.008	–.044	.016	.043
VxLxP	–.029	–.012	–.002	.015	–.011	–.037
VxAxP	.090	.059	.064	–.011	.048	.099

5. Discussion

These results are clearly at odds with expectations arising from earlier studies and the pilot data in Experiment 1. The pilot data showed that F$_0$ was reliably higher following [t^h] than [d] regardless of prosodic context, but that F$_0$ only differed reliably between vowels of different heights in accented syllables. This latter finding replicated those reported for German by Ladd and Silverman (1984) and American English by Steele (1986). In the more extensive data collected from four speakers in Experiment 2, F$_0$ doesn't differ consistently between [t^h] and [d], although it is higher when the vowel onset is breathier, and the vowel onset is reliably breathier when preceding consonant is [t^h] than when it's [d]. Moreover, the robust effects of vowel height on F_0 in the middle of the vowel obtained in this study are statistically independent of prosody. Given that I was the speaker in Experiment 1 and was aware of and interested in the outcome, only the results obtained in Experiment 2 are of probative value.

The results Ladd and Silverman's and Steele's studies motivated the choice of method in these experiments: manipulating a segment's prosodic context to see whether its effects on F_0 remained unchanged. If its effects did not change, then the segment's effects were to be interpreted as automatic, but if they occurred in accented, i.e. prominent syllables alone, then they were instead to be interpreted as controlled. If the segment's effects were larger in prosodically prominent contexts, as they were in Ladd and Silverman's and Steele's results, their larger size could be interpreted as exaggeration of the contrast between that segment and a minimally contrasting one in a site where the local information content is high. This exaggeration was expected to occur when a vowel bears a pitch accent, particularly a contrastive pitch accent, because pitch accents pick out the words with high local information content.

For both vowel height and consonant voicing, these data suggest that neither vowel nor consonant pronunciations differ as function of information content. The essential difference between the effects of vowel height vs. consonant voicing is that VF_0 differences are consistently present in all contexts, while CF_0 differences are more variable and also dependent on breathiness. Otherwise, by the reasoning that motivated the choice of method in this study, the results indicate that the effects of both segmental contrasts on F_0 are automatic and not controlled.

What is to be done? The first thing to determine is whether these speakers pronounced the vowels and consonants differently in other ways between prominent vs. not-prominent contexts. If they did not, then manipulating the prosodic context was (surprisingly) unsuccessful at eliciting differences in pronunciation and other manipulations or methods must be tried.

5.1 Vowel height

For vowel height, the F1 values at the middle of the vowel are the obvious choice for this purpose. If the articulations that implement vowel height contrasts are exaggerated, then F1 differences should be larger in prominent contexts than they are in contexts that aren't prominent. Accordingly, multiple regression analyses were run in which the dependent variable was the mean F1 spanning the middle 30 ms of each vowel. The same predictors and model hierarchy were used in this analysis as in the analyses of vowel height's effects on F_0. Vowel is included in the first model in the hierarchy and this model unsurprisingly accounts for an enormous proportion of the variance, from 0.861 in the version of model 1 where F2's data are left out, to 0.877, in the version where M1's data is. The proportion of variance accounted for increases only by tiny amounts as additional predictors are added, by a total of just .009 in the version of model where F2's data are left out and by a total of just .013 in the version where M1's data is.

Table 8 is a list of the coefficients that were at least marginally significant in two or more jackknifed models. The negative coefficient for Vowel shows the expected effect of vowel height on F1: it's higher in lower vowels than higher ones. The positive coefficient for H1–H2 shows that breathier vowels have higher F1 values. This result is surprising given that the coefficient for H1–H2 was also positive in the analysis of vowel height's effects on F_0 reported above. That analysis showed that F_0 is not only higher in breathier vowels but also in higher vowels, which creates the expectation that lower vowels, i.e. those with higher F1 values, should be less not more breathy.

The effects of vowel height have, however, already been accounted for by the predictor Vowel, so these coefficients represent those effects of H1–H2 on F_0 or F1 that are independent of vowel height, and their values show that both F_0 and F1 increase with breathiness.

F1 is also higher in accented than unaccented syllables, suggesting that the mouth is opened wider in accented vowels (Harrington, Fletcher, and Beckman 2000; Erickson 2002), but this effect is significant in only one model and marginally significant in one other. The coefficients representing the interactions of Level with Accent or Position are both negative but significantly so in only one jackknifed model for Level by Accent and in only two for Level by Position. Given the values of the Level by Accent predictor (Table 3), the sign of this coefficient indicates that F1 is lower, i.e. the vowel is higher when F_0 is higher. The negative coefficient for the Level by Position predictor indicates that F1 values in syllables in higher H and R vs. lower F and L contexts are pulled together early in the utterance but they are pushed apart late.

The coefficients representing the interaction between Vowel and Accent are more consistently significant. They are also uniformly negative, which means that an accented high vowel has a lower F1 value and an accented low vowel has a higher F1 than these vowels would have if they were unaccented. That is, height differences are exaggerated in accented syllables compared to unaccented ones. Yet F_0 differences are no greater between high and low vowels when accented than unaccented – the Vowel x Accent interaction wasn't even marginally significant in the analysis of vowel height's effect on F_0. The presence of an accent does alter how much vowels differ in tongue height, but despite doing so its presence doesn't alter how much they differ in F_0.

Table 8. Model coefficients for F1: The first column includes all cases. The second and subsequent columns include only those for which absolute value of the standardized difference in fit value <= 0.19206 (= 2*(p/n)½ , where the number of parameters p = 16 and n=1735). The third and subsequent columns exclude cases for each speaker in turn (in each of these models cases are excluded whose standardized difference in fit value > 0.19206). The numbers in parentheses at the top of each column are the number of cases in the models.

Predictor	all (n=1735)	|dfit| < .19206 (1631)	– M1 (1200)	– M2 (1223)	– F1 (1228)	– F2 (1224)
constant	.110	.113	.100	.143	.089	.110
Vowel	–1.249	–1.278	–1.264	–1.284	–1.306	–1.267
H1–H2	.028	.055	.036	.093	.048	.052
Accent	.098	.204	.223	.155	.064	(.201)
AxL	–.109	(–.191)	–.212	(–.182)	.007	–.182
LxP	–.029	–.035	–.042	–.034	–.052	–.008
VxA	–.322	–.266	(.046)	–.264	–.277	–.265

This result does therefore support the hypothesis that F_0 differences between vowels of different heights are automatic. However, a problem remains. The most plausible explanation of these differences has been that the raising the tongue somehow stretches the folds, while lowering it somehow slackens them. This explanation is supported by Ohala and Eukel's (1987) finding that F_0 differences between vowels of different heights were larger when the jaw was propped open by a bite block and speakers had to raise the tongue more independently of the jaw to get it close enough to the palate to successfully produce a high vowel. The problem is this: the presence of an accent has just been shown to exaggerate F1 differences between high and low vowels, and this exaggeration is presumably achieved by raising the tongue more in high vowels and lowering it more in low vowels, yet despite these larger differences in tongue height, F_0 differences are no greater in accented than unaccented vowels.

Why doesn't raising and lowering the tongue more also raise and lower F_0 more if F_0 differences between vowels are a byproduct of the tongue pulling on the vocal folds? The answer may be that the larger F1 differences are a byproduct of raising or lowering the *jaw* more in accented than unaccented syllables, and that the tongue itself is not raised or lowered more independently of the jaw than it is in unaccented syllables. Harrington et al. (2000) and Erickson (2002) both show that the effect of an accent is to increase jaw movement and that the tongue actually moves independently of the jaw to ensure that the desired vowel target is still reached. So the finding that the size of F1 but not F_0 differences varies with the presence of an accent may not be such a serious problem after all.

5.2 Consonant voicing

For consonant voicing, voice onset time (VOT) might differ more in prominent than non-prominent contexts. *Z*-transformed VOT values from the four speakers were used as the dependent variable in the same model hierarchy as was used to analyze the effects of consonant voicing on F_0 at vowel onset, except that the Consonant predictor had only two values, 1 for [t^h] and –1 for [d], because no VOT measurements can be made on [l].

Unsurprisingly, the first model in the hierarchy, whose only predictor is Consonant, accounts for a very large proportion of the variance in VOT, ranging from 0.801 for the model which includes all the cases to 0.891 for the model that excludes data from speaker M1. The increments in the variance accounted from adding all the other predictors are modest, ranging from just 0.019 to 0.033. Again, the model including all the cases accounts for the smallest proportion of the variance at step 5 in the hierarchy, 0.839, and the one excluding speaker M1 accounts for the most at this step, 0.910.

Table 9 lists the coefficients that were at least marginally significant in two or more of the jackknifed models. The results are strikingly uniform and straightforward. Naturally, the coefficient for Consonant is positive as VOT is longer in [t^h] than [d]. The coefficient for H1–H2 is consistently negative, which indicates that VOT is shorter when following vowel onsets are breathier. This tradeoff is not surprising because VOT and breathiness are both correlates of glottal spreading. Tokens differ from one another in whether voicing begins while the glottis is still appreciably spread, in which case the vowel will begin with a breathier voice quality and VOT will be shorter, or voicing doesn't begin until the glottis is no longer particularly spread, in which case the vowel will begin with a more modal, less breathy voice quality following a longer VOT. The positive coefficient for Accent shows that VOT values are longer in accented than unaccented syllables. The coefficient for Level is consistently negative, although only significantly so in two of the jackknifed models. The sign of this coefficient shows that VOT values are shorter in H or R than F or L contexts. The positive coefficient for Position indicates that VOT values are longer early in the utterance than late. Level and Position interact significantly and the coefficient for this interaction is positive. VOT values are thus longer in H and R than F or L contexts early in the utterance, but shorter late. Finally, a single interaction between the segmental and prosodic predictors is significant, Consonant x Position, and its coefficient is consistently positive. VOT values differ more between [t^h] and [d] early in the utterances than late.

The significance of the coefficient for Accent doesn't indicate that speakers exaggerate VOT differences between [t^h] and [d] in accented syllables compared to unaccented ones. They merely make them longer. Exaggeration of differences would only be indicated by a significantly positive coefficient for the interaction between Consonant and Accent. Though uniformly positive, this interaction's coefficient was at best marginally significant and that in just one model, the one including all the cases. VOT differences between [t^h] and [d] are only consistently exaggerated early in the utterance compared to late, as shown by the significantly positive coefficient for the Consonant by Position interaction. An early position in the utterance would often, perhaps even typically, have higher information content than a later position since content is less predictable early, but an early position is not one of greater prominence like an accented syllable. These speakers don't exaggerate VOT differences between [t^h] and [d] in prominent syllables any more than they do F_0 differences in the vowels following these consonants.

Table 9. Model coefficients for VOT: The first column includes all cases. The second and subsequent columns include only those for which absolute value of the standardized difference in fit value <= 0.23767 (= $2*(p/n)^{½}$), where the number of parameters $p = 16$ and n=1133). The third and subsequent columns exclude cases for each speaker in turn (in each of these models cases are excluded whose standardized difference in fit value > 0.23767). The numbers in parentheses at the top of each column are the number of cases in the models.

Predictor	all (*n*=1133)	|dfit| < .23767 (1067)	– M1 (812)	– M2 (788)	– F1 (805)	– F2 (796)
constant	–.131	–.117	–.111	–.098	–.117	–.129
Consonant	.867	.877	.892	.913	.889	.849
H1–H2	–.070	–.066	–.061	–.109	–.059	–.062
Accent	.220	.180	.171	.177	.285	.193
Level	–.043	–.033	–.043	–.025	–.016	–.044
Position	.045	.051	.036	.045	.064	.057
LxP	.047	.058	.057	.051	.056	.067
CxP	.049	.054	.036	.053	.064	.068

5.3 Summary and concluding remarks

Taken together, these findings show that prominence doesn't prompt speakers to increase the distinctiveness of the [voice] contrast between these two consonants, at least as measured by VOT and F_0. These findings also dovetail with the conclusions drawn from the further investigation of VF_0 differences, which also weren't exaggerated in prominent syllables, although another correlate of vowel height contrasts, F1, was. In short, unlike Ladd and Silverman (1984) or Steele (1986), I was unable to get manipulations of prosody to alter articulations of segments in ways that would affect F_0, for either vowel height or consonant voicing.

Again, the question can be asked: what is to be done? The answer is to try manipulating prosodic context such that one can be confident that the presence or absence of a pitch accent does reliably correlate with the waxing and waning of local information content. The materials used in this study were designed to produce such a correlation, and the speakers certainly hyper-articulated accented words in these utterances and hypo-articulated unaccented ones. However, those differences in the extent of hyper- vs. hypo-articulation may not have reflected variation in local information content because each sentence was simply a permutation of the names. New materials are being designed to correct this error. An approach of this kind is a more sensitive test of the hypothesis that speakers control F_0 to enhance contrasts in syllables whose information content is high than encouraging them to hyper-articulate globally by having them try to transmit their messages in the context of background noise or other signal degradations.

Even if this error can be corrected, the results of Experiment 2 (though not Experiment 1) remain a problem for interpreting Ladd and Silverman's and Steele's findings that VF_0 differences apparently depended solely on the presence of a pitch accent. The speakers in Experiment 2 certainly produced some words with H* and L* pitch accents and just as certainly produced others without any such accents, yet VF_0 differed as much in the unaccented as the accented syllables. This finding suggests that these differences are not after all dependent on a pitch accent. Instead, in anticipation of the results of the new manipulations of prosodic context, we might conclude that these differences and the presence of a pitch accent are independently controlled means of conveying local information content.

---

1 For very helpful comments, advice, and criticisms I am grateful to the audiences at four earlier presentations of this paper, at: the Methods in Phonology Conference at Berkeley in May 2004, the Tone and Intonation in Europe Conference on Santorini in September 2004, at the Phonetics and Phonology Workshop in Tokyo in December 2004, and finally the Linguistics Department Colloquium at SUNY, Stonybrook in March, 2005. The comments and criticisms of an anonymous reviewer have also helped me a great deal in improving this paper, particularly in making its central argument more intelligible. I retain responsibility for any errors that might have survived.

2 What I consider to be plausible, even most plausible, is surely considered to be outlandish by others who've tried to explain these F_0 differences as automatic.

The domain of prominence is the phonological word in the Danish materials analyzed by Reinholt Petersen, not the larger prosodic constituent that's the domain of the pitch accents in the German and English materials analyzed by Ladd and Silverman (1984) and Steele (1986). Their results show that VF_0 differences aren't found in syllables that are merely the most prominent syllables in phonological words in these languages. The most likely explanation for this difference is that the prominent syllable in a phonological word in German and English doesn't necessarily bear a pitch accent (Beckman 1986), while a comparable syllable does bear a pitch accnt in Standard Copenhagen Danish. We might predict that a pitch accent is a necessary condition for the realization of VF_0 differences, whether that accent arises in the intonation or the lexicon. This prediction could be tested by measuring F_0 in vowels differing in height in deaccented words in Standard Copenhagen Danish. Reinholt Petersen doesn't report such measurements, nor am I aware of any other study that does so. I am grateful to the editors for drawing this prediction to my attention.

---

4 I am indebted to the reviewer and the editors for bringing this alternative to my attention.

5 The names *Rene* and *Renee* were pronounced identically. The shorter spelling is intended to indicate a man and the longer a woman, i.e. the marriages being discussed are between opposite sexes.

6 103 outliers are excluded by this criterion in the first analysis, i.e. just a little more than 6%. The number and percentage of outliers are equally small in the other analyses reported here.

7 The only other significant predictors of H1–H2 values were Position, where the coefficients were significantly negative in all models and the interactions between Consonant and Position and between Consonant and Level, where for each interaction, the coefficients were significantly positive in all but one model: Consonant x Position wasn't significant for the model from which M2's data were excluded and Consonant by Level wasn't for the model from which F2's data were excluded. The negative sign of the coefficient for Position shows that vowel onsets are breathier late in the utterance than early. The positive sign of the coefficient for the interaction of Position with Consonant shows that they are breathier for an early than a late [t^h] but breathier for a late than an early [d]. That is, the difference in breathiness between these consonants is greater early than late. Finally, the positive sign of the coefficient for the interaction of Level with Consonant is evidence that breathiness is greater following [t^h] in the higher H and R contexts but higher for [d] in the lower F and L contexts.

References

Atkinson, James E. 1978 Correlation analysis of the physiological factors controlling fundamental frequency. *Journal of the Acoustical Society of America* 63: 211–222.

Beckman, Mary E., and Gayle Ayers Elam 1997 Guidelines for ToBI labelling. (version 3.0). Manuscript. The Ohio State University.

Boersma, Paul 2001 Praat, a system for doing phonetics by computer. *Glot International* 5(9/10): 341–345.

Boersma, Paul, and David Weenink 2005 Praat: doing phonetics by computer (Version 4.3.01) Computer program. http://www.praat.org/

Cho, Taehong, and Patricia A. Keating 2001 Articulatory strengthening at the onset of prosodic domains in Korean. *Journal of Phonetics* 28: 155–190.

Collier, René 1975 Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America* 58: 249–255.

Connell, Bruce 2002 Tone languages and the universality of instrinsic F_0: Evidence from Africa. *Journal of Phonetics* 30: 101–129.

DiCristo, Albert, Daniel. J. Hirst, and Yukihiro Nishinuma 1979 L'estimation de la F_0 intrinsèque des voyelles: Etude comparative. *Travaux de L'Institut de Phonétique D'Aix-en-Provence* 6: 149–176.

Dyhr, Neils 1990 The activity of the cricothyroid muscle in the intrinsic fundamental frequency in Danish vowels. *Phonetica* 47: 141–154.

Erickson, Donna, Kiyoshi Honda, Hiroyuki Hirai, and Mary Beckman 1995 The production of low tones in English intonation. *Journal of Phonetics* 23: 179–188.

Erickson, Donna 2002 Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59: 134–149.

Ewan, William G. 1976 Laryngeal behavior in speech. Ph. D. diss., University of California, Berkeley.

Fischer-Jørgenson, Eli 1990 Intrinsic F_0 in tense and lax vowels with special reference to German. *Phonetica* 47: 99–140.

Fougeron Cecile and Patricia A. Keating 1997 Articulatory strengthening at edges of prosodic domain. *Journal of the Acoustical Society of America* 106: 3728–3740.

Halle, Morris and Kenneth N. Stevens. 1971 A note on laryngeal features. *Quarterly Progress Report, Research Laboratory in Electronics* 101: 198–213. Cambridge, MA: Massachusetts Institute of Technology.

Harrington, Jonathan, Janet Fletcher, and Mary E. Beckman 2000 Manner and place conflicts in the articulation of accents in Australian English. In *Papers in laboratory phonology V: Acquisition and the lexicon,* Michael Broe, and Janet B. Pierrehumbert (eds.), 40–51. Cambridge, UK: Cambridge University Press.

Hombert, Jean-Marie 1977 Consonant types, vowel height, and tone in Yoruba. *Studies in African Linguistics* 8: 173–190.

Hombert, Jean-Marie 1978 Consonant types, vowel quality, and tone. In *Tone: A linguistic survey*, Victoria Fromkin (ed.), 77–111. New York: Academic Press.

Hombert, Jean-Marie, John J. Ohala, and William G. Ewan 1979 Phonetic explanations for the development of tones. *Language* 55: 37–58.

Honda Kiyoshi, and Osamu Fujimura 1991 Intrinsic vowel F_0 and phrase-final lowering: Phonological vs. biological explanations. In *Vocal fold physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms*, Jan Gauffin, and Britta Hammerberg (eds.), 149–157. San Diego: Singular Publishing Group.

Honda, Kiyoshi, Hiroyuki Hirai, Shinobu Masaki, and Yasuhiro Shimada 1999 Role of vertical larynx movement and cervical lordosis in F_0 control. *Language and Speech* 42: 401–411.

Keating, Patricia A., Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu 1999 Domain-initial articulatory strengthening in four languages. *UCLA Working Papers in Linguistics* 97: 139–156.

Kingston, John 1985 The phonetics and phonology of the timing of oral and glottal events. Ph. D. diss., University of California, Berkeley.

Kingston, John 1991 Integrating articulations in the perception of vowel height. *Phonetica* 47: 149–179.

Kingston, John 1992 The phonetics and phonology of perceptually motivated articulatory coordination*. Language & Speech. Festschrift for John J. Ohala* 35: 99–113.

Kingston, John, and Randy L. Diehl 1994 Phonetic knowledge, *Language* 70: 419–454.

Kingston, John, Neil A. Macmillan, Laura Walsh Dickey, Rachel Thorburn, and Christine Bartels 1997 Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America* 101: 1696–1709.

Kingston, John, Randy L. Diehl, Cecilia J. Kirk, and Wendy A. Castelman. submitted. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*.

Ladd, D. Robert, and Kim E. A. Silverman 1984 Vowel intrinsic pitch in connected speech. *Phonetica* 41: 31–40.

Lehiste, Ilse 1970 Suprasegmentals. Cambridge, MA: MIT Press.

Liberman, Mark Y., and Janet B. Pierrehumbert 1984 Intonational invariance under changes in pitch range and length. *Language sound structure*, Mark Aronoff, and Richard Oehrle (eds.), 157–233. Cambridge, MA: MIT Press.

Löfqvist, Anders, Thomas Baer, Nancy S. McGarr, and Robin Seider Story 1989 The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85: 1314–1321.

Ní Chasaide, Ailbhe, and Christer Gobl 1993a Dynamic variation of the voice source in VCV sequences: Intrinsic characteristics of selected consonants. *Proceedings of the first review* *meeting of the Esprit Basic Research Action, No. 6975: SPEECH MAPS, 2*: 44. Grenoble: Institut de la Communication Parlée.

Ní Chasaide, Ailbhe, and Christer Gobl 1993b Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36: 303–330.

Ní Chasaide, Ailbhe, and Christer Gobl 1997 Voice source variation. In *The handbook of the phonetic sciences*, W. J. Hardcastle, and John Laver (eds.), 427–461. Oxford: Blackwells.

Ohala, John J. 1970 Control and production of speech. University of California, Los Angeles, *Working papers in phonetics* 15.

Ohala, John J. 1973 The physiology of tone. In *Consonant types and tone, Southern California Occasional Papers in Linguistics*, 1st ed, Larry M. Hyman (ed.), 1–14.

Ohala, John J., and Brian Eukel 1987 Explaining the intrinsic pitch of vowels. In *In honour of Ilse Lehiste*, Robert Channon, and Linda Shockey (eds.), 207–215. Dordrecht: Foris.

Pitrelli, John F., Mary E. Beckman, and Julia Hirschberg 1994 Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the 1994 International Conference on Spoken Language Processing*, v. 1, 123–126. Yokohama, Japan.

Reinholt Petersen, Niels 1978 Intrinsic fundamental frequency of Danish vowels. *Journal of Phonetics* 6: 177–189.

Riordan, Carol J. 1980 Larynx height during English stop consonants. *Journal of Phonetics* 8: 353-360.

Sapir, Simon 1989 The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. *Journal of Voice* 3, 44–51.

Shadle, Christine 1985 Intrinsic fundamental frequency of vowels in sentence context. *Journal of the Acoustical Society of America* 78: 1562–1567.

Silverman, Kim E. A. 1987 *The structure and processing of fundamental frequency contours*. Ph. D. diss., Cambridge University.

Silverman, Kim E. A., Mary E. Beckman, John F. Pitrelli, Mari Ostendorf, Colin Wightman, Patti J. Price, Janet B. Pierrehumbert, and Julia Hirschberg 1992 TOBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing, Vol. 2*, 867—870. Banff, Canada.

Steele, Shirley A. 1986 Interaction of vowel F_0 and prosody. *Phonetica* 43: 92–105.

Vilkman, Erkki, Olli Aaltonen, Ilkka Raimo, P. Arajävi, and H. Oksanen 1989 Articulatory hyoid-laryngeal changes vs. cricothyroid muscle activity in the control of intrinsic F_0 of vowels. *Journal of Phonetics* 17: 193–203.

Whalen, Douglas H., and Andrea G. Levitt 1995 The universality of intrinsic F_0 of vowels. *Journal of Phonetics* 23: 349–366.

Whalen, Douglas H., Bryan Gick, Masanobu Kumada, and Kiyoshi Honda 1999 Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F_0. *Journal of Phonetics* 27: 125–142.

Zee, Eric 1980 Tone and vowel quality. *Journal of Phonetics* 8: 247–258.