Martine Toda & Marzena Zygis (eds), *Turbulent sounds: An interdisciplinary guide*. Berlin/New York: Mouton de Gruyter. 143–166.

Kim, Mi-Ryong & San Duanmu. 2004. Tense and Lax stops in Korean. *Journal of East Asian Linguistics* 13. 59–104.

Kim, Soohee. 1999. *Subphonemic duration difference in English /s/ and few-to-many borrowing from English to Korean.* Ph.D. Dissertation, University of Washington.

Kim-Renaud, Young-Key. 1974. *Korean Consonantal Phonology.* Ph.D. Dissertation, University of Hawaii.

Kingston, John. 1990. Articulatory binding. In John Kingston & Mary E. Beckman (eds), *Papers in Laboratory Phonology I,* 406–434. Cambridge, UK: Cambridge University Press.

Kingston, John & Randy L. Diehl. 1994. Phonetic Knowledge. *Language* 70. 419–454.

Klatt, Dennis. 1974. The duration of [s] in English words. *Journal of Speech and Hearing Research* 17. 51–63.

Leben, W. 1973. *Suprasegmental Phonology.* Ph.D. Dissertation, MIT,.

Lee, Arong & Gregory K. Iverson. 2006. Variation in Korean integration of English word-final /s/. *Euhakyenkwu* [Language Research] 42. 239–251. Seoul : Seoul National University.

Lisker, Leigh & Arthur S. Abramson. 1964. Cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization*, Ph.D. Dissertation, Unviersity of Massachusetts at Amherst.

Lombardi, Linda. 1995. Laryngeal features and privativity. *The Linguistic Review* 12. 35–59.

McCarthy, John J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry* 17. 207–263.

Murray, Robert W. & Theo Vennemann. 1983. Sound change and syllable structure in Germanic phonology. *Language* 59. 514–528.

O'Shaughnessy, Douglas. 1981. A study of French vowel and consonant durations. *Journal of Phonetics* 9. 385–406.

Silva, David. 1992. *The Phonetics and Phonology of Stop Lenition in Korean*. Ph.D. Dissertation, Cornell University, Ithaca, N.Y.

Silva, David. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23. 287–308.

Skalicková, A. 1960. *The Korean Consonants*. Prague, CSAV, Rozpray.

Stevens, Kenneth N. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica* 34. 264–279.

Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Cambridge, MA: the MIT Press.

Stevens, Kenneth N. & Sheila E. Blumstein. 1981. The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas & Joanne L. Miller (eds), *Perspectives in the study of speech*, 1–39. Hillsdale: Lawrence Erlbaum.

Stevens, Kenneth N. & Samuel J. Keyser. 1989. Primary features and their enhancement in consonants. *Language* 65. 81–106.

Stevens, Kenneth N., Samuel J. Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonologoical theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds), *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Erlbaum.

Vaux, Baux. 1998. The laryngeal specifications of fricatives. *Linguistic Inquiry* 29. 497–511.

Vennemann, Theo. 1988. *Preference Laws for Syllable Structure*. Berlin: Mouton de Gruyter.

Yip, Moria. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19. 65–100.

Automaticity vs. feature-enhancement in the control of segmental F0

Phil Hoole & Kiyoshi Honda
Ludwig-Maximilians-Universität München/Laboratoire de
Phonétique et Phonologie

Several phonological oppositions are typically accompanied by F0 differences. F0 is higher following voiceless consonants than voiced; it is also higher for high than low vowels. Analysis of cricothyroid activity aimed to determine whether these F0 differences are automatic effects contingent on the basic articulatory manoeuvres required for the oppositions of voicing and vowel height, or whether the differences reflect active enhancement strategies. Results for both oppositions suggest a hybrid model: The articulatory contingency is at the heart of the F0 differences, but these differences may be reinforced by active laryngeal adjustments. Additional analysis focused on German tense vs. lax vowels. Higher cricothyroid activity in lax vowels could explain why these vowels do not follow typical intrinsic F0 patterns. Tentative support was found.

1. Introduction

It has long been known that some linguistic distinctions involve consistent differences in fundamental frequency, even though the basic distinctions are not prosodic in nature. This applies to consonant voicing, where F0 is typically higher following voiceless than voiced consonants, and also to vowel height, where F0 is typically somewhat higher in high than low vowels. The latter is often referred to as intrinsic F0 (abbreviated in the following to IF0). It is by no means clear, however, what the cognitive status of these F0 differences is. Are they simply contingent on the articulatory manoeuvres required for the basic linguistic contrasts (e.g. control of offset and onset of voicing for consonants, control of tongue height for vowels), and thus automatic and not directly planned by the speaker? Or do they represent strategies that speakers can use to actively enhance these distinctions? An example of an enhancement that is clearly not contingent on the basic lingual articulatory distinction is that of lip-rounding for /ʃ/ vs. /s/ in English (and other languages). As discussed by Keyser & Stevens (2006) as part of their extensive consideration of enhancement mechanisms, this use of lip position ensures the salience of the spectral differences induced by the differing anteriority of the tongue-blade constriction for the two fricatives. Similarly, for the F0 patterns of interest here Diehl and Kingston have argued that these patterns reinforce auditorily relevant characteristics of the oppositions involved, but are not directly contingent on the defining articulatory manoeuvres for voicing in consonants on the one hand, and height in vowels on the other (see Diehl 1991; Kingston 1992; Kingston & Diehl 1994). If correct, such an account would provide support for the primacy of the auditory representation of speech sounds. In fact, we will be arguing that the account most consistent with the experimental evidence may be a hybrid one which allows us to get the best of both worlds. In other words, the basic F0 effects are indeed a mechanical consequence of more fundamental articulatory manoeuvres for voicing and vowel height. But some speakers (some of the time) may latch onto these effects and reinforce them with active muscular adjustments.

In order to disentangle the relative weight of passive mechanical and active muscular contributions to F0 direct recordings of laryngeal muscle activity by means of electromyography (EMG) can have an important role to play. In this paper, then, we present an analysis of cricothyroid (CT) activity. This is the intrinsic laryngeal muscle that, acting on the cricothyroid joint, regulates the relative position of cricoid and thyroid cartilages. Increasing CT activity thus lengthens the vocal folds, and concomitantly, increases the longitudinal tension in them. Accordingly, CT is generally regarded as the laryngeal muscle most unambiguously involved in the control of F0 (see Honda 1994, for a recent review of the physiological mechanisms in F0 control).

Before turning to details of the experimental methods we will give further background to the two main areas of consonant voicing and vowel intrinsic F0. The latter will include as a sub-area reference to a further linguistic opposition, namely the tense-lax distinction in German, since this has proved an interesting challenge for teasing apart automaticity and enhancement in F0 control.

After presentation of the experimental results the concluding discussion will consider the implications for theories of enhancement, for example with respect to the division in the model of Keyser & Stevens into feature-defining and feature-enhancing gestures, and with respect to the observation that they refer to themselves as "unexpected" (op. cit., p. 57) that in strongly overlapped running speech feature-defining gestures are more likely to be obliterated than enhancing ones.

1.1 Consonant voicing

In this section we present the physiological background to the association of voiceless consonants with higher F0 and voiced consonants with lower F0. While the basic acoustic patterns are readily observable, we will outline why there has been room for disagreement as to the source of the effects, and thus as to what they reveal about how speakers plan the realization of one of the most fundamental of all phonological distinctions.

It is not necessary here to review all the literature that has accumulated on the simple fact that there are voicing-related differences in F0. A good source of early discussion is in the book on tone edited by Fromkin (1978), particularly the chapters by Hombert and Ohala. Convenient sources showing typical patterns are Ohde (1984) and the paper by Löfqvist et al. (1989) discussed in more detail with respect to the physiological findings below (see Hoole 2006a, for further references and extensive illustrations based on the German material used in the present investigation). Typically, the F0 at the midpoint of the vowel following a voiceless consonant is slightly but consistently higher (of the order of 10 Hz) than following a voiced consonant. However, the effects are almost always substantially larger at the onset of voicing for the vowel. Thus, a common pattern is for the differences to gradually weaken over the course of the vowel, but they may remain visible even right up to the end of the vowel.

A key earlier paper for understanding why F0 differences might occur is the seminal paper on laryngeal features of Halle & Stevens (1971). Their work made clear that increasing the tension of the vocal folds can support the suppression of voicing – and this increased tension might then also manifest itself as higher F0 in the neighbouring vowel. Conversely, slackening of the vocal folds (leading to a reduction in F0) can be advantageous for supporting voicing in adverse aerodynamic conditions, e.g. when the vocal tract is occluded. The carefully balanced corpus in recent acoustic analyses of Hanson (2009) supports the view that voicing-related differences in English (and we assume a similar situation for our German material) are probably best viewed as a raising of F0 on voiceless consonants, rather than a lowering on voiced consonants.

EMG support for voicing-related regulation of vocal-fold tension in consonants has, however, been very slow to accumulate. The clearest evidence to date that voiceless consonants attract higher CT activity has come from Löfqvist et al. (1989; see also Hutters 1985; Dixit & MacNeilage 1980; – and – on the negative side Collier et al. 1979). Nevertheless, as Whalen et al. have more recently remarked (1993, p. 2158) the number of languages investigated is too small to allow firm conclusions.

It is worth pointing out here that the other main muscle responsible for regulating vocal fold tension, namely the vocalis, is unlikely to provide the link between devoicing and high F0 since on balance the evidence suggests that it is suppressed during voiceless consonants (e.g. Hutters 1985, Collier et al. 1979) – probably because it tends to support vocal fold adduction. This thus provides further motivation for focusing on CT.[1]

The most basic aim of the present investigation is to determine how robust the difference in CT activity is between voiceless and voiced consonants. Only if the pattern expected from previous work proves to be a consistent one (increased CT muscular activity in the voiceless case) will it make sense to go on to the more far-reaching aim of determining whether this increased activity is simply part of the physiological control of voicelessness that happens to spill over onto the following vowel in the form of increased F0, or whether this increased F0 on the vowel represents an independent auditory property that speakers actively manipulate to enhance (and extend in time) the acoustic salience of the voicing distinction. The interpretation can thus hinge quite finely on the precise timing of the CT activity. The timing in turn is a delicate matter because estimates for the delay between CT activity and its effects on the acoustic output are not very consistent, ranging from about 20 to at least 80ms (we return to this point in the methods section).

Löfqvist et al. are firmly of the opinion that increased CT activity supports devoicing, i.e. it is not planned by speakers primarily to increase F0 in the following vowel, because peak CT activity is located squarely in the consonantal phase, and voiced-voiceless CT differences are weakest around the midpoint of the vowel. Nevertheless, it remains slightly unclear why F0 differences on the vowel are very consistently found.

Moreover, as pointed out by Kingston & Diehl (1994, p. 440) there is perhaps also a tension between the assumption that the CT activity helps to suppress voicing – which would mean concentrating the effect at the start of the consonant – and the assumption that it can also account for clear F0 differences after the end of the consonant. In fact, Kingston & Diehl are of the opinion, based on a consideration of the voicing contrast in a large number of languages, that F0 differences are not contingent on the specific laryngeal adjustments for the consonant at all, but depend on the consonant's specification with respect to a more abstractly defined [voice] feature. In other words, the F0 differences can be seen as an auditorily driven enhancement of the voicing contrast – enhancing the contrast by increasing the low frequency emphasis in the spectrum in the [+voice] context (any voiced sound will generally have its main concentration of energy in the spectrum lower than for a voiceless sound; this basic pattern can be emphasized by keeping F0 low after a voiced consonant):

---

[1] But it must also be admitted that in the laboratory setup available to us CT was much easier to obtain than vocalis (and also less painful for the subject).

> F0 is uniformly depressed next to [+voice] stops, regardless of how the [voice] contrast is otherwise realized … the F0 differences are a product of articulations that are controlled independently of the timing and size of glottal articulation (Kingston & Diehl 1994, p. 432)

We believe that careful consideration of the timing patterns in our data will in fact reinforce the interpretation that F0 effects are contingent on the articulatory adjustments for consonantal devoicing. At the same time, in order to resolve some of the unclear issues above, we will try to identify mechanisms that can help understand how the effect of the increased CT activity propagates to the following vowel, and we will also reconsider whether the consonantal CT activity is a specific adjustment to suppress voicing.

Finally, we will entertain the possibility that even though the CT activity is rooted in the contingencies of consonantal articulation, speakers may develop an enhancement strategy by reinforcing this effect – a hybrid approach that in a sense combines the accounts of Löfqvist et al. on the one hand, and Kingston & Diehl on the other.

Since the empirical part of this paper will be focusing on the underlying physiological and biomechanical mechanisms it is worth pointing out before continuing that a logical prerequisite for the assumption of an enhancement strategy is that the F0 differences are perceptually relevant. While investigation of the cue-value of consonant-related F0 differences in the perception of the voicing distinction has given somewhat mixed results (see e.g. Haggard et al. (1970), Abramson & Lister (1985), Kohler (1985), Schiefer (1986), Silverman (1986), Whalen et al. (1990), Whalen et al. (1993)) our understanding of Whalen et al. (1993) is that F0 differences of typical magnitude do indeed reach listeners' perceptual systems, and are thus potentially available for enhancement. The latter authors explicitly discuss their results with respect to tonogenesis in which one common process is assumed to be that the consonant-induced F0 perturbations on the vowel remain as a tonal contrast on the vowel after the original voicing distinction has been lost (see e.g. Hombert et al. 1979). But if – as we will be arguing here – the F0 differences are rooted in the articulatory manoeuvres required for the voice-voiceless contrast then – as Whalen et al. further point out – tonogenesis logically requires a process of enhancement to have taken place because otherwise the F0 differences would disappear in parallel with the loss of the voicing distinction.

1.2 Vowel intrinsic F0

The purpose of the present section is analogous to that of the previous section, but considers now the status of F0 differences related to vowel height. The basic point to be made is that previous investigations allow room for disagreement as to whether the F0 differences represent an active strategy by speakers to enhance the basic phonological opposition of vowel height, or whether the differences emerge as a mechanical consequence of movements required in any case for the articulation of the different vowel categories. In addition, an important methodological insight of previous work will be introduced that will help to maximize the usefulness of new EMG data in resolving this disagreement.

As in the previous section, it is not necessary here to review all of the literature on the basic F0 findings. Based on the extensive range of languages and investigations summarized in Whalen & Levitt (1995) high vowels have, on average, an F0 that is of the order of 15 Hz higher than in low vowels. Specifically for German, in previous work we have found differences of the order of 30 Hz for speech material similar to that used in the present investigation (Hoole & Mooshammer 2002).

The comprehensive review of Whalen & Levitt is not just an exercise in data collection, however. The theoretical thrust of their paper is to use the universality of IF0 as an indication that it is an automatic, mechanical consequence of vowel articulation, and not an auditorily driven strategy used by speakers to enhance vowel contrasts. The latter view has been put forward, for example, by Kingston (1992). Enhancement could work by exploiting the known relevance of the F1-F0 distance in the perception of vowel height (e.g. Traunmüller 1981), i.e. high F0 and low F1 for high vowels, and vice-versa for low vowels, but as pointed out by Whalen et al. (1998) note that there are presumably limits on this effect in running speech because vowel height is not misperceived when a syllable receives F0 prominence.

Why, more specifically, might universality speak against enhancement? Briefly, (see Whalen & Levitt for more discussion), IF0 is present in languages where it is probably not useful for enhancement (e.g. small vowel inventory), or, at first sight, even counter productive (e.g. tone languages; but see also Connell 2002). Also, IF0 tends to disappear for utterances at the lower end of the speaker's F0 range. If the effect were an actively controlled one there would be no particular reason to expect this, whereas the mechanical status of the larynx is undoubtedly different at different F0 levels.

Clearly, EMG evidence is once again crucial, since an active enhancement mechanism must predict laryngeal EMG differences related to vowel height. Before summarizing results of previous investigations, it is necessary, in turn, to pin down the potential sources of mechanical effects.

Various versions of what might be termed tongue-pull theories have been proposed (see e.g. Dyhr (1990), Sapir (1989), Honda (2004), Hoole (2006a) for reviews). Perhaps the most generally accepted account is that proposed in Honda (1983), by which contraction of the genioglossus posterior to support raising of the tongue exerts a forward pull on the hyoid bone which in turn causes the thyroid cartilage to rotate forward and lengthen the vocal folds.

Regarding possible differences in CT activity between high and low vowels we turn first to Honda & Fujimura (1991). Interestingly, they report results for CT (one subject) showing a close relationship with IF0 (about as close as that between genioglossus posterior and IF0). This close match is referred to as "unexpected" but they go on to give an interpretation that will be very relevant for the hybrid approach argued for here: They entertain the possibility that mechanical and active effects may co-exist. "Thus the cricothyroid activity associated with high vowels "emulates" the biologically natural F0 rise due to hyoid bone movements" (p. 151). This biological predisposition could in effect lead to a process of phonologization.

One of the more extensive investigations to date of the relationship between CT and IF0 is that of Dyhr (1990; 4 speakers). Although Whalen et al. (1998) are rather sceptical about the interpretability of the results we think, in fact, that they are presented in sufficient detail for it to be clear that there is a tendency for higher CT activity for high vowels.

Thus one might be tempted to view IF0 as an active adjustment. However, such a conclusion would be premature, or at least incomplete. Whalen et al. (1998) failed to find evidence for higher CT activity in higher vowels. Perhaps more crucially, however (since the results are based on very unusual utterances in which steady-state vowels are matched to a series of pre-given pitches), they argue very cogently that when interpreting positive results it is necessary to look closely at the regression line linking F0 and muscular activity. If results for different vowels lie on different regression lines then it is hazardous to simply compare activation levels across vowels. For example, interpretation becomes a delicate matter if a given change in EMG has a different effect on the F0 of [i] compared to that of [a]. Put another way, if two vowels lie on different regression lines, for example such that at the same CT value a different F0 is predicted, then clearly there remain factors in addition to CT that affect F0. These methodological considerations will prove essential in our attempts below to disentangle mechanical and enhancement contributions to F0 in our own data and thus demonstrate the usefulness of a hybid account: put in a nutshell, a difference in CT activity between low and high vowels does not rule out the possibility that mechanical effects are still relevant.

1.3 German as a test case: The tense-lax opposition

There is a further very specific issue related to vowel IF0 for which EMG is also crucially required, and, more particularly, for German speakers. The German vowel system divides fairly neatly into two series, with members of the tense series being more peripheral and longer than corresponding members of the lax series. With the exception of the /a/ vowel pair, the lax member of a vowel pair is generally lower than the tense cognate. The interest of the situation in German resides in the following: It has become fairly clear that German tense and lax vowels do not differ much in IF0 whereas they do differ clearly in tongue height. (We will use 'tenseness' as a convenient label for this distinction, although as will be seen as the discussion progresses it may not be a felicitous one.) For example, the tongue height of lax /ɪ/ is not only lower than that of its tense counterpart /iː/ but is in fact often even lower than that of tense /eː/. Depending on the investigation, the IF0 of lax /ɪ/ is sometimes a bit higher and sometimes a bit lower than that of the tense cognate /iː/, but crucially it is never as low as (or lower than) tense /eː/. Accordingly, this situation has been quoted as a problem for mechanical accounts of intrinsic F0 (Diehl 1991; Kingston 1992, 2007). Fischer-Jørgensen (1990) gives an extensive discussion of possible reasons (see also Hoole 2006a) and also provides a summary of IF0 measurements of German. Hoole & Mooshammer (2002) present both IF0 data and EMA measurements of tongue position. Earlier sources of tongue-position data are given in Wood (1982).

There is in fact an interesting scenario that could still salvage the mechanical hypothesis: Assume that mechanical effects of the kind outlined above in Section 1.2 are actually at work. This would normally result in a lower F0 for lax /ɪ/ than tense/iː/ (in fact, as just mentioned, it could even be expected to be lower than tense /eː/). However, speakers may actively raise F0 on the (so-called!) lax vowels, giving a very similar final outcome at the acoustic surface. If EMG data provided evidence for this then one long-standing problem regarding IF0 would be resolved. In addition, it would have interesting implications for our understanding of the articulatory components of the tense-lax distinction. For example, active raising of F0 on lax /ɪ/ might help to distinguish it from tense /eː/, to which it is quite similar in terms of tongue height. Or the resulting rather similar F0 for /ɪ/ and/iː/ might function as a prosodic marking that the two vowels are the highest representatives of the tense and lax series respectively.

1.4 Summary of the issues

Summarizing the issues raised in the introductory sections above, it has been argued that the question of automaticity versus enhancement with respect to F0 control is relevant – and as yet unresolved – in two main areas, namely consonant voicing and vowel height, the latter area being subdivided with respect to straight height distinctions on the one hand, and tense-lax distinctions on the other. This motivates the acquisition and analysis of muscle activation patterns as a particularly direct route to resolution of these issues, and thus to improving our understanding of the articulatory adjustments involved in key phonological oppositions.

2. Experimental procedures, speech material and subjects

Recordings of CT muscle activity were made for three native speakers of German. The speech material consisted of pseudo-words consisting of tense or lax vowels in a voiced or voiceless consonantal context. Vowel height was also systematically varied. These target words were uttered in a short carrier phrase. There were slight differences between the three subjects regarding the precise corpus and some other experimental details. This will be outlined for each subject below.

EMG was recorded by means of hooked-wire electrodes (platinum-iridium alloy, 0.002″ diameter, A-M Systems). The insertions for all three subjects were into a fairly posterior location of the CT muscle, estimated to be either in the pars obliqua or at the transition between pars recta and pars obliqua (see e.g. Hirano & Ohala 1969, for illustrations of insertion paths into CT). Signals were recorded with an 8 Channel Grass Model 15LT system with Model 15A54 amplifiers. Filter settings consisted of a high-pass filter at 30 Hz, a notch-filter at 50 Hz to suppress mains hum, and a low-pass filter at 3000 Hz. Additional filtering was carried out in software as outlined in Section 2.2 below.

2.1 Subject-specific details

Subject CK

The vowels consisted of four tense-lax vowel pairs, namely/iː yː uː ɑː/ (tense), and / ɪ ʏ ʊ a/(lax). The vowels were embedded in the stressed syllable of the two pseudo-words /gəlVbə/ (voiced context) vs. /gəfVpə/ (voiceless context).[2]

The target words were in turn embedded as ⟨word1⟩ in the carrier phrase "*Ich habe ⟨word1⟩ nicht ⟨word2⟩ gesagt*" (= "I said ⟨word1⟩ not ⟨word2⟩"; ⟨word2⟩ also contained phonetically varying material that will not be analyzed here). 10 randomized repetitions of every sentence were recorded.

Subject CG

For this subject the corpus was expanded relative to that used for subject CK, who had been the first subject recorded. Preliminary results for CK had indicated that differences related to the tense-lax distinction might be present, so for subsequent subjects an additional tense-lax pair was included, in order to give more comprehensive data on this point. These results had also indicated that differences related

---

[2] Whenever it is useful to refer to tense-lax pairs together, we will use the term 'vowel category' (e.g. in the statistical analysis in the results below) and as notation will use the basic Roman letter without length mark, e.g. /i/ refers to /iː, ɪ/ (a-vowels will be disambiguated verbally).

to consonant-voicing could be present, so an additional phonetically maximally simple voiceless vs. voiced context was now included (and the carrier phrase was simplified to make it easier to accommodate this additional material).

Accordingly, for subject CG the corpus was divided into two parts (and will be referred to as Corpus 1 and Corpus 2 in the results section). In both parts the vowels were /iː yː eː uː ɑː/ (tense), and /ɪ ʏ ɛ ʊ a/ (lax), i.e. the tense-lax pair /eː, ɛ/ was added, this also introducing an intermediate vowel height level compared to the corpus used for the first subject CK. In the first part the target pseudo-words were /lVbə/ and /fVpə/, i.e the same consonantal contexts as for CK. In the second part, symmetric plosive contexts were used: /bVbə/ (voiced) and /pVpə/ (voiceless). The carrier phrase was simpler than that used for CK: "*habe* ____ *besucht*" (= "I visited ____"). For both corpora 10 repetitions of each utterance were recorded.

Subject SF

For this subject the intention was to record the same two-part corpus as for CG. However, some problems were encountered with movement artefacts (including microphonic effects at the voice frequency) in the EMG signal. This became worse in the second part of the corpus and the recording had to be terminated after about 6 repetitions had been collected. For this reason, detailed statistical results will be reported here only for the first corpus (containing 10 repetitions). Despite these problems the signal was considered to be of acceptable quality after additional high-pass filtering at 500 Hz (see Rischel & Hutters (1980) for confirmation that high-pass filtering at quite high cutoff frequencies can be beneficial for laryngeal EMG in speech experiments). Some more qualitative observations using material from the second corpus will also be included as part of the discussion of consonant voicing effects (Section 4.2).

2.2 Processing of the EMG data: Estimating strength of muscle activation

In order to be amenable for analysis the high-frequency EMG interference pattern (bandwidth of the order of a couple of kHz) must be converted to a low bandwidth signal reflecting the bandwidth of speech movements (of the order of 20 Hz). Analogously to the calculation of an amplitude envelope from a raw audio signal the most basic and probably most popular way of processing the raw EMG signal is to calculate the RMS amplitude over a sliding window. One variant of this suggested by the discussion in Rischel & Hutters (1980) already noted above is that high-frequency emphasis of the EMG signal may provide better differentiation of speech-related patterns. A simple way of achieving this is to simply take the first difference of the raw waveform. Accordingly, we calculated the RMS amplitude both for the raw and the first-differenced waveform. The details of the

calculations were: Window length 40ms; window shape, square; window shifted in steps of 2.5ms (giving an output sample rate of 400 Hz); additional smoothing of the 400 Hz signal with a Kaiser FIR filter, with Fc = 15 Hz.

As an additional means of capturing the patterns of activity in the signal, we also calculated the zero-crossing rate using the same window and filtering parameters just given for RMS, and again performed the calculations both on the raw and first-differenced waveform. The motivation for use of zero-crossing rate with EMG signals has been that it may give a measure of motor-unit activity in the muscle that is reasonably independent of position of the electrodes relative to the motor-units being picked up. Amplitude-based methods may be dominated by high-amplitude spikes from one (or just a few) motor-units very close to the recording site (Fromkin & Ladefoged 1966; see also Clancy et al. 2002, Zhou & Rymer 2004, for recent discussion of EMG processing parameters. Computing the zero-crossing rate from the differentiated raw signal is equivalent to peak-counting, which has also frequently been used as an EMG parameterization).

Preliminary inspection of the recordings indicated that use of the first-differenced signal for the RMS amplitude and the zero-crossing rate did indeed give stronger correlations with F0, and greater sensitivity to the experimental variables, so only the EMG parameterizations based on the first-differenced signal will be considered further here. In addition, it also emerged that the relative sensitivity of RMS amplitude and zero-crossing rate varied quite noticeably over speakers and electrode insertions. Accordingly, for the detailed analyses below we decided to use for each speaker whichever measure gave the strongest correlation with F0 in the target vowel. For CK and SF this resulted in a clear choice in favour of zero-crossing rate. For CG the differences were less marked, but came out in favour of RMS amplitude.[3]

---

[3] Further background and illustrations pertaining to the different approaches to processing the raw EMG data can be found in Hoole (2006a). In fact, all the ANOVAs reported in the main results section below were run with all variants of the processed EMG, i.e. RMS and zero-crossing for both non-differentiated and differentiated raw signal. There were only isolated cases where the choice of processing strategy made a marked difference to the results of the key experimental comparisons. One example of this was the effect of consonant voicing on the strength of the EMG signal in the C1 segment of speaker CK. This was highly significant for the zero-crossing parameterization reported below (see table of ANOVA results in appendix), but not significant for the RMS approach. It should be emphasized that the choice of parameterization for each subject was made independent of the patterns in the ANOVA results but rather, as noted above, simply referred to the strength of the correlation between muscular activity and F0, which appears justified in view of the overall aim here of studying the muscular control of F0.

2.3 Time alignment of EMG activity with F0

In order to be able to interpret the EMG patterns in terms of the acoustic consequences intended by the speaker in the speech signal, it is necessary to have some estimate of the delay between measurable electrical activity in the muscle and the point in time at which the speech signal is actually affected by the associated changes in articulatory configuration. In the present case this is a somewhat delicate matter, since some of the speech segments in which we are interested are quite short (e.g. high lax vowels) and the estimates for the delay to be found in the literature vary by amounts comparable to the duration of such segments (from about 20ms to over 100ms).

It would be beyond the scope of the current paper to try and disentangle the possible sources of the diverging estimates (e.g. Atkinson 1978, Baer 1981, Sawashima et al. 1982; Larson et al. 1987, Sapir et al. 1984, Dyhr 1990, Herman et al. 1999). Since it appeared essential to derive our own empirical estimate that would be appropriate to the speech tasks of interest here, we adopted the following procedure:

Preliminary inspection of EMG and F0 contours suggested that EMG led F0 by a time in the range of 30 to 120ms. After processing the EMG as described in Section 2.2 above we shifted the EMG signal in steps of 10ms over this range, and at each time step determined over each target vowel the average EMG and median F0 for that token, and then calculated the correlation coefficient between these two parameters over the complete vowel sample. Depending on speaker and insertion, correlations peaked at time steps from 50 to 70ms, so for further analysis we used a time-shift of 60ms for all speakers.

3. Results

In many earlier EMG studies of speech, presentation of the results was often confined to qualitative discussion of ensemble averages. We were encouraged by the approach in the investigations of Löfqvist et al. (1989) and Whalen et al. (1998) to envisage carrying out classical statistical analysis such as ANOVA in which each token constitutes an observation. To extract measures amenable to such analysis we determined separately for each token the average signal amplitude for the two acoustically-defined segments of principal interest here, namely the consonant preceding the target vowel (C1) and the target vowel itself (V).

The structure of the material for this experiment involves three independent variables, namely voicing of the consonant context, vowel category (4 vowel

categories for CK, five for CG and SF), and vowel tenseness.[4] These three factors will be referred to as VOICE, VOWEL, and TENSE, respectively, and the results will be presented in that order in three main sections corresponding to these factors. Analyses of Variance were carried out separately for each speaker (using the GLM function of SPSS). The two parts of the corpus for speaker CG were also analyzed separately, and will be referred to as speakers CG(1) and CG(2). A complete tabulation of the ANOVA results is given as an appendix.

3.1 Consonant voicing

The ultimate aim of the systematic manipulation of consonant voicing in the design of the experiment is to determine whether any differences in F0 on the vowel following the consonant can be directly attributed to articulatory manoeuvres necessary for the regulation of consonant voicing, or whether such F0 differences are better attributed to supplementary adjustments to the vowel itself. Resolving this issue actually needs to proceed in several steps. The first task is to determine whether there is a reliable difference in CT activity related to consonant voicing at all (since previous work is not comprehensive enough to be sure in advance on this point). The present section will give the results for this basic point (together with information on the F0 differences actually found in the present material). Higher CT activity (in the context of voiceless consonants) located not just on the consonant itself but also on the following vowel would be the most direct evidence in favour of an enhancement effect, since there is no obvious reason why control of voicing for the consonant mechanically requires different CT activity on the following vowel. The more likely result, in the light of previous work, is that CT activation differences will essentially be confined to the consonant. In this case, a final decision on the 'articulatory contingency' versus 'enhancement' explanations will be postponed to the discussion section, since it will be necessary to consider in more detail the precise purpose of the CT activity as well as the physiological background that would account for how its effect propagates to the following vowel (and we will also make use of an analytical tool that is more conveniently introduced in the section on vowel IF0).

Turning now to the results, the segment of initial interest here is the consonant (the C1 segment). We will consider this first and then go on to consider whether voicing-related effects remain visible in the vowel for the EMG signal, and how strong voicing-related F0 differences are.

---

[4] In other words, as outlined in Section 2.1 (details of material for each subject), 'vowel category' refers to tense-lax vowel pairs.

The result for the main effect of VOICE in C1 was very simple: It was significant at $p < 0.001$ for all speakers, and always in the expected direction of more activity for voiceless than voiced. Thus this study adds a substantial body of evidence for a language not previously investigated that the cricothyroid is involved in the articulatory manoeuvres distinguishing voiceless and voiced consonants. It is worth pointing out that a similar pattern was found for CG(1) and CG(2), so the results do not appear to change radically depending on the specific voiced and voiceless consonants used for the comparison: /l/ vs. /f/ for Part 1 and /b/ vs. /p/ for Part 2. Thus, it would be more precise to say that voiceless vs. voiced in this context refers to consonants that have a clear laryngeal abduction-adduction gesture (/f, p/) vs. those that do not (/l, b/).

The ANOVA did, however, also reveal some significant interactions: with TENSE at $p < 0.01$ for CK and $p < 0.05$ for CG(1), and with VOWEL at $p < 0.01$ for SF and CG(1). (The 3-way interaction was not significant; in fact, it was not significant in any of the ANOVAs carried out in this study, so it will not be referred to explicitly again.)

The VOICExTENSE interaction expressed itself as a larger difference between voiceless and voiced in the context of tense vowels (see Figure 1), an effect that we will encounter from several points of view in the course of these results.

The VOICExVOWEL interaction for SF and CG(1) was mainly due to rather smaller voiceless-voiced differences for the /a/ vowels. However, nowhere in any of the material did any of the interactions lead to a reversal of the direction of the voiceless vs. voiced comparison.

Turning to the vowel segment, the interesting question here is whether any voiceless vs. voiced differences are so robust that they could be seen as evidence that speakers actively plan to generate voicing-related differences in F0 on the vowel.

In the ANOVA a significant main effect of VOICE was found for CK at $p < 0.01$, SF at $p < 0.001$ and CG(2) at $p < 0.01$. The result for CG(1) was not significant. The overall pattern is thus that significant voiced-voiceless differences still occur, but tend to be less strong than on the consonant itself. This is illustrated in Figure 1, which shows for the consonant and vowel segments the average *differences* between matched voiced and voiceless target items (i.e. matched for vowel category and tenseness). The differences were calculated as 'voiceless' minus 'voiced', so values above the dotted zero line are consistent with the hypothesis of more activity for voiceless consonants. For CG(1) and CG(2) there was a general decrease in the size of the differences from consonant to vowel. For CK and SF this only applied to the matched pairs involving lax vowels. Accordingly, the VOICExTENSE interaction was significant at $p < 0.01$ for CK and at $p < 0.05$ for SF (the only other significant interaction was for VOICExVOWEL that was significant at $p < 0.05$ for SF).

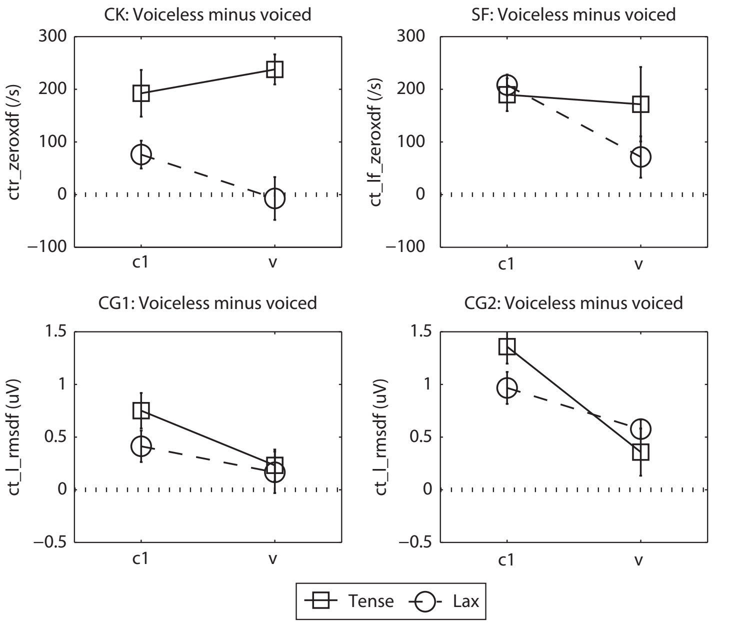

Figure 1. Difference in CT activity between corresponding segments (C1 and vowel on the abscissa) in voiced and voiceless contexts. Positive values (above the dotted line) indicate more activity in voiceless context. Calculated separately for tense and lax vowel contexts. First, average EMG activity for each word form was calculated. Then, each data point was calculated as the average difference between word pairs contrasting only with respect to voicing category. Error bars are the standard error of the mean over the corresponding number of word pairs: 4 for CK, 5 for all other speakers. See text for further details. (The EMG units reflect the different procedures used for estimating activity level: zero-crossing rate (i.e. per second) for CK and SF, RMS amplitude for CG ('uV' = 'microVolt'); in all cases based on the differentiated signal.)

We have now seen that there are clearly higher levels of CT activity associated with voiceless consonants and extending to some extent from the consonant into the following vowel. It is now necessary to determine whether the expected differences in F0 are actually found. For the purposes of statistical analysis we extracted the median F0 in each vowel token and carried out an ANOVA with the same factors as for the EMG analysis.[5]

---

[5] F0 was calculated by means of the YIN algorithm of de Cheveigné & Kawahara (2002). It uses a modified autocorrelation approach that coped well with the less than ideal audio recording conditions in the EMG experiments. It can be downloaded from A. de Cheveigné's webiste: http://audition.ens.fr/adc/sw/yin.zip (accessed on 10.12.2009).

Since the F0 contours in the recorded utterances were usually monotonically rising (or fairly flat) the median value can be considered as representative for the situation at about the midpoint of the vowel. This can be considered in turn as providing a fairly conservative estimate of the robustness of voicing-dependent differences, since inspection of ensemble averages of the F0 contours indicated that differences were typically greater at vowel onset and tended to decrease over the course of the vowel (cf. similar pattern in Löfqvist et al. 1989). The basic result was once again very simple: The main effect of VOICE was in every case significant at $p < 0.001$ in the expected direction of higher F0 values in the voiceless context. Figure 2 shows the average voiceless vs. voiced differences for each speaker.[6]

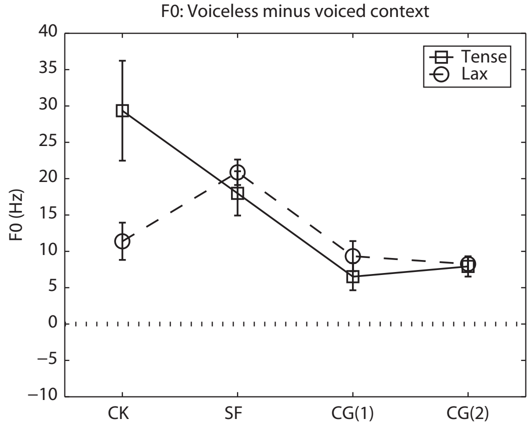

Figure 2. Difference in F0 between corresponding segments in voiced and voiceless contexts. Positive values (above the dotted line) indicate higher F0 in voiceless context. Other details as for Figure 1

The only interactions were VOICE*TENSE for CK at $p < 0.01$ and VOICExVOWEL for SF at $p < 0.05$, which both parallel the EMG results.

More significant for the discussion is that there is one clear case where the EMG result does not parallel the F0 result, namely CG(1) where the EMG result was not significant. This means that it is apparently possible for F0 on the vowel to differ depending on the voicing context, without CT activity to match it. In a sense, then, the F0 in the vowel in voiceless consonant contexts is higher than would be predicted on the basis of the EMG activity in the vowel. In the next section (on intrinsic F0 in vowels) we will present a technique for capturing this observation more quantitatively, and in the concluding discussion it will be necessary to consider what mechanisms could be responsible for the propagation of effects related to activity in the preceding consonant onto the vowel. The other issue that will be discussed further later is whether those cases where voicing-related differences in EMG on the vowel itself do occur (e.g. tense vowels for speakers CK and SF) can be seen as enhancement of an existing effect.

3.2 Vowel intrinsic F0

As discussed in the introduction, the key question revolves around whether the very robust intrinsic F0 difference between low and high vowels involves an active component in terms of laryngeal muscle adjustments (specifically here whether there is a higher level of CT activity for high vowels), or whether the automatic biomechanical consequences of vowel articulation provide sufficient explanation.

We will first consider briefly for the vowel segment itself the statistical results for the main effect of VOWEL in the EMG measurements. We will then look in more detail at the relationship between EMG and F0: The precise nature of the F0 vs. EMG regression for the different vowel categories will play an important role in putting the results into proper perspective. The regression analysis is the most direct route for assessing the automatic account, since it allows the prediction of this account to be tested that high vowels should show higher F0 than low vowels at a comparable EMG level. More importantly, a potential result of higher CT activity for high vowels does not logically exclude the possibility that automatic, mechanical effects are *also* present. This situation, too, is most conveniently uncovered by appropriate regression analysis.

Turning to the detailed results for each subject, for CK the high vowels as a group show quite consistently higher CT activity than /a/, and the main effect of VOWEL was significant at $p < 0.001$. For SF the basic pattern is similar, but is more pronounced for the tense vowels; accordingly, the main effect of VOWEL was again significant at $p < 0.001$, but the TENSE*VOWEL interaction was also significant at $p < 0.001$ (there was also a VOICE*VOWEL interaction: $p < 0.05$). The vowel /e/ (not in the corpus for CK) had a rather variable pattern but appeared on the whole closer to /a/ than to the high vowels /i y u/ (see Figure 3 for details). The pattern for CG was different. In CG(1) no significant effect of VOWEL was found ($p = 0.1823$), while in CG(2) /a/ appeared to have *higher* activity than the other vowels (and the main effect was significant at $p < 0.001$). This effect was stronger for the tense vowels, giving a significant TENSE*VOWEL interaction ($p < 0.01$).

---

[6] It is observable here, and again below in Section 3.2 that CK is the speaker showing the strongest modulation of F0. This may be because he spoke the target words in a carrier phrase involving contrastive stress (see Section 2.1 above).

Average EMG values for each linguistic condition can be found in Figure 3. The main purpose of Figure 3, however, is to give a first appreciation of the relationship between F0 and EMG. The F0 values, considered in isolation, are unremarkable, showing the expected intrinsic F0 pattern. For speakers SF and CG, who had /e/ in the corpus, this vowel is generally intermediate between /a/ and the high vowels /i y u/.

Considering more specifically the relationship between F0 and EMG the patterns that emerge follow on from the ANOVA results, the main difference being between CK and SF on the one hand, and CG on the other. For CK and SF there is a clear positive relationship between CT activity and F0. This is particularly strong for CK, while for SF the emergence of this relationship is particularly due to the tense vowels (plotted in squares and circles) having a clear association between low EMG and low F0, and vice-versa. The lax vowels (diamonds and triangles) cluster in the mid region of the plot and show this association somewhat less clearly. Nevertheless, we can certainly say for both these speakers that there is a strong relationship between EMG and F0 simply because it is apparent even without subdividing the data in detail by consonant context and vowel tenseness. Since, furthermore, high vowels have stronger CT activity activity than low vowels (i.e. in the direction of intrinsic F0) it is tempting to conclude for these speakers that intrinsic F0 is in fact the result of an active laryngeal adjustment. We will see immediately below that this conclusion is premature. In fact, the results for CG already weaken the generality of this explanation. For the first corpus there is simply no obvious relationship between CT activity and F0, while for the second corpus the coupling of the unexpectedly high CT activity for /a/ (already mentioned in the ANOVA results) with completely typical intrinsic F0 values result in a tendency to a *negative* relationship between CT activity and F0.

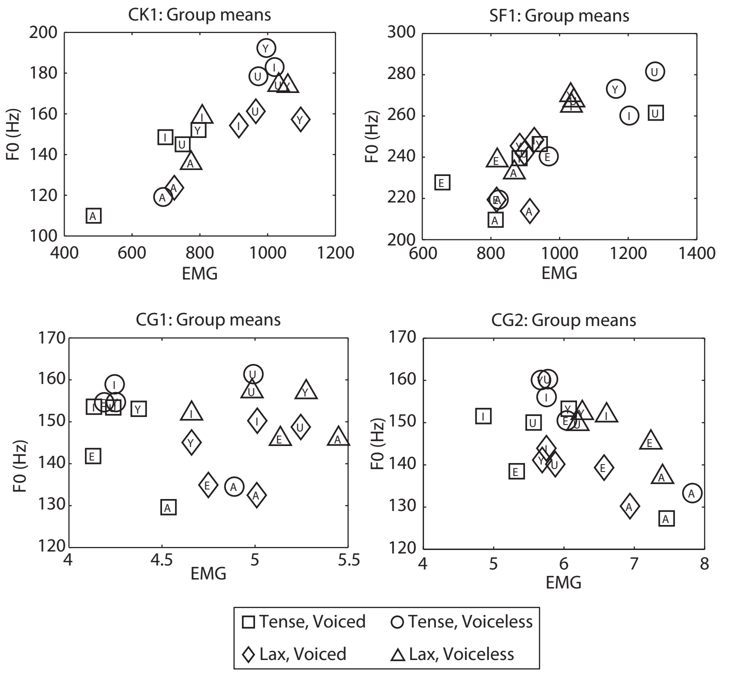

Figure 3. Scatter plot of F0 vs. EMG activity based on mean value of each linguistic category in the corpus (vowel category * tenseness * consonant voicing). See Figure 1 for explanation of EMG units. See Hoole 2006a, Figure 5.23 for a colour version of this figure available online

Figure 3 by no means captures all relevant aspects of the relationship between CT activity and F0. The crucial methodological insight derived from the paper by Whalen et al. (1998) is that it is essential to determine whether data from different vowel categories lie on the same regression line for the prediction of F0 from CT activity. If this is not the case, then F0 must be being influenced by factors in addition to CT activity.

To make the procedure clear we will look at a detailed example from one speaker, before summarizing the complete material for all speakers.

Figure 4 shows the trajectories in the F0 vs. EMG plane of ensemble-averaged data for the tense vowels in voiceless context of speaker SF (i.e. each individual trajectory represents the average over 10 repetitions). This figure shows features already seen in the basic analysis above, i.e. higher F0 for the higher vowels, and also higher CT activity (contours for the higher vowels are displaced slightly to the right relative to /a/). The important point is, however, that by no stretch of the imagination can the traces for the individual vowels be regarded as falling on the same regression line. But this would have to be the case if the CT activity differences alone account for the F0 differences. In other words, there is probably some additional factor contributing independently to the higher F0 in high vowels. The obvious candidate for this is the tongue-pull effect outlined in the introduction. The data shown here for speaker SF has been chosen for illustration as it is a clear example, but it is by no means untypical.

To summarize the robustness of the effects over all material and speakers the following procedure was adopted: Each ensemble-averaged trajectory in the F0-EMG space (i.e. for every possible combination of vowel category with tenseness and voicing context) was fit by means of linear regression. Then the average EMG level was calculated for each group of vowels matched for tenseness and consonant context (e.g. over the five cases shown in Figure 4). At this EMG level the F0 predicted by the regression equations was determined for each vowel. These

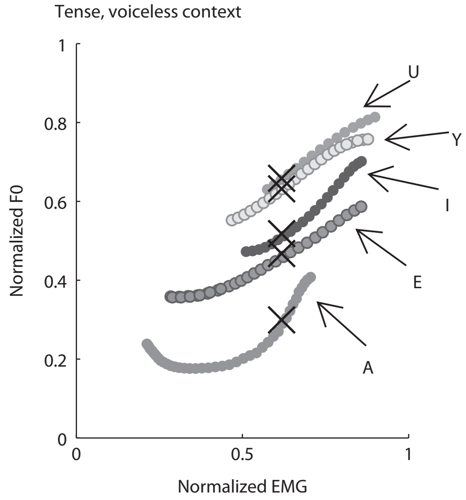

Figure 4. Relationship between ensemble-averaged F0 and EMG over the vowel segment. Speaker SF. Both F0 and EMG have been normalized to a range of 0 to 1. The cross on each trajectory indicates F0 at the overall mean EMG value and thus allows an estimate of the F0 expected for each vowel at a constant value of CT activity (see also Figure 5 below)

are the data points marked by an 'X' on the trajectories in the figure. In effect, this gives us a direct test of the mechanical account of F0: The account predicts that at the same EMG value different values of F0 will result, depending on vowel height.

The results for all material for all subjects are shown in Figure 5.

The results are exactly as would be predicted by a mechanical account of intrinsic F0: In every experimental condition the corresponding /a/-vowel has the lowest F0 value. For the two subjects (SF and CG) for whom /e/ was recorded this mid vowel, in turn, is roughly intermediate between /a/ and the high vowels /i, y, u/. Moreover, the magnitude of the difference between low and high vowels is generally somewhat smaller for the lax vowels (shown in the right half of each panel) compared to the tense vowels (in the left half). This is to be expected because the lax vowels as a group can be regarded as centralized relative to the tense vowels; thus they show a smaller range of differences in tongue height (and tongue root advancement) than the tense group. And finally, and crucially, a mechanical account would predict very similar patterns across speakers, which is precisely what is seen in Figure 5, whereas in contrast as seen in Figure 3 the active muscular activation patterns vary substantially across speakers.

We thus assume that the tongue-pull mechanism is the basic mechanism underlying intrinsic F0 and follow Whalen et al. in rejecting an active enhancement

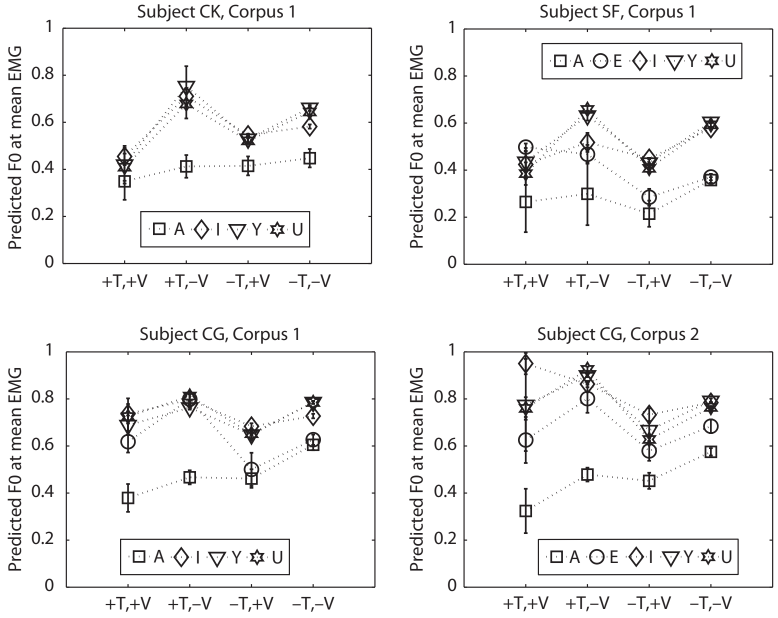

Figure 5. Predicted F0 at mean EMG activity level comparing vowel categories (coded by symbol). Based on regression analysis of ensemble-averaged data. Error bars show 99% confidence interval. Positions on the abscissa represent the possible combination of vowel tenseness (+/–T) with consonant voicing (+/–V). F0 values are normalized, as in the ensemble-average trajectories in Figure 4 above. See Hoole, 2006a, Figure 5.25 for a colour version of this figure available online

mechanism as the core source of the effect. However, the fact remains that two speakers did also have higher CT activity activity on the high vowels. Accordingly, unlike Whalen et al. we consider that a hybrid explanation could be attractive, in effect as suggested by Honda & Fujimura 1991 (see introduction): speakers may learn to actively support a biomechanically given effect, perhaps as no more than part of individual speaking style, but perhaps also as an element of clear speech that is at their disposal. In this view, enhancement effects could thus be expected to vary over speakers, whereas mechanical effects should be consistent, exactly as found here.

In fact, using the information in Figure 5 we can make a rough estimate of the proportion of the F0 differences between vowels visible in Figure 3 that is attributable to mechanical effects.

First we need to extract from Figure 5 the size of the normalized F0 difference between the low vowel /a/ on the one hand, and the group of high vowels (/i u y/) on the other. Confining attention just to the tense vowels (i.e. in the left half of each panel of Figure 5), since these are the vowels where the largest mechanically related differences are expected, we obtain difference values of about 0.2, 0.25, 0.35 and 0.4 for CK, SF, CG1 and CG2 respectively.

The corresponding scaling values used for each subject to obtain the normalized F0 values from the raw values were 125, 126, 74 and 71 (derived from the difference between the maximum and minimum F0 value in Hz over all ensemble-averaged F0 contours of each speaker). Multiplying these two series of values together gives the value in Hertz estimated to be attributable to the mechanical contribution: 25 for CK, 31.5 for SF, 26 for CG1 and 28 for CG2 (to nearest 0.5 Hz).

If these values are then placed in relation to the F0 differences between vowels seen in Figure 3 it will be seen that this estimated mechanical contribution accounts in the case of speaker CG for essentially the whole range of F0 differences (of the order of 30 Hz), consistent with the absence of vowel-related CT effects for this speaker. On the other hand, for speakers CK and SF under half of the observed F0 differences appears attributable to the mechanical contribution, again consistent with the fact that these were the speakers who did show vowel-related differences in CT activity.

3.2.1 *A brief return to consonant voicing*

In the immediately preceding section we have introduced a technique for comparing F0 at the same EMG level over different sounds. The rationale for such a technique is particularly compelling in the context of vowel intrinsic F0, but we will now also use it to come back to an observation made in the analysis of consonant voicing effects, namely that the F0 differences on the vowel (over differences in pre-vocalic consonant voicing) may be larger than would be expected from the CT activity observable in the vowel itself.

This can now be assessed more quantatitively by simply reapplying the technique in the previous section, after re-arranging the material appropriately. Rather than comparing groups of five vowel categories (four for CK) matched for tenseness and consonant voicing context, we now compare vowel pairs that are matched for vowel category and tenseness but differ in the voicing context.

The results are extremely consistent: For every subject it is quite clearly the case that at the same CT activity level F0 in the vowel is slightly higher in voiceless consonant context. This is summarized in Table 1 by computing the difference in normalized F0 between matched pairs.

Table 1. Pairwise differences between predicted normalized F0 at mean CT activity level for vowels matched for vowel category and tenseness but differing in voicing context. Positive difference values indicate higher F0 in voiceless context. Significance determined by Wilcoxon signed ranks test. $n = 10$ for SF and CG (5 vowels, each tense and lax); $n = 8$ for CK

Subject	mean difference	sd of differences	significance
CK	0.097	0.073	$p < 0.05$
SF	0.103	0.063	$p < 0.01$
CG(1)	0.091	0.043	$p < 0.01$
CG(2)	0.094	0.051	$p < 0.01$

There is a striking parallel to the vowel intrinsic F0 results: This indicates that for all speakers part of the higher F0 following voiceless consonants cannot be directly accounted for by the CT activity level on the vowel itself, i.e. this does not just apply to speaker CG where the CT activity differences on the vowel were relatively weak. It is also striking that the mean differences are very similar over speakers as expressed here in terms of normalized F0, i.e. for all speakers they are about 10% of the F0 range found over the target vowels. This consistency could speak for an effect propagating mechanically from the preceding consonant onto the vowel. As mentioned in the first section of the results, we will consider in the final discussion what the precise mechanism could be.

3.3 Tense vs. lax vowels

In this section we look at the results from the point of view of the tense-lax opposition. The basic hypothesis is that there is higher CT activity in the lax vowels. If confirmed, this would further contribute to the viability of tongue-pull explanations of IF0 since it could explain why the F0 of tense and lax vowels does not differ as expected from knowledge of their respective tongue positions. In other words, hitherto German tense-lax vowels had appeared puzzling for tongue-pull theories because despite the clearly lower tongue position for the lax cognates there are only inconsistent differences in F0. In effect, this discrepancy would only be an apparent one if mechanical effects – on the basis of the previous section (3.2) they clearly seem to be present – were being overlaid by active muscular adjustments that raise F0 above its 'mechanically' expected value in the lax vowels.

We look first at the basic patterns of CT activity by considering the main effect of TENSE in the 3-way ANOVA. In addition, Figure 6 summarizes the trends in the data in a way similar to that in which Figure 1 summarized the results for VOICE: here by computing the difference between matched pairs of lax and tense vowels in such a way that positive values indicate trends consistent with the hypothesis, i.e. the more positive the values the greater the extent to which activity for lax vowels exceeds that for tense vowels.

Overall, the results do point in the direction of more activity for the lax vowels, but the results are less clear-cut than for the effect of VOICE in the first section of the results above. Looking in detail first at the vowel segment (here clearly the crucial segment), for speaker CK the main effect of TENSE was significant at $p < 0.01$, but there was also an interaction with VOICE (at $p < 0.01$). As is clearly to be seen in Figure 6, this is because there is a very large effect in the hypothesized direction in the context of voiced consonants, but effectively no difference in voiceless context. For CG the main effect was significant for both corpora ($p < 0.001$ for CG(1) and $p < 0.01$ for CG(2)). For CG(2) the effect was somewhat more variable over the various vowel categories, reflected in a significant TENSExVOWEL interaction ($p < 0.01$). Finally, for SF, the overall trend is actually contrary to the hypothesis. Figure 6 shows values varying around zero in voiced contexts, but clearly negative for voiceless contexts. The ANOVA details were: TENSE $p < 0.01$ (contrary to the expected direction), TENSExVOICE $p < 0.05$, TENSExVOWEL $p < 0.01$.

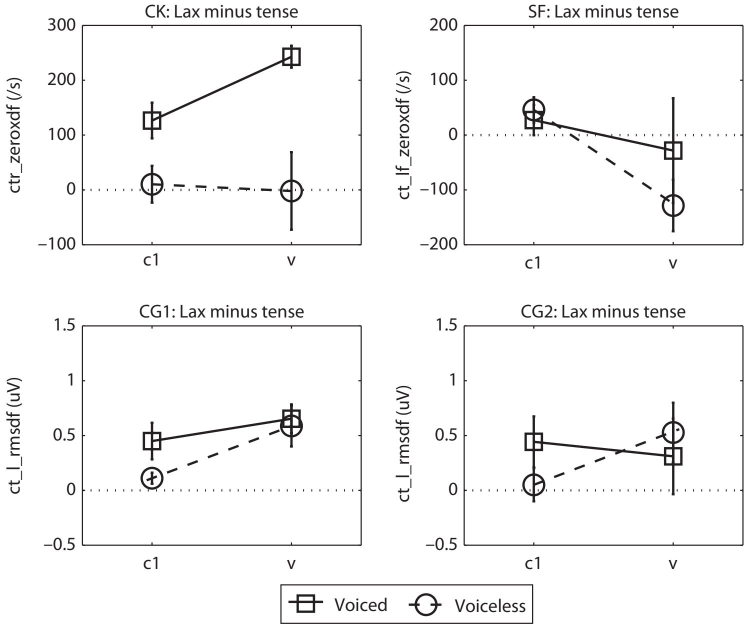

Figure 6. Difference in CT activity between corresponding segments in lax and tense vowel contexts. Positive values (above the dotted line) indicate more activity in lax context. Calculated separately for voiced and voiceless consonant contexts. Other details as for Figure 1

The figure also shows the results for the C1 segment. Clearly, taken on its own this segment is much less important for the basic hypothesis than the vowel segment. Nonetheless, given that the latter segment did show some indication of the expected trend towards more activity for lax vowels, it was considered interesting to examine whether this might already be visible in the syllable onset. And indeed, as seen in Figure 6 and confirmed by ANOVA, the differences are not always very large but nonetheless consistently in favour of more activity in the lax vowel contexts. Note that this applies not only to CK and CG but also to SF.

The ANOVA details for the main effect of TENSE are:

CK and SF $p < 0.01$, CG(1) $p < 0.001$, CG(2) $p < 0.05$.

Significant interactions were:

TENSExVOICE, CK $p < 0.01$, CG(1) $p < 0.05$; TENSExVOWEL, SF $p < 0.01$.

The most appropriate summary of this set of results, in our view, is that an active increase of CT activity on the lax vowels does appear to represent a plausible scenario, but that it would certainly not be fair to claim that it has been conclusively demonstrated. Given the latter concession, we will refrain in the present paper from further discussion of the implications of the results for our understanding of the tense-lax distinction itself. For this the reader is referred to Hoole (2006a) and Hoole & Mooshammer (2002).[7] We would, however, like to emphasize here the possible implications of the tense-lax results for vowel intrinsic F0 in general. At the outset of our study the hypothesis of increased CT activity for the lax vowels simply seemed to represent one logical possibility. But one potential result – given the complete absence of previous relevant data – could have been that none of the subjects showed any indication of this increased activity. This would have been quite a strong indication that previous mechanical accounts of vowel IF0 are either misguided, or at least lacking some major as yet unidentified component. The present results, however, despite being as already admitted not conclusive, indicate that the hypothesis of increased CT activity for the lax vowels cannot be ruled completely out of court. Thus any use of the F0 relationships in the German tense-lax contrast as an argument against mechanical accounts of IF0 must be viewed with caution since clearly a plausible alternative scenario exists that removes the foundation for that argument.

---

[7] Hoole (2006a) also includes some discussion of why the measurement procedures might incorporate a slight bias against finding higher CT activity on the lax vowels (related to interaction of segmental and prosodic influences on CT). But for present purposes it appeared preferable to adhere to procedures that can safely be regarded as conservative with respect to the hypothesis being tested.

4. General discussion

Having considered the special case of vowel IF0 in the German tense-lax distinction at the end of the immediately preceding section we will now in the general discussion move to a consideration of the wider implications of the study for IF0 in general, i.e. with respect to vowel height. The second, and probably more intricate part of the discussion will examine consonant voicing. As will be recalled from the introduction, the theme common to the intrinsic F0 and consonant voicing discussion is what Kingston (1992) has referred to as covariation, i.e. what phonetic properties change in parallel when a given linguistic distinction is realized: Is this covariation mechanical and unavoidable given the properties of the human speech apparatus, or does it reveal that speakers can home in on unrelated articulations as a means for mutually enhancing a specific auditory property of the speech signal, e.g. F1-F0 distance or low-frequency emphasis in the signal?

4.1 Intrinsic F0 and vowel height

We have intimated that our results suggest a compromise solution between the more radical points of view that have been taken in the literature. The crucial feature of the approach was to take seriously the methodological strictures in Whalen et al. (1998) and look closely at the properties of the function linking EMG to F0 over the different vowel categories.

The nature of this relationship showed rather persuasively that a mechanical effect must be at work. Nevertheless, the results also showed that an active difference in CT activity between high and low vowels is certainly also to be reckoned with. As discussed in the introduction, previous evidence for higher CT activity for high vowels has been somewhat ambivalent. The key point here is that even if future investigations were to provide even more consistent evidence that this is a typical laryngeal muscular adjustment for high vowels, then this does not rule out the importance of a mechanical substrate for IF0.

Kingston has argued that many cases of covariation – in his view certainly the one here (but see also the discussion below of the results of his 2007 investigation) – are not based on articulations that are contingent on each other, but rather that speakers make use of independent articulations that can advantageously be combined to emphasize a specific auditory property. Based on the hybrid result found here, we would favour a rather different scenario: Speakers may choose to enhance higher pitch on high vowels, but if they do so the impetus comes from a contingent effect on F0 that is part of normal vowel articulation. Thus speakers may notice that low vowels typically 'come out' with lower F0 than higher ones; having made this identification they then in effect have access to a property that could potentially be heightened – for example, it could become part of their phonetic repertoire for clear speech.

In this perspective we do not necessarily assume that speakers are initially aiming to enhance a specific property like F1-F0 distance in order to make vowel height contrasts more salient for the listener (cf. background to this idea in the introduction above). Nevertheless, this auditorily useful effect could perhaps provide reinforcement when such a covarying pattern starts to emerge.

In short, we suggest that the driving force behind this pattern of covariation is the articulatory contingency, and not the auditory perceptual goal. However, as suggested by Honda & Fujimura (1991) speakers may latch onto these effects and push them further as part of their linguistic behaviour.

4.2 Consonant voicing

We have seen that voiceless consonantal contexts reliably attract more CT activity. Given that previous results in the literature had not been all that consistent the results thus certainly contribute to tipping the balance in favour of CT activity as a typical component of voiceless consonants.

What still needs to be considered in more detail, however, is the precise purpose of this activity.

We saw in the introduction that there is a possible discrepancy between the common assumption that tensing the vocal folds contributes to the suppression of vibration and the fact that the voicing status of the consonant clearly affects the following vowel more than the preceding one: these two aspects would seem to require different timing patterns. As also indicated in the introduction, for this and other reasons Kingston & Diehl had questioned whether the F0 patterns can be directly contingent on laryngeal articulations required for voiceless consonants, and suggested that they could be only related systematically to a more abstractly defined [voice] feature. As reiterated recently by Kingston (2007: 174):

> "F0 is raised next to voiceless unaspirated stops in languages such as French where this is the pronunciation of the [-voice] stops, but lowered next to such stops in languages such as English where this is the pronunciation of [+voice] stops."

In our opinion, however, the strength of the latter argument is questionable: Even though English /b/ and French /p/ may both surface as voiceless unaspirated it is hazardous to assume that their laryngeal adjustments are identical (as is acknowledged by Keating 1984, who otherwise – similarly to Kingston – proposes a rather abstract conception of the phonological feature [voice]). There is actually very little data on laryngeal manoeuvres in French voiceless stops, but the work of Benguerel et al. (1978) indicates clearly the presence of an active glottal abduction gesture (confirmed in as yet unpublished data of our own), whereas such a gesture is probably never present in an English phonologically voiced stop.[8]

Returning to the discrepancy just mentioned (suppression of voicing should target consonant onset, whereas F0 effects are more obvious post-consonantally), we will look at the relevant timing patterns in more detail immediately below, but one simple first step towards removing this discrepancy would be to question the relevance of using increased CT activity to assist the suppression of vocal fold vibration at the onset of voiceless consonants. Given a completely occluded vocal tract and a glottal abductory movement which not only weakens vocal fold vibration directly but also allows a rapid rise of pressure in the supraglottal cavity then conditions for maintenance of voicing deteriorate very rapidly. Thus there is probably no necessity for the speaker to introduce yet a further mechanism to make voicing stop even quicker. The one or two weak residual vibrations that often occur as the glottis opens and pressure builds up are unlikely to be communicatively relevant given that offsets are generally less auditorily salient than onsets, and any sound generated is radiated from an already strongly occluded vocal tract.

Even if there is, then, no need for the CT activity to be focussed on the consonant onset, what mechanism would nonetheless allow its effect to propagate so strongly from some later location in the consonant onto the vowel?

The answer here could be that the delay for the effect of relaxation of CT is longer than that for activation. The delay estimates given in Sawashima et al. (1982) suggest that this is the case, and this possibility is also mentioned in passing by Löfqvist et al. (1989).

This would be consistent with the example in the following figure taken from our own data, where clearly the decline in F0 lags behind the decline in CT activity. In other words, during the segment marked by vertical dashed lines (corresponding to the stressed target vowel) F0 increases more or less in parallel with increasing CT activity. But once F0 has increased it stays at a fairly high level even after CT activity has started to decline noticeably (roughly over the period from 400ms to 600ms on the time axis). This kind of hysteresis effect of F0 relative to CT would provide a mechanism for the propagation of the effect

---

[8] A further indication that these two classes of sounds are clearly distinct laryngeally comes from consideration of clusters: It is readily observable that /r/ devoices substantially in French /pr/ clusters whereas this is not the case in English /br/ clusters. This is mysterious if we do not assume that French stops have an active glottal abduction gesture.

on F0 of increased CT activity to later time locations. Recall from the analysis of predicted F0 for an average EMG level in the vowel that voiceless contexts come out higher than voiced ones (see Section 3.2.1 above). A mechanism of the kind just illustrated by means of Figure 7 could account for this very clear result.[9]

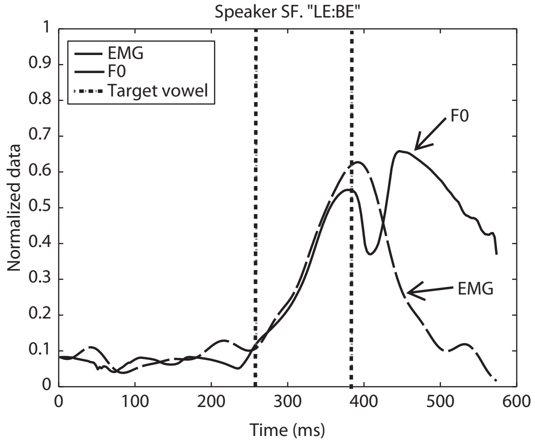

Figure 7. Overlaid normalized F0 and EMG traces centered on the stressed vowel of the target word /leːbə /. Vertical dashed lines mark acoustic onset and offset of the target vowel. The slight downward inflection of the F0 trace around 400ms on the time axis is related to the occlusion phase of /b/

It remains to make a specific suggestion regarding the timing of increased CT activity for the voiceless consonant: We estimate that it may be quite closely linked to the time-course of glottal ab- and adduction. This is illustrated in the

---

[9] There may be an additional candidate for the effect that F0 after voiceless consonants is higher than would be predicted from CT activity in the vowel: In EMG studies that have also recorded the vocalis muscle it is noticeable that after being suppressed for the voiceless consonant this muscle can 'kick in' again quite abruptly at the start of the following vowel. This can be observed, for example, in Löfqvist at al (1984), and also for the aspirated plosives in Hutters (1985). The immediate purpose of the vocalis 'injection' is certainly to ensure strong glottal closure for modal voicing as the vocal folds start to vibrate again, but given that it is activated quite strongly at the consonant-vowel transition it might also contribute to higher F0.

following figure, which shows ensemble averages of CT activity in the vicinity of the target consonant.

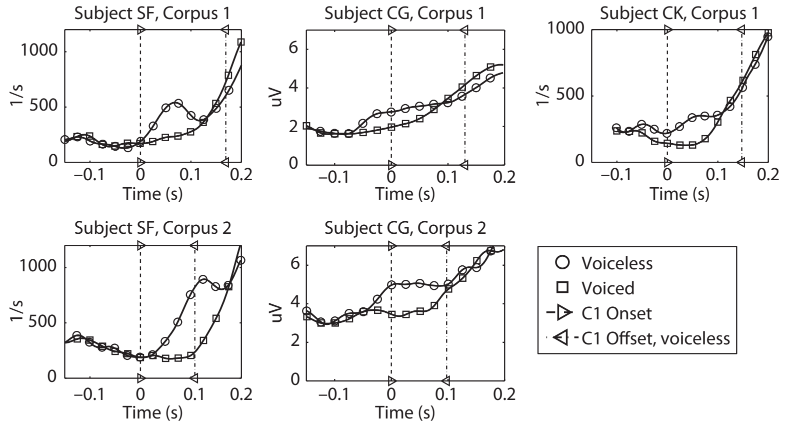

Figure 8. Comparison of ensemble-averaged data for voiced and voiceless consonants. Traces aligned at acoustically defined onset of consonant (= left vertical dashed line, at zero on the time axis). Right vertical dashed line marks offset of the oral occlusion phase of the voiceless consonant: /f/ in corpus 1 (top panels), aspirated /p/ in corpus 2 (bottom panels). Signals based on zero-crossing rate for SF and CK, and on RMS amplitude for CG (see Figure 1 for further explanation of y-axis units)

The averaging was run over all vowel contexts, simply distinguishing by voicing in the consonant. The most interesting aspect is to compare Corpus 1 and Corpus 2 because the consonants can be assumed to differ with respect to the timing of the glottal opening gesture relative to the oral occlusion phase. From transillumination experiments we can expect peak glottal opening to be roughly in the middle of the occlusion phase for the fricative (/f/, Corpus 1) and close to the end of the occlusion phase of the aspirated plosive (/p/, Corpus 2). In the ensemble averages we can observe from Corpus 1 to Corpus 2 a shift in the timing of the location of the maximum difference between the voiceless and voiced curves that could parallel the glottal opening differences, very clearly for SF, less so for CG.[10] The clearer result for SF may be because she has a longer aspiration phase for /p/, averaging about 73ms compared with 40ms for CG. The other point to note in this figure is that for SF and CK (for whom Corpus 2 was not available) the main voiceless-voiced differences are clearly not located at consonant onset. This is not so clear for CG, but it turns out that even this finding has a counterpart in the control of glottal opening: We have recorded extensive transillumination data for SF and CG (see Hoole et al. 2003; more detail in Hoole 2006b): a very stable feature was that CG initiated glottal abduction for voiceless consonants earlier than SF (relative to formation of the occlusion for the consonant).

We believe that this link with glottal timing is also to be observed in the Löfqvist et al. (1989) data: For their English speakers the difference between the voiced and voiceless CT traces is maintained up to a later point with respect to the stop release line-up (for the plosives) than with respect to the frication offset lineup (for the fricatives), reflecting the aspiration phase of the plosives, whereas for the Dutch speaker this difference between fricatives and stops is not apparent (the voiceless stops being essentially unaspirated).

Why could this pattern of coordination be advantageous in the production of voiceless consonants? The basic idea being followed here is nothing more complicated than the formula for the area of a triangle: If half the area of the glottis is assumed to be roughly given by a triangle with the midline of the glottis as its base and the position of the vocal process of the arytenoid cartilage as its apex then greater CT activity could increase the length of the base (or at least stabilize it). Thus we would like to argue that the purpose of the CT activity is to increase the mechanical efficiency of the abductory motion of the arytenoids with respect to the resulting glottal aperture. There may in fact be an interesting side-effect of this manoeuvre with respect to vocal fold tension, but not in terms of suppression of vibration but rather in terms of controlling the re-initiation of vibration at the onset of the following vowel, i.e. at a location that is most likely auditorily more salient and communicatively more relevant. The following scenario is based on experiments by Hanson & Stevens (2002) on the control of voicing in articulatory synthesis. It seems particularly plausible in the light of the discussion above as to how effects of CT activation can propagate to later points in time.

As one of their examples they discuss the synthesis of voiceless aspirated plosives and concentrate particularly on modulation of the compliance of the vocal folds. They find it useful to modulate the compliance roughly in parallel with glottal opening, the aim being "to gain more control over the *onset* of voicing" (p. 1173; our emphasis). The aerodynamic background to this is that intraoral pressure is assumed to drop rapidly at release, and the model predicts that under default conditions the vocal folds would be able to vibrate when the glottal area has reduced to about 12mm^2, i.e. before the glottis has completely closed. By reducing the compliance, phonation threshold pressure is raised and the vocal folds are not able

---

[10] It seemed justifiable to use data from corpus 2 of SF for this purpose because the ensembles are not sub-divided by vowel category. So a substantial amount of data is available for averaging (about 30 tokens each for the voiced and voiceless contexts).

to vibrate until the glottal area (in this example) has further reduced to about 5.5mm^2. These measures result acoustically in a fast, clear transition from voiceless aspiration noise to modal phonation.

4.3 A different approach to the study of enhancement

Before coming to the final conclusions of our investigation it is of interest to consider the results of a completely different approach to weighing up the roles of automaticity and enhancement for precisely those F0 phenomena that have been focused on here. Kingston (2007) looks to find evidence for enhancement as predicted from his earlier work discussed above. We look first at the design and the results for IF0 related to vowel height, and then, secondly and more briefly, at F0 effects related to consonant voicing.

The essential feature of the approach involves analyzing how IF0 strength waxes and wanes as the prosodic context of the target vowels is manipulated, the basic argument behind this being that an automatic effect (such as tongue-pull) should not vary with prominence (assuming non-prominent syllables still have full vowel quality), whereas it would be very natural that an active enhancement effect would strongly target prominent vowels with presumably high information load. This thus forms an interesting counterpart to our own study (which focuses on the physiological substrate), since the approach is now explicitly via the assumed communicative relevance of enhancement. The expectation that the approach could lead to evidence against the automaticity of IF0 follows up the earlier work of Kingston mentioned in the introduction, and was specifically derived from various indications in the literature that IF0 is only present in intonationally prominent syllables (e.g. Ladd & Silverman 1984; Steele 1986).

However, non-prominent syllables often have low F0, and it is certainly not immediately clear that the tongue-pull mechanism necessarily has the same effect on F0 over the full pitch range of the speaker (see Whalen & Levitt 1995, for more discussion; one possibility is that in low tones the activity of the strap muscles may counteract the tongue-pull mechanism). Kingston himself discusses the case of tone languages where it appears that IF0 is more apparent for high tones than low tones. The crucial test, then, is to compare material where both high and low tones can be both prominent and non-prominent, and determine whether strength of IF0 effects follows from prominence more closely than from simple pitch height. Steele (1986) had found indications that non-prominent high tones might not show IF0 effects (see also Reinholt Petersen 1978, for prominent low tones in Danish). Steele argued that there might be interactions between the strength of tongue-pull effects on IF0 and the level of subglottal pressure, thus keeping IF0 within the scope of articulatorily contingent effects. But assessment is hampered here by the lack of relevant physiological data on subglottal pressure (probably even more acute a lack than in the case of muscle activation data).

The pattern of results in Kingston's own experiments was quite complex, but in the more extensive experiment (with naive subjects) the results appeared more consistent with the automatic account, i.e. contrary to the original hypothesis. This preliminary conclusion was cross-checked by looking for evidence that speakers were actually using articulatory adjustments to mark prominent syllables. If this were not the case then the experiment may simply have been unsuccessful in eliciting appropriate utterances. In fact, F1 values indicated that prominent syllables had actually been given articulatory prominence, supporting the above conclusion: The speakers simply had not enhanced vowel-related F0 differences on syllables that were indeed prominent, so any F0 differences present have best to be assumed to be automatic. Nevertheless, the story may not stop here: Although the speakers were clearly distinguishing prominent and non-prominent syllables, there may not have been a genuine modulation of local information content.

Turning now more briefly to the effect of consonant voicing on vowel F0, Kingston's (2007) experiments applied the same logic of prosodic manipulation as just discussed for vowel height effects. Unfortunately, the results were inconclusive: No positive evidence in favour of the controlled hypothesis was found. But in the most extensive experiment the consonant-related F0 effects were overall very weak, which is also an awkward result for the automatic account (and unexpected in the light of the bulk of the previous literature).

5. Conclusion

In this concluding section let us first briefly summarize the principal results, and then consider the implications for theories of enhancement.

With regard to consonant voicing there was a robust finding of higher CT activity for the voiceless consonants. It was suggested that the time course of this increased CT activity is quite closely linked to the time course of glottal ab- and adduction for the consonant. A possible physiological mechanism was put forward to explain how differences in F0 on the following vowel can result, even if differences in CT activity are confined to the consonant itself. The more variable part of the results for the voicing contrast was in fact that differences in CT activity may also occur on the vowel, but less consistently than on the consonant.

With regard to vowel height and IF0 the robust result, emerging from the regression analysis, was that a mechanical influence of vowel articulation on F0 can be assumed. Once again, this robust result was accompanied by a more variable result (i.e. less constant over speakers), namely that mechanical influences on F0 may be overlaid by muscular adjustments acting in the same direction.

The third linguistic opposition considered here, namely tense vs. lax, gave least evidence of robust results. The hypothesized higher CT activity for the lax vowels could not be conclusively demonstrated. Nevertheless, indications were found that this scenario is a realistic one, which means in turn that it may not be justified to use F0 patterns in German tense vs. lax vowels as an argument against mechanical accounts of the general IF0 phenomenon. Thus even these less clear results fit in well with the main findings on vowel height and F0.

We will now consider the three linguistic oppositions more specifically from the point of view of enhancement. We will reverse the order of the above summary to consider first, and more briefly, the tense-lax opposition.

For reasons given in detail elsewhere (Hoole & Mooshammer 2002) we believe that a key characteristic of lax vowels in German is the close coupling between the vowel and the following consonant. Taken together with the short duration of a complete CVC movement this results in the force input for the supraglottal movements being concentrated strongly in the centre of the vowel. For this we have used the term ‘pulsatile’ force input, contrasting with ‘distributed’ force input in the tense vowels. Increased CT activity for the lax vowels could be seen as a ‘spilling over’ of the pulsatile force pattern to the laryngeal system. With regard to enhancement the situation is particularly unclear because even if the potentially increased CT activity is responsible for the unexpectedly high F0 in lax vowels then it is unknown at present whether this is actually useful to listeners in terms of enhancing the salience of the vowel oppositions (e.g. to use an example from the introduction, /eː/ vs. /ɪ/). Given the current state of the data it would be premature to consider whether the situation for the tense-lax distinction is compatible with the distinction made by Keyser & Stevens (2006) between feature-defining and feature-enhancing gestures, and their contention that feature-enhancing gestures are in a sense more robust, since they may remain present precisely in the cases where feature-defining gestures are in danger of being obliterated in heavily overlapped running speech.

Turning now to to vowel-height and IF0, discussion is more straightforward because of the clearer pattern in the results.

To the extent that IF0 is a purely mechanical effect it would not be appropriate to regard it as an enhancing gesture, since in both the approaches of Keyser & Stevens, as well as of Kingston & Diehl enhancement clearly implies some form of additional planning by the speaker. To the extent that IF0 can also involve an active muscular contribution then enhancement would be an appropriate term to use. Our main difference from Kingston is that we nevertheless emphasize the articulatory contingency as the source of the enhancing effect. With respect to Keyser & Stevens, active regulation of IF0 appears at first sight to fit in with their approach since they clearly envisage a situation in which the realization of an opposition acquires an acoustic property (here F0) not related to the defining features of the opposition (for vowel height perhaps expressed in terms of formant frequencies). It is less clear whether a further component of their approach is supported here (already alluded to immediately above in connection with the tense-lax distinction): It seems unlikely that IF0 as enhancement will prove more robust than basic vowel height differences. In other words, it seems unlikely that it would be introduced by speakers in graded fashion just in cases where the salience of defining gestures is in danger. For example, speaker CG showed no active enhancement of IF0. We suspect that he would not then start to make use of this strategy in a different situation involving less clear lingual realization of vowel height. Our guess is rather the reverse (but clearly should be tested over more widely varying speaking situations): speakers CK and SF showed a good deal of enhancement precisely because they viewed the experimental situation as calling for a very clear speaking style; CG does not use it *even* in a typical clear-speech situation, and so most likely does not use it at all. Thus further differentiation of the areas in which enhancement can manifest itself may be called for.

Turning finally to the effects of the consonantal voicing contrast, the discussion was devoted in large part to making a case for F0 differences on the vowel as emerging from typical articulatory patterns for the preceding consonant. The discussion was perhaps somewhat more intricate than for intrinsic F0 effects in vowels. So it is important to emphasize that the ultimate conclusion we would like to reach is nonetheless exactly equivalent: The driving force comes from the articulatory contingency, but once established speakers can deliberately emphasize its effects. There was evidence of this in the EMG patterns of speakers CK and SF, where at least in the case of the tense vowels, not just the consonantal phase but also the vocalic phase exhibited robust EMG differences. So it seems possible that our speakers on these occasions are actively enhancing the difference during the vowel segment.

This potential enhancement of an articulatorily driven pattern is, as discussed in the introduction, relevant to the more general issue of tonogenesis: the development of a tonal contrast from a voicing contrast logically requires a process of enhancement of the F0 effects of the voicing contrast, since otherwise the incipient tonal contrast would not outlive the loss of the voicing contrast itself.

We have just emphasized the parallelism of the vowel IF0 and consonant voicing results; there is, nonetheless an important difference between enhancement related to consonant voicing and that related to vowel IF0: The core of the basic vowel IF0 effect is, as repeatedly emphasized, completely mechanical, and does not require the intervention of CT activity at all. By contrast, there is of course a difference in CT activity at the core of the consonant voicing effect; it is rather its propagation to the vowel that probably has to be regarded as mechanical. And of course the proposed close connection between the time course of CT activity and glottal abduction in voiceless consonants is also not a mechanical link in the sense in which movement of the tongue for vowel articulation can mechanically influence position of the hyoid bone and the laryngeal cartilages. In languages such as Hindi the voiced aspirated plosives require a combination of glottal abduction with very *low* CT activity in order to support the continuation of voicing with slack vocal folds (Dixit & MacNeilage 1980). Coming back to the framework proposed by Keyser & Stevens, they point out that potential enhancement of consonant voicing effects through F0 is a different type of enhancement to that discussed in the introduction with respect to rounding of /ʃ/. In the latter case, a completely new articulatory gesture (rounding) is introduced which serves to enhance the saliency of the already present anteriority-related spectral differences between /s/ and/ʃ/. In the case of consonant voicing, by contrast, it is the defining feature (in their view) of [+stiff vocal folds] that can manifest itself in qualitatively different acoustic properties: voicelessness on the consonant, raised F0 on the vowel, the latter then constituting enhancement because the details of the F0 contour (including, perhaps, language dependence) may be under speaker control (Stevens & Keyser 2009). The observations made in the present investigation are broadly consistent with this point of view. Our main difference from Stevens and Keyser with respect to consonant voicing is that we would assign a much more central role to the timing of the glottal abduction gesture in the realization of voicing distinctions (in a tradition going back at least to Löfqvist 1980). In their framework (see especially Keyser & Stevens 2006, p. 42), the feature [spread glottis] is not relevant for the phonological representation of English obstruents, but rather is introduced as an enhancement gesture (enhancing the defining feature of [+stiff vocal folds]) in cases where a plosive is realized with aspiration. We have tended to emphasize the parallelism between these two laryngeal adjustments. In other words, for languages such as English and German but also for languages such as French and Dutch in which the voiceless consonants are traditionally regarded as unaspirated, we assume that glottal ab/adduction and CT activity form a close functional synergy, in which a clear division into a defining and enhancing feature is not straightforwardly possible. Final resolution of this issue is of course beyond the scope of the present investigation (but see Hoole 2006a, for a small amount of additional data and discussion).

As a final conclusion uniting both the consonant voicing and the vowel IF0 issues we suggest that it is a general feature of movement planning to take advantage of physical forces, where possible ("go with the flow"). As a guideline for future work, it can be hypothesized that whenever an effect is assumed to be automatic and mechanical then it should be possible to demonstrate that it is fairly constant over speakers. On the other hand, the adoption of enhancement strategies will probably be more variable, reflecting the fact that speakers differ in clarity and, perhaps, in their sensitivity to acoustic differences.

Acknowledgements

Thanks to Emi Murano for insertions, Rafael Laboissière for lab facilities, Suse Fuchs for logistical support. Thanks to Anders Löfqvist, Doug Whalen, Rachid Ridouane and above all Nick Clements for comments on an earlier version of the manuscript.

Appendix

3-way ANOVA analysis of CT activity for each subject and segment. Tabulation of F, degrees of freedom and p values for each main effect and two-way interaction.

	C1 Segment							
	CK		**SF**		**CG1**		**CG2**	
	F(DoF)	**p**	**F(DoF)**	**p**	**F(DoF)**	**p**	**F(DoF)**	**p**
Voice	41.98 (1,140)	< 0.0001	272.45 (1,180)	< 0.0001	77.6 (1,180)	< 0.0001	91.4 (1,180)	< 0.0001
Vowel	6.94 (3,140)	0.0002	4.82 (4,180)	0.001	6.12 (4,180)	0.0001	8.66 (4,180)	< 0.0001
Tense	10.89 (1,140)	0.0012	9.26 (1,180)	0.0027	18.01 (1,180)	< 0.0001	4.1 (1,180)	0.0444
Voice x Vowel	1.02 (3,140)	0.3847	3.82 (4, 180)	0.0052	3.56 (4,180)	0.0081	0.49 (4,180)	0.7434
Voice x Tense	7.88 (1,140)	0.0057	0.59 (1,180)	0.4426	6.59 (1,180)	0.0111	2.6 (1,180)	0.109
Vowel x Tense	0.47 (3,140)	0.703	3.59 (4,180)	0.0076	1.24 (4,180)	0.2941	1.41 (4,180)	0.2335

	Vowel Segment							
	CK		SF		CG1		CG2	
	F(DoF)	p	F(DoF)	p	F(DoF)	p	F(DoF)	p
Voice	9.7 (1,140)	0.0022	24.33 (1,180)	< 0.0001	1.93 (1,180)	0.1668	10.66 (1,180)	0.0013
Vowel	14.19 (3,140)	< 0.0001	21.03 (4,180)	< 0.0001	1.58 (4,180)	0.1823	18.3 (4,180)	< 0.0001
Tense	10.61 (1,140)	0.0014	10.1 (1,180)	0.0017	18.84 (1,180)	< 0.0001	8.6 (1,180)	0.0038
Voice x Vowel	0.14 (3,140)	0.9377	2.96 (4,180)	0.0212	0.36 (4,180)	0.8374	1 (4,180)	0.4103
Voice x Tense	10.97 (1,140)	0.0012	4.14 (1,180)	0.0434	0.05 (1,180)	0.8186	0.58 (1,180)	0.4463
Vowel x Tense	1.24 (3,140)	0.2986	6.88 (4,180)	< 0.0001	0.13 (4,180)	0.971	4.28 (4,180)	0.0025

References

Abramson, Arthur & Leigh Lisker. 1985. Relative power of cues: F0 shift versus voice timing. In Victoria Fromkin (ed), *Linguistic Phonetics, Essays in honor of Peter Ladefoged*, 25–31. New York: Academic.

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

Baer, Thomas. 1981. Investigation of the phonatory mechanism. In Christy L. Ludlow & Molly O'Connell Hart (eds), *Proceedings of the conference on the assessment of vocal pathology, = ASHA Reports* 11. 38–46.

Benguerel, André-Pierre, Hajime Hirose, Masayuki Sawashima & Tatsujiro Ushijima. 1978. Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study. *Folia Phoniatrica* 30. 175–198.

Clancy, A Edward, Evelyn L. Morin & Roberto Merletti. 2002. Sampling, noise-reduction and amplitude estimation issues in surface electromyography. *Journal of electromyography and kinesiology* 12. 1–16.

Collier, René, Leigh Lisker, Hajime Hirose & Tatsujiro Ushijima. 1979. Voicing in intervocalic stops and fricatives in Dutch. *Journal of Phonetics* 7. 357–373.

Connell, Bruce. 2002. Tone languages and the universality of intrinsic FO: Evidence from Africa. *Journal of Phonetics*, 30 (1). 101–129.

De Cheveigné, Alain & Hideki Kawahara. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America* 111 (4). 1917–1930.

Diehl, Randy L. 1991. The role of phonetics within the study of language. *Phonetica* 48. 120–134.

Dixit, Prakash R. & Peter F. MacNeilage. 1980. Cricothyroid activity and control of voicing in Hindi stops and affricates. *Phonetica* 37. 397–406.

Dyhr, Niels. 1990. The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. *Phonetica* 47 (3–4). 141–154.

Fischer-Jørgensen, Eli. 1990. Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica* 47. 99–140.

Fromkin, Victoria A. (ed). 1978. *Tone: A linguistic survey*. Academic Press.

Fromkin, Victoria A. & Peter Ladefoged 1966. Electromyography in Speech Research. *Phonetica* 15. 219–242.

Haggard, Mark, Stephen Ambler & Mo Callow. 1970. Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47. 613–617.

Halle, Morris & Kenneth N. Stevens. 1971. A note on laryngeal features. *Quarterly progress report* 101, 198–212. Cambridge, MA: Research Laboratory of Electronics, MIT.

Hanson, Helen M. & Kenneth N. Stevens. 2002. A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *Journal of Acoustical Society of America* 112 (1). 1158–1182.

Hanson, Helen M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *Journal of the Acoustical Society of America* 125 (1). 425–441.

Herman, Rebecca, Mary Beckman & Kiyoshi Honda. 1999. Linguistic models of F0 use, physiological models of F0 control, and the issue of ''mean response time''. *Language and Speech* 42. 373–399.

Hirano, Hirano, Minoru & John Ohala. 1969. Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech Language and Hearing Research* 12. 362–373.

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

Honda, Kiyoshi. 1983. Relationship between pitch control and vowel articulation. In Diane M. Bless & James H. Abbs (eds), *Vocal Fold Physiology*, 286–299. San Diego: College-Hill Press.

Honda, Kiyoshi & Osamu Fujimura. 1991. Intrinsic vowel F0 and Phrase-final F0 lowering: Phonological vs. Biological Explanations. In Jan Gauffin & Britta Hammerberg (eds) *Vocal Fold Physiology,* 149–157. San Diego: Singular Publishing Group.

Honda, Kiyoshi. 2004. Physiological factors causing tonal characteristics of speech: from global to local prosody. *Speech Prosody 2004 Nara*, 739–744.

Hoole, Philip & Christine Mooshammer. 2002. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles & Helmut Spiekermann (eds), *Silbenschnitt und Tonakzente*, 129–152. Tübingen: Niemeyer.

Hoole, Philip, Susanne Fuchs & Klaus Dahlmeier. 2003. Interarticulator timing in initial consonant clusters. In Sallyanne Palethorpe & Marija Tabain (eds), *Proceedings of the Sixth International Seminar on Speech Production, Macquarie University: Sydney, Australia*, 101–106.

Hoole, Philip. 2006a. Experimental studies of laryngeal articulation. Part I: Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing. Unpublished habilitation thesis, Ludwig-Maximilians-Universität, Munich. http://www.phonetik.uni-muenchen.de/~hoole/pdf/habilemg_chap_all.pdf (accessed 4.3.2011).

Hoole, Philip. 2006b. *Experimental studies of laryngeal articulation. Part 2: Laryngeal-Oral Coordination in Consonant Sequences.* Unpublished habilitation thesis, Ludwig-Maximilians-Universität, Munich. http://www.phonetik.uni-muenchen.de/~hoole/pdf/habilpgg_chap_all.pdf (accessed 4.3.2011).

Hutters, Birgit. 1985. Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42. 1–24.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60 (2). 286–319.

Keyser, Samuel J. & Kenneth N. Stevens. 2006. Enhancement and overlap in the speech chain. *Language* 82 (1). 33–63.

Kingston, John & Randy Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kingston, John. 1992. The phonetics and phonology of perceptually motivated articulatory covariation. *Language and Speech* 35. 99–113.

Kingston, John. 2007. Segmental influences on F0: Controlled or automatic? In Carlos Gussenhoven & Tomas Riad (eds), *Tones and Tunes*, Vol. 2. 171–210. Mouton de Gruyter.

Kohler, Klaus J. 1985. F0 in the perception of lenis and fortis plosives. *Journal of the Acoustical Society of America* 78. 21–32.

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

Larson, Chuck, Gail B. Kempster & Michael Kistler. 1987. Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. *Journal of Speech and Hearing Research* 30. 552–558.

Löfqvist, Anders. 1980. Interarticulator programming in stop production. *Journal of Phonetics* 8. 475–490.

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

Löfqvist, Anders, Nancy S. McGarr & Kiyoshi Honda. 1984. Laryngeal muscles and articulatory control. *Journal of the Acoustical Society of America* 85. 951–954.

Ohala, John. 1978. Production of tone. In Victoria Fromkin (ed), *Tone: A linguistic survey*, 5–39. Academic Press.

Ohde, Ralph N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *Journal of the Acoustical Society of America* 75. 224–230.

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

Rischel, Jørgen & Birgit Hutters. 1980. Filtering of EMG signals. *ARIPUC (Copenhagen)* 14. 285–316.

Sapir, Shimon, Michael D. McClean & Erich S. Luschei. 1984. Time relations between cricothyroid muscle activity and the voice fundamental frequency (F0) during sinusoidal modulations of F0. *Journal of the Acoustical Society of America* 75. 1639–1641.

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

Sawashima, Masayuki, Hajime Hirose, Hirohide Yoshioka & Shigeru Kiritani. 1982. Interaction between articulatory movements and vocal pitch control in Japanese word accent. *Phonetica* 39 (4–5). 188–198.

Schiefer, Liselotte. 1986. F0 in the production and perception of breathy stops: Evidence from Hindi. *Phonetica* 43. 43–69.

Silverman, Kim.1986. F0 segmemtal cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43. 76–91.

Steele, Shirley. 1986. Interaction of vowel F0 and prosody. *Phonetica* 43. 92–105.

Stevens, Kenneth N. & Samuel J. Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38. 10–19.

Traunmüller, Hartmut. 1981. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69. 1465–1475.

Whalen, Doug H., Arthur S. Abramson, Leigh Lisker & Maria Mody. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47 (1–2). 36–49.

Whalen, Doug H., Arthur S. Abramson, Leigh Lisker & Maria Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America* 93. 2152–2160.

Whalen, Doug H., Bryan Gick, Masanobu Kumada & Kiyoshi Honda. 1998. Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27 (2). 125–142.

Whalen, Doug H. & Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23 (3). 349–366.

Wood, Sidney. 1982. X-ray and model studies of vowel articulation. *Working Papers 23*, Lund University.

Zhou, Ping & William Z. Rymer. 2004. Can standard surface EMG processing parameters be used to estimate motor unit global firing rate? *Journal of Neural Engineering* 1. 99–110.