Experimental studies of laryngeal articulation

Philip Hoole

Part I: Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing

Experimental studies of laryngeal articulation

Habilitationsschrift
zur Erlangung der Venia Legendi für das Fach Phonetik

Fakultät für Sprach- und Literaturwissenschaften
der Ludwig-Maximilians-Universität

vorgelegt von

Dr. phil. Philip Hoole

München, Oktober 2006

Part I: Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing

Preface

This study presents two sets of experiments on laryngeal function, entitled "*Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing*" (Part I), and "*Laryngeal-Oral Coordination in Consonant Sequences*" (Part II)[1].

Instrumental studies of articulatory processes are inevitably a collaborative endeavour, and it is now a great pleasure to acknowledge all those involved in these experiments:

The EMG experiments grew out of a set of pilot experiments carried out at ATR Labs, Kyoto, after I had approached Kiyoshi Honda with the idea of performing laryngeal EMG to find out whether German really is a problem for theories of intrinsic pitch. I am very grateful to him for the open ear he lent to my first tentative suggestions. However, it would never have been feasible to carry out a full set of EMG recording sessions in Germany if I had been the only person interested in this kind of data. The crucial elements, in addition to Kiyoshi Honda and Emi Murano who did the needlework, were the interest and support of the phonetics lab of Zentrum für Allgemeine Sprachwissenschaft, Berlin, together with the facilities of Rafael Laboissière's lab for sensorimotor coordination at the Max-Planck-Institute for Psychological Research, Munich. This gave us the opportunity to record data relevant not just for the questions I was originally interested in, but in particular also for lingual and mandibular function. Special recognition is owed here to Christian Kroos who was the first subject for both laryngeal and lingual EMG and thus boldly went - as far as I am aware - where no German subject had gone before, at least in the context of hooked-wire EMG for phonetic research.

The second group of experiments, which used transillumination and fiberoptic filming to study laryngeal kinematics, was carried out in Berlin at the ZAS phonetics lab, using an experimental setup I had implemented some years ago. The specific topic of consonant sequences which is focussed on here represents just one aspect of the many recordings we have carried out together over the last few years. The long-suffering subjects for the present recordings were Ralf Winkler, Suse Fuchs and Christian Geng (the latter two also gave freely of their neck and tongue for the EMG experiments). Jörg Dreyer kept the lab running smoothly, and Dr. Klaus Dahlmeier wielded a mean fiberscope.

Finally, this work is dedicated to Hans-G. Tillmann on the occasion of his retirement. His unflagging enthusiasm for our subject over all the many years that he was head of phonetics at Munich have been a tremendous motivation.

[1]Part II includes as an appendix the extended version of a very old conference paper on velar and glottal activity in Icelandic. I prepared this extended version following the conference but never published it. It seemed worth including here because the present experiments in effect apply to German some ideas originally developed in that paper.

Table of Contents for Part I

1 Introduction

The main motivation for embarking on this investigation of laryngeal EMG (specifically the cricothyroid muscle) was that this seemed the most direct route towards resolving a longstanding problem with regard to the phenomenon of vowel intrinsic pitch: In German there is surprisingly little difference in F0 between tense-lax vowel pairs. On the basis of typical mechanical models of intrinsic pitch one would expect lower F0 in the lax vowels. If it could be shown that in lax vowels there is muscular activity at the laryngeal level actively raising F0 then this would be an interesting contribution to our understanding of the phonetic implementation of the German vowel system. However, the implications would go beyond German since, in so doing, it would reinforce the viability of mechanical models of intrinsic pitch. As will be seen, evidence for an effect in the hypothesized direction was found.

While the immediate impetus for the investigation came from the specific question of intrinsic pitch in tense-lax vowel contrasts the recorded material also, of course, provides much relevant information on intrinsic pitch in general, i.e the extremely regular feature of higher F0 in high vs. low vowels. The main alternative to mechanical models of intrinsic pitch is that it represents a strategy by the speaker to actively enhance vowel distinctiveness (e.g through F1-F0 distance). The availability of EMG information makes it possible to weigh up alternative explanations directly:

> Do speakers show more laryngeal EMG activity for high vowels than low vowels - a necessary but (as we will see) perhaps not sufficient condition for the presence of active enhancement?
>
> At comparable EMG levels is F0 higher for high vowels than for low vowels - a necessary consequence of a mechanical explanation?

It turns out that these explanations are certainly not mutually exclusive: a mechanical effect is certainly present, but some speakers appear to actively reinforce it. The results thus give a more balanced picture of an extremely pervasive phonetic phenomenon.

The EMG experiments planned for the intrinsic pitch questions were further exploited as an opportunity to address an additional question of very broad relevance to the phonetic implementation of linguistic distinctions, namely the role of vocal fold tension in the voicing distinction for consonants. There are known to be very robust differences in the F0 following voiced and voiceless consonants, and it has been hypothesized that regulating vocal fold stiffness is a crucial element in regulating the presence or absence of voicing in consonants. However, the EMG evidence for corresponding active laryngeal adjustments is still rather fragmentary - and for German has not been investigated at all. Thus alternative explanations cannot yet be considered completely ruled out. This area, too, raises interesting questions about the extent to which a basic pre-existing effect may become enhanced for linguistic purposes:

> Are F0 differences in the vowel more extensive than they would be if they were just 'fall-out' from the task of regulating voicing in the consonant?

To answer such questions any increase in the pool of relevant EMG data is highly desirable. It will be seen that the results confirmed very clearly that the presence of voiceless consonants attracts a higher level of cricothyroid activity. Interpreting the exact *purpose* of this higher activity hinges rather crucially on its precise timing. Discussion of this turns out to be somewhat less straightforward.

Although this investigation was planned to shed further light on two apparently unrelated issues (vowel intrinsic pitch, consonant voicing) we encountered an unexpected effect running through the whole investigation, namely that there was often a statistically significant interaction between the two main independent variables - tenseness of the vowel, voicing of the consonant - such that the voicing effect was clearer before tense vowels, the tenseness effect clearer after voiced consonants. This has interesting implications for how a specific physiological system, in this case the cricothyroid muscle system, reacts to multiple linguistic influences - and also has implication on the methodological side for how investigations of this kind should be designed in the first place.

The outline of the work is as follows:

Chapters 2 and 3 review previous investigations and present the detailed motivation for the present investigation, Chapter 2 with respect to intrinsic pitch (starting with a general overview, and then going on to the special status of German as a test case), Chapter 3 with respect to the relationship between consonant voicing and F0.

Chapter 4 first gives some general background to electromyography, then outlines the experimental procedures used (including discussion of various issues in the processing of EMG data), and in particular summarizes the linguistic material and the basic characteristics of the EMG signals available for each subject.

Chapter 5, which presents the results, is the most extensive chapter. First of all, the results relevant to the issue of consonant voicing are covered. This order was chosen because the results for this topic are initially somewhat more clear-cut and make it easier to get to grips with the various analysis techniques used throughout the chapter. Following this, the results for the tense-lax opposition are presented and then the more general intrinsic-pitch question of a possible difference in laryngeal activity between high and low vowels. Finally, a brief section is devoted to a very specific question related to the functional organization of the cricothyroid muscle, namely whether the two compartments of this muscle, the pars obliqua and pars recta, show differential activation in linguistically relevant tasks (the two compartments probably correspond in turn respectively to translation and rotation at the cricothyroid joint). For example, it has been suggested that pars obliqua activity corresponds to relatively slow changes in F0. This question was not introduced in the preceding paragraphs because it did not form part of the initial motivation for the experiment, but as the experiments were being planned it was decided to use them as an opportunity to also try and address this question. In fact, in the end relevant data was only successfully acquired for one subject, so, given this slim database, presentation both of the background and the results is concentrated in this single section.

Chapter 6 provides the concluding discussion of the complete results, relating them to the previous state of our knowledge and discussing avenues for future research.

2 Intrinsic Pitch in vowels: German as a test case

2.1 Introduction

In this section we will discuss the explanations that have been offered for the widespread phenomenon of intrinsic pitch (IF0) in vowels, and on this background will show the motivation for our own experiments by outlining why German appears to be a good test case - from the point of view both of our understanding of intrinsic pitch and also of our understanding of the articulatory implementation of the German vowel system - particularly the so-called tense-lax opposition[2].

The basic phenomenon of IF0 is simply stated: This is that high vowels such as [i] and [u] tend to have a higher fundamental frequency than low vowels. In an extensive cross-language review (to be discussed in more detail immediately below) Whalen & Levitt (1995) come up with average figures for the magnitude of the effect of about 15 Hz or 1.5 semitones; thus the effect is not huge, but by no means negligible[3].

IF0 has attracted considerable attention because it presents a very interesting challenge to our understanding of how the various components of the speech production process interact, and potentially has significant implications for our understanding of how linguistic distinctions are signalled.

2.2 IF0. The basic issue: mechanics vs. active enhancement

A convenient starting point for more detailed discussion is the just-mentioned paper by Whalen & Levitt. They assembled all possible data from the literature that gives information on IF0, the resulting coverage amounting to 31 languages from quite a large cross-section of the world's main language families. The main conclusion was that IF0 is universally present in the languages

---

[2]We will use the terms 'tense' and 'lax' as a convenient means of referring to this characteristic property of the German vowel system, but without initially presupposing a specific phonological analysis or assuming a specific phonetic implementation.

[3]Here a tribute to E.A. Meyer is in order: he is credited by Whalen & Levitt as being the first to notice the effect (Meyer, 1896/1897). He was also probably the first to make one of the key observations about the articulation of the German vowel system that play a central role in the discussion below.

of the world. There is a very specific theoretic thrust behind this paper, i.e it is not just an exercise in data collection: the authors use the apparent universality of IF0 as an argument in favour of the idea of IF0 as a mechanical consequence of vowel articulation, and as an argument against an alternative hypothesis (put forward, for example, by Kingston (1992); further references in Whalen & Levitt) that IF0 is introduced by speakers deliberately to enhance vowel contrasts (there is thus a parallel here to the topic of the next chapter, that of F0 differences related to consonant voicing: is this an inevitable by-product of turning voicing on and off, or do speakers enhance F0 differences on the vowel to provide more robust signalling of the voicing contrast to the listener?) . The enhancement hypothesis would suggest that speakers actively raise F0 on e.g [i], which at first sight should be detectable in cricothyroid activity. The evidence for vowel-specific differences in CT-activity turns out to be conflicting, and will be looked at in more detail below (there is a further paper by Whalen and co-authors specifically on this issue). In order to put this into perspective it is necessary first to consider the quite long history of more mechanically oriented explanations for IF0. And before leaving Whalen & Levitt's paper it is worth giving a more specific idea of why the simple pervasiveness of IF0 increases the plausibility of mechanical explanations, and decreases that of active enhancement.

- First, magnitude of IF0 does not seen to depend on vowel inventory size. If it were an enhancement effect selectively employed by specific languages, then one might expect it to be more salient in languages with large inventories, i.e where finer contrasts need to be made.
- IF0 appears to be reliably present in tone languages (but see Connell (2002) for more discussion). If it were under speaker control then one might expect it to be eliminated as a potential source of interference with tonal perception. At any rate, it is not clear why e.g Mandarin speakers, who have a simple vowel system, would want to use IF0 to enhance vowel height perception when it could conflict with reliable tonal perception.
- Whalen's perspective on speech perception follows that of Fowler (to which we will have occasion to return below (p. 14); see especially Fowler & Brown (1997); also Fowler (1981)). In this perspective, with which we have considerable sympathy, listeners are able to parse F0 into vowel-production and tone-production related components, so it would not be necessary to expect tone perception to be compromized.
- IF0 effects tend to disappear in the lower region 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.
- We could also include here an argument frequently quoted in favour of an active component in F0, which is, however, given very short shrift by Whalen & Levitt: this is the intriguing finding that speakers of oesophageal speech may exhibit IF0, i.e in the absence of a larynx. However, Whalen & Levitt argue that this says absolutely nothing about the background to IF0 in normal speech: oesophageal speakers may simply discover that introducing IF0 deliberately helps to improve the naturalness of their speech.

We turn now to a consideration of the explanations that would have IF0 emerge as part and parcel of the normal process of vowel articulation. The main hypothesis is the so-called tongue-pull hypothesis, which actually comes in at least two different flavours. After looking at this we will then consider more briefly an acoustic coupling hypothesis (relationship between F0 and F1).

Useful reviews of the various models that have been proposed can be found in Dyhr (1990), Sapir (1989), Honda (2004).

2.3 Tongue-pull theories

Perhaps the earliest version of the tongue-pull hypothesis emphasized the vertical dimension (Ladefoged, 1964, p. 41). While there is a tendency for larynx height to correlate with F0, this is only when other things are equal. As Ohala & Eukel (1987, p. 209) put it, in vowel production other things are not equal. Thus there is a clear tendency for the high vowel [u] to have low larynx position, while the relative position of [i] and [a] is not so clear. So a simple relationship between larynx height and intrinsic pitch is clearly untenable (and later Ladefoged abandoned his original view (Ladefoged et al., 1972)). Nevertheless Ohala outlined a possible way in which vertical tension could still be involved, based on tension in the soft tissues directly, rather than being mediated by the vertical position of hard structures such as hyoid and thyroid. His proposal was that raising of the tongue dorsum could act via the aryepiglottic folds on the false vocal folds, thereby either increasing the vertical tension in the vocal folds, or - by increasing the size of the laryngeal ventricle - reducing any damping effect of the false vocal folds on vocal fold vibration. Additional light on this line of reasoning was contributed by Ewan (1979), firstly by pointing out that the typical low larynx position for [u] could (combined with the high tongue position) additionally contribute to increased vertical tension in the soft tissues. This could explain the fact that there is a fairly consistent tendency for higher F0 in [u] than [i]. Secondly, Ewan emphasized the lowering of F0 for low vowels rather than the raising for high vowels, but with a similar argument, namely that retraction of the tongue into the pharynx could lead to a compression of the ventricular folds, slackening of laryngeal tissues, and increased mass involved in vibration.

Ohala & Eukel backed up their version of the tongue-pull hypothesis by exploiting the classic bite-block speech paradigm: their reasoning was that if the jaw is fixed in an unusually open position, then speakers will be forced to make more use of the tongue for high vowels. This should enhance intrinsic pitch effects. At least for the largest bite-block used (10mm) this turned out to be the case (result based on 7 speakers of American English)[4]. Ohala & Eukel's experiment is of considerable relevance for our further discussion since there have been suggestions that take quite a different tack, namely that intrinsic pitch is more closely related to jaw height than to tongue height. As we will see shortly, this plays a particular role in results related specifically to German.

A further influential version of the tongue-pull hypothesis has been propagated in a number of publications by K. Honda. This might be regarded as a horizontal version of the hypothesis. The basic mechanism postulated is shown in the following figure. Posterior Genioglossus is known to be active for high vowels such as [i] and [u]. Upon contraction it can move the hyoid bone forwards, which in turn, through its linkage with the thyroid cartilage, leads to rotation at the

[4]It should be noted that in an earlier version of this experiment (although referred to by Ohala & Eukel as a replication) Lubker, McAllister & Lindblom(1977; "Vowel fundamental frequency and tongue height", JASA 62, S16-S17) failed to find any effect; this is put down by Ohala & Eukel to the fact that unlike them Lubker et al. based their analysis on unnormalized F0 data (as this was an ASA meeting abstract I have not checked this reference).

cricothyroid joint, and hence to lengthening of the vocal folds and an increase in pitch. In more recent work Honda has suggested that this basic mechanism may be supplemented by an effect of jaw opening in low vowels working in the opposite direction: opening the jaw may have the effect of moving the hyoid somewhat posteriorly.

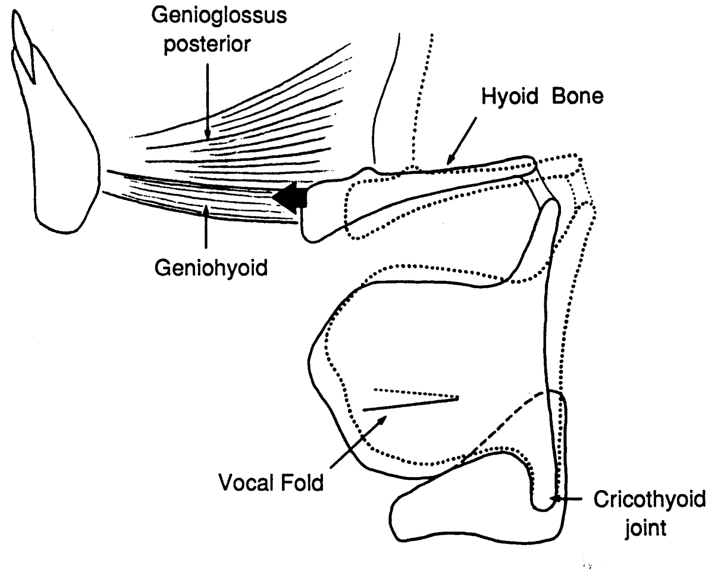


Fig 1. Supplementary tensor mechanism (Honda, 1986) which explains the biomechanical interaction between the tongue and the larynx. The geniohyoid muscle pulls the hyoid bone forward and causes the thyroid cartilage to rotate forward by pulling its superior horns, resulting in raising F0. Tongue gestures involve hyoid bone movement and thus activate this mechanism.

Fig. 2.1: Reproduction of Fig. 1 from Honda & Fujimura, 1991, illustrating mechanical link between genioglossus activity and lengthening of the vocal folds

The horizontal tongue-pull account sounds basically quite plausible. The genioglossus is the major extrinsic muscle of the tongue, so its strong contraction for high vowels may be expected to have widespread repercussions. However, it is not totally straightforward to document the horizontal movements of the hyoid expected by this account. Honda (1983a) shows some figures of hyoid movement (measured externally by means of a rod with LED applied to the hyoid) which point in the expected direction, but no scale is given in the figures, so it is difficult to estimate their magnitude. Rossi & Autesserre (1981) used xeroradiography to record four isolated vowels for four French speakers. They emphasize the mechanism of tongue root advancement for [i] and [u]; the movement of the hyoid is as expected from Honda's model, but the amount of movement is much smaller than of the tongue root itself, and the link between hyoid and thyroid movement also seemed to be somewhat tenuous. They see the soft tissue connections as the more relevant mechanism: advancement of the tongue root enlarges the laryngeal vestibule and stretches the membranous connections between epiglottis and arytenoids and between epiglottis and vocal folds[5].

A clinical study by Vilkman et al. (1989) also raised some problems for a central role of the hyoid. They found basically normal IF0 for vowels in a patient with an anatomical anomaly of

---

[5]These authors are also concerned to explain the higher intrinsic pitch of [u] which is somewhat peripheral to our concerns here. Here they assume an additional vertical tension mechanism, in which the larynx lowers more than the hyoid.

the hyoid-laryngeal region, such that there was only a very weak connection between hyoid and thyroid[6].

We will next look briefly at the acoustic coupling hypothesis, before considering in detail the evidence that active CT adjustments are the principal source of IF0. On the background of this plethora of conflicting hypotheses we will then be well-equipped to handle a situation in which German may add even further to the puzzles.

2.4 Acoustic coupling

The possible role of coupling between F1 and F0 is succinctly reviewed in Ohala & Eukel, and Dyhr. The original idea was that in high vowels F1 can become low enough to approach F0, and that F1 may then in effect dictate preferred frequencies of vocal fold vibration (F0 would then get pulled up towards the slightly higher F1; in low vowels F1 would be simply too high to cause any effect). As pointed out by Ohala, however, although it is quite easy to demonstrate source-tract coupling effects when phonating into a long tube, it is not so easy to model these effects in more normal vocal tracts. In fact some models would actually predict that F0 shifts down, when F1 is low (see reference to Guérin & Boë (1980) in Dyhr (1990)). As a more direct test, Ohala reports on the effects of helium speech; this shifts formant frequencies up very substantially, but F0 only slightly, so any coupling effects should be reduced; in fact intrinsic pitch effects seem to be largely unaffected. In addition, Ewan (1979) reported on intrinsic pitch in nasals. Here, too, the results did not appear consistent with a coupling hypothesis. Intrinsic pitch effects were found that could be simply related to the *articulation* of flanking vowels, whereas it was not clear how they could be explained by *resonance* conditions during the nasal itself.

2.5 EMG evidence for an active muscular contribution

We now consider to what extent available EMG investigations of the cricothyroid support the idea that IF0 is directly related to vowel-specific patterns of muscle activity. As already indicated above, this would be the most direct source of evidence that speakers use IF0 as a mechanism to enhance vowel height perception on the part of the listener (we will accordingly need to go briefly below (p. 14) into the separate evidence that there is an interaction between F0 and F1 in vowel perception). We will first look at a number of investigations that give some prima facie evidence for a contribution of CT activity to IF0, and then look at some more recent EMG evidence from Whalen et al. (1999) that failed to confirm this, and also look at the arguments of these latter authors that the evidence from the preceding investigations may be misleading.

---

[6]The authors also consider the patient's hyoid movements in speech to be anomalous, thus reinforcing the conclusion that hyoid-thyroid coupling is not crucial to normal IF0. However, this only applies to the vertical direction (hyoid appeared to go up for /**u**/ and down for /**a**/). In the possibly more important horizontal direction, the patient may not have been so anomalous: /**i**, **u**/ were further forward than /**a**/.

2.5.1 Honda & Fujimura (1991)

Honda & Fujimura (1991) present CT activity for the vowel productions of one speaker of American English. The data was actually recorded some years earlier, and the authors seem themselves to have been surprised how close the match was between IF0 and CT activity: They refer to it as “unexpected” (p. 150) after in the same paper showing the close relationship between intrinsic pitch and activity of the posterior genioglossus. We juxtapose the two corresponding figures below[7]. One might be tempted to think of a synergistic pattern of activation between GGP and CT. Given that Honda has been one of the most influential figures in elucidating the biomechanical substrate of IF0 it is interesting to quote their conclusion:

> *Our interpretation is that while the biological circumstances create a phonetic tendency for high vowels, for example, to be associated with high F0, such a tendency has to go through a process of phonologization to create a language-specific rule, in order to account for all aspects of observed characteristics associating intrinsic F0 with vowels in different languages. Once such a phonologization takes place, not only can the same F0 effect be used perceptually as one of the cues for vowel features, but also the same effect can be attained through other biological mechanisms that may be available. Thus the cricothyroid activity associated with high vowels “emulates” the biologically natural F0 rise due to hyoid bone movements. (p.151)*

We find the idea of the motor system exploiting a biological predisposition a very attractive one (“go with the flow” in colloquial terms; in a different context see discussion of forward movement of the tongue in velar closure in Hoole et al., 1998).

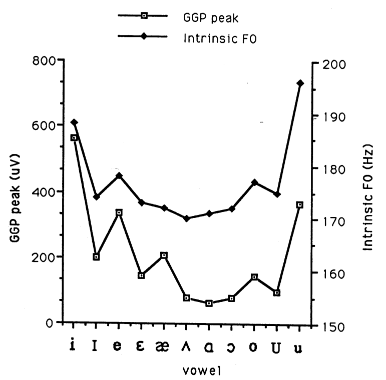


Fig. 2. The EMG activity of the posterior genioglossus muscle (GGP) during production of 11 English vowels in isolated /əpVp/ utterances (open squares) in comparison with the intrinsic vowel F0 (closed diamonds). "GGP peak" represents peak values of the ensemble average waveform of the EMG signals from GGP for ten repetitions, produced by one male speaker. "Intrinsic F0" is computed from the data by Lehiste and Peterson (1961), by averaging the F0 values of the vowels following /p/, /t/, and /k/.

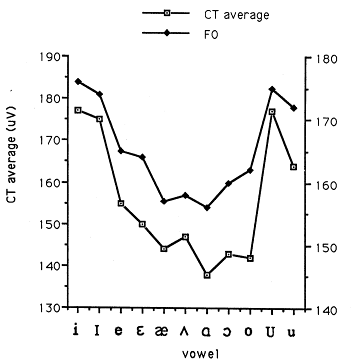


Fig. 3. The EMG activity of the cricothyroid muscle (CT) during productions of the vowels in isolated /əpVp/ utterances (open squares) in comparison with the measured vowel F0 (closed diamonds). "CT average" indicates the segment average values of the ensemble average EMG waveform from CT. The segment averages are computed using a 150-ms time window before the voice onset. "F0" represents the peak F0 values in the ensemble average waveform of vowel F0 in the utterances.

Fig. 2.2: Reproduction of Figs. 2 and 3 from Honda & Fujimura, 1991, showing vowel intrinsic pitch in parallel with genioglossus posterior (left) and cricothyroid (right)

[7] F0 values in the GGP figure are based on figures from Lehiste & Peterson (1961), whereas F0 in the CT figure is based on the actual utterances from the EMG experiment. If our reading of the the experimental description is correct, the tongue and larynx EMG were recorded in the same experimental session. A further experiment by Honda (1983b) with the same subject also found a quite close relationship between F0 and CT (for the more posterior of two CT insertions) over the vowels [i] and [a] (op. cit. Fig. 10-5)

2.5.2 Autesserre et al. (1987)

Autesserre et al. (1987) studied CT activity in one speaker of French and come to the tentative conclusion that intrinsic pitch could represent an active process. Whalen et al. question the relevance of this study because the subject was required to perform pitch melodies well beyond the normal range of speech; however, my reading of the paper is that these melodies were performed in a different part of the investigation, and that the intrinsic pitch results are based on a variety of stressed syllables from various natural intonation patterns of French. But it is true that the precise nature and amount of the speech material on which the result is based is not clear.

2.5.3 Vilkman et al. (1989)

Vilkman et al. (1989) in the clinical study already mentioned above also examined CT activity in one normal speaker of Finnish. In an analysis of variance CT activity was able to explain a substantial amount of the F0 changes over vowels (although the authors also feel that other factors, such as vertical tension in the larynx, probably have a role to play). They offer a novel interpretation for this vowel-related CT activity as a kind of compensatory effect:

> *... the increased cricothyroid muscle activity for high vowels occurs as a part of a complex motor pattern, in which the upward movement of the tongue is compensated by this muscle activity in order to avoid opening of the cricothyroid visor during increased vertical pull in the laryngeal region. This compensation results as a by-product to increased longitudinal tension of the vocal folds and a rise in F0. (p. 202).*

As Fischer-Jørgensen (1990) remarks, however, it is "*not obvious why it should be important to avoid this consequence*" (p.101), i.e the opening of the cricothyroid visor.

2.5.4 Dyhr, 1990

We now come to the investigation (Dyhr, 1990) with the most extensive range of normal speech material. Since Whalen et al. cast severe - but to my mind rather exaggerated - doubts on the interpretability of the results, detailed consideration is necessary.

This investigation analyzes recordings that were originally made in the 1970s in Copenhagen as part of an extensive EMG study of the larynx. Material was available from a total of 5 recording sessions. Four speakers were involved, one speaker being recorded twice, about three years apart. From the complete material (which was originally recorded for other purposes) items were selected allowing high (/**i**/, /**u**/) and low vowels (/**a**, **æ**/) to be compared in comparable contexts (all items were real Danish words spoken in a carrier phrase). Both short and long vowels were included in the corpus. In order to indicate how the results were presented, the figure below reproduces Fig. 1 from Dyhr's paper.

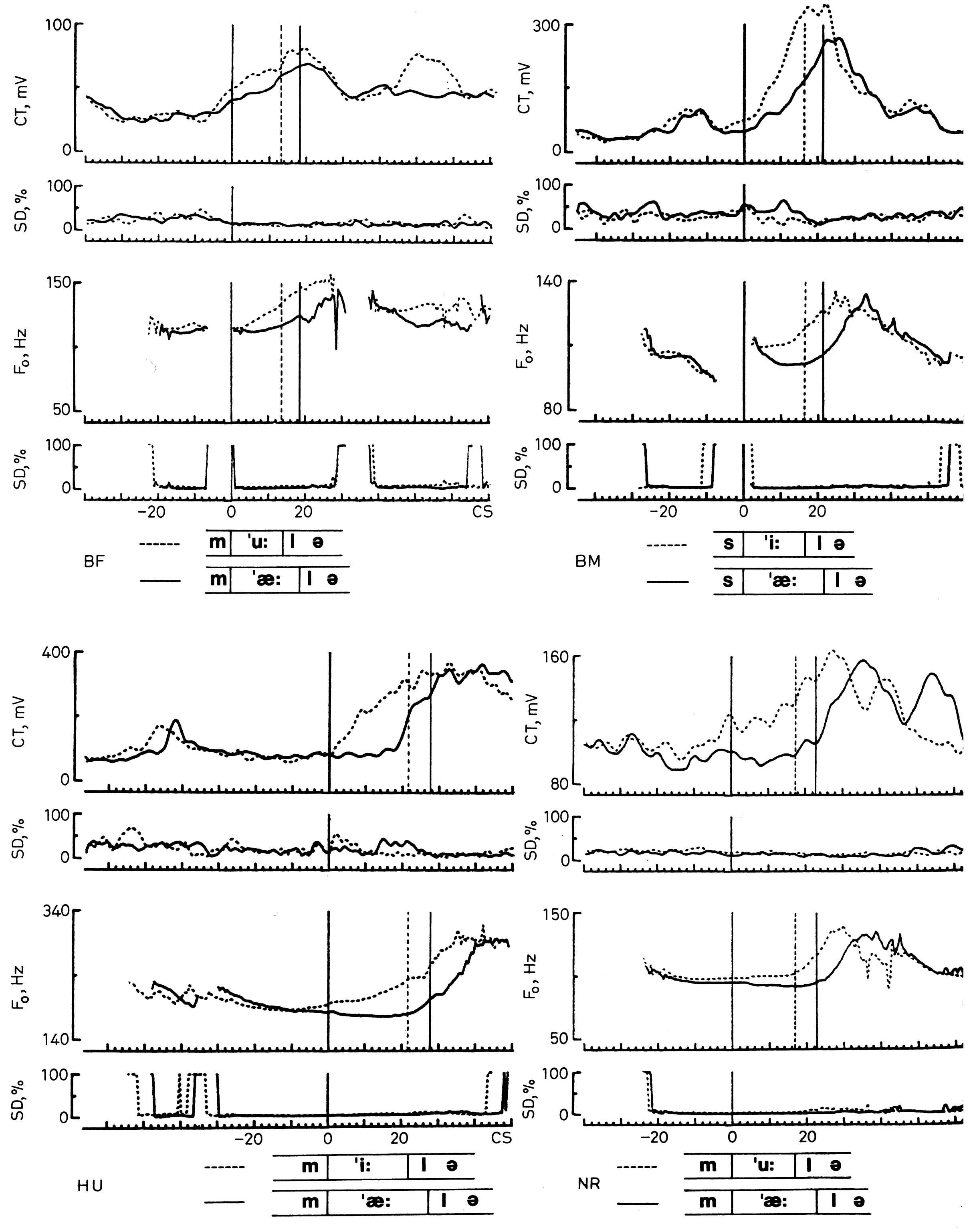


Fig. 2.3: Reproduction of Fig. 1 from Dyhr, 1990. Original legend: "*Comparison of superimposed average CT curves (top), F0 curves (bottom) and corresponding standard deviation curves. The curves are drawn with a broken line for high vowels and a solid line for low vowels. The line-up point (0) is the onset of the stressed vowel, the offset is marked with a broken line for high vowels and a solid line for low vowels.* [List of test sentences omitted here (P.H.)] *All examples show a stronger and earlier CT activity for high vowels than for low ones. Notice the difference in CT activity in low vowels; B.F. and B.M. have rising activity, while H.U. and N.R. have level or slightly falling activity*"

It will be seen that each of the four panels in the figure shows the results for one word pair for one speaker. Each curve is the ensemble average of (usually) 6 repetitions per word. In all the examples shown here CT activity is higher for the high vowel over pretty well the whole course of the vowel. Even though this is not the complete material, one might suspect that a fairly robust effect is emerging. Strangely, Whalen et al. criticize Dyhr for showing only a very small proportion of his data (and no statistics). They claim that only 16 individual vowels (in 8 pairs) of the total corpus of 306 vowels are shown. However, in fact 12 pairs are shown, and they seem to have overlooked the fact that these are ensemble averages, and not individual utterances (so roughly 12*2*6 =144 items contribute to the displayed results). Another way of looking at this is that 8 word-pairs for 1-4 speakers gave a total of 20 comparisons, of which 12 are shown. Thus quite a substantial proportion of the material is available, and the other two figures not reproduced here also give a very consistent picture of higher CT activity for the higher vowels. It is true - and unfortunate - that Dyhr does not give any statistical analysis of the results. But assuming the displayed cases are representative then even a primitive sign test counting the number of pairs with higher ensemble-averaged activity for the high vowels would give a significant result. At the level of individual pairs it is more difficult to estimate what differences could be significant. It will be seen in the figures that there is a curve labelled 'SD %' associated with each average curve. This presumably respresents the standard deviation over the (usually) 6 repetitions, but the text does not explain why it is given in percent, so it is not clear how to use it to estimate how reliably the average curves are separated at any given point in time. Whalen et al. also criticize the way in which the CT curves have been aligned with F0. Again there are some uncertainties in Dyhr's procedure but I think Whalen et al. are unduly pessimistic about any deleterious effect this might have on the results. Specifically, Dyhr took the estimated delay between CT activity and its effect on F0 into account in preparing the figures. The delay was estimated from time-lag between EMG peak and F0 peak in each utterance, and generally gave figures between 70 and 100ms. Whalen et al. have some understandable reservations about basing the time-lag on lining up peaks, and my reading of the paper is that Dyhr may have computed a separate time-lag for each utterance, which seems hard to justify physiologically. But nonetheless the typical figure he quotes has support in the literature, and I find it hard to believe that uncertainty in the alignment of maybe 20-30ms will impact very much on the results[8]

In short, I feel that while on the one hand Dyhr could probably have made a more watertight case for higher CT activity in high vowels than he actually did, on the other hand the results are by no means as uninterpretable as Whalen et al. would have.

2.5.5 Whalen et al., 1999

While the balance of the investigations reviewed hitherto is of higher activity in the high vowels we will see during the consideration of Whalen et al.'s own investigation that they still have one quite telling argument for urging caution in the interpretation of this apparently higher activity.

---

[8]A further possible minor quibble involves the chosen line-up point. This is referred to as vowel onset (of the target vowel), but probably means voice onset. It is not inconceivable that in aspirated consonants VOT could be longer for the high than for the low vowel, and this could have the effect of localizing the line-up effectively later in the rising intonation contour, giving a slight bias towards higher activity in the high vowels. But aspirated plosive contexts represent only 2 out of the 8 word pairs.

They investigated 4 speakers of American English with the explicit aim of testing the reliability of higher CT activity in high vowels, but used an unusual experimental procedure: the material did not consist of normal speech utterances at all, but rather of isolated vowels that had be spoken matched in pitch to a series of previously recorded target tones. This proves to have both advantages and disadvantages for the interpretability of the results. The target tones were chosen based on measurement of the intrinsic pitch interval (/**a**/ vs. /**i**,**u**/) for each subject in a pre-test, and ranged in steps of the intrinsic pitch interval from one interval below the speakers' normal /**a**/ pitch to 3 intervals above the normal i/u pitch. CT activity was assessed as the average activity over the 150ms preceding voice onset. There were essentially three strands to the results. Before considering them, a word on the subjects. Of the four subjects, one showed very atypical CT activity, with negative correlation between CT and F0. This subject happens to be one of those showing effects in the direction hypothesized by the authors, i.e no higher activity for the high vowels, but I think it would have been preferable to exclude this subject from further consideration (and she will not be mentioned in my summary below). For a further subject, two successful insertions were available (left and right sides), but the insertion on the left gave a rather weak signal. The results for this insertion are tabulated and displayed, but not included in the group statistics as being "near the lower limit of resolution" (p. 131). However, it is noticeable that this is the insertion with the clearest pattern *contrary* to the authors' hypothesis. The first strand to the results involved straight comparison using ANOVA of CT levels at matching points on the scale defined by the intrinsic pitch interval. This gave a mixed pattern of statistically significant results. The subject with two insertions had higher activity on the high vowels, one subject has lower activity on the high vowels, and one subject was mixed (/**i**/ higher, and /**u**/ lower relative to /**a**/). The more innovative part of their analysis becomes visible at the next stage of the results. Here an analysis of covariance was carried out, with F0 as covariate. Their reasoning, which I find very pertinent, is as follows:

> *If the CT levels are comparable across vowels, there should be no residual effect left over after F0 has been partialled out. While this assumption is not made explicit in any previous publications, it must at least be partially true for there to be any sense in comparing the levels across the different vowels. (p.132)*

If significant differences emerge after partialling out F0, then this indicates a possible vowel-dependent effect on F0 that is independent of CT. Specifically, higher vowels would be expected to have lower activity, i.e to be located below the overall regression line between F0 and CT. One subject had a clear effect of this type, for one subject the effect was significant though not very large, and for the subject with two insertions the results went in opposite directions. By way of illustration, we show a plot of the subject with the clearest results in terms of the authors' hypothesis. Unfortunately, it is necessary in one's mind's eye to flip the x and y axes to match the above discussion. However, one relevant effect can be directly gleaned from the figure as it stands: at any given level of CT, F0 tends to be higher for the high vowels. The third strand to the results, related, to a certain extent, to the previous one, can be also be derived from this figure. This is the question of whether the slope of the regression between CT and F0 differs over the different vowel categories. There turned out to be significant differences between the slopes for all subjects, consistently in the direction that slopes were shallower for /**a**/ than for the high vowels. The authors were able to derive corresponding estimates from the earlier work of Honda and of Vilkman et al, which basically confirmed these trends[9].

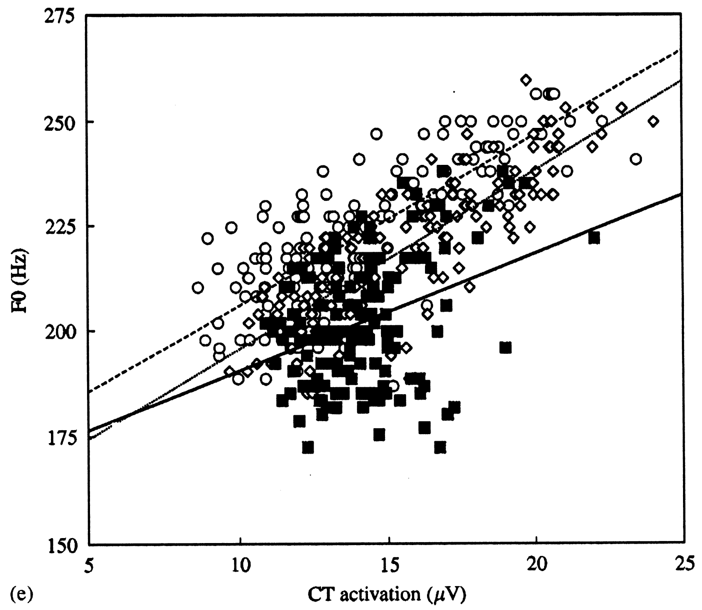


Fig. 2.4: Reproduction of Fig. 1e from Whalen et al., 1999. Relationship between CT and F0 for /**a**/ (filled squares, solid line), /**i**/ (diamonds, dotted line), /**u**/ (circles, dashed line)

What the last two strands of the results boil down to is that when results over the vowel categories do not fall on the same regression line, then it is difficult to compare CT levels across vowels (a given CT increment does not necessarily have the same effect on F0 across different vowels) and there may be influences on F0 that are dependent on vowel but independent of CT. Thus Whalen et al. introduce an important methodological refinement to the discussion (that we will make use of in a different context below) and provide support for the relevance of biomechanical effects in IF0. At the same time, I do not think they have succeeded in ruling completely out of court the possibility that speakers may ***also*** have an active CT contribution to IF0. As the authors themselves note (p.140), the tone-matching task may not result in completely typical F0 behaviour. In the light of our quote above from Honda & Fujimura, one may wonder whether utterances that are close to non-speech simply may not engage mechanisms that speakers have adopted as part of their repertoire for signaling phonological contrasts.

There is a further technical possibility that may explain part of the results. This study recorded CT from the pars recta, and there is some evidence that the closer relationship with F0 is found from the pars obliqua. This is consistent with results from the small amount of data where we were able to compare these two partitions of the muscle in our experiments, as discussed below in the final section of the results (p. 97ff; see also Honda, 2004, Fig. 4). In the work reported in Dyhr (1990), for example, insertions were taken from the pars obliqua. It is noticeable that many of Whalen et al.'s figures (not reproduced here) show a fairly weak relationship between CT and F0 (even leaving aside the speaker with abberrant negative correlations). Data for the high vowels of the single speaker shown here gave in fact the highest correlations found anywhere in the material.

---

[9]Curiously, in the paper abstract it is reported that "when F0 was shifted by an amount equivalent to that seen in IF0, it was found that the high vowels needed more CT activity to effect a change than the low vowels did", which seems to me to imply the shallower slope for the high vowels.

2.6 Vowel enhancement revisited

On the background of this review of a search for an explanation of IF0 we are in a position to now look specifically at whether German introduces a further puzzle.

It was not the central concern of our experiments to answer the question of active versus passive contributions to intrinsic pitch, however our experiments do supply some clearly much-needed additional data on a question with interesting implications for the extent of our understanding of the control of speech production.

Before doing so, it is worth providing a small amount of additional background on the enhancement hypothesis, for which vowel-specific CT-differences would be the most direct evidence. Essentially, this goes back to the demonstration by Traunmüller (e.g 1981) that vowel quality stays very similar over a wide range of F0 if the distance between F0 and F1 (in Bark) is kept constant.

Whalen et al. make, however, the simple point that we do not continually misperceive vowel height when the vowel is given an F0 prominence.

An alternative account, to which Whalen et al. would subscribe, regarding the relationship between IF0 and perception has been developed by Fowler (Fowler & Brown, 1997). This applies the concept of perceptual parsing to the area of pitch perception in a similar vein to which it had been applied to the perception of vowel quality in the seminal investigation of Fowler (1981). The essential finding was that if /**i**/ and /**a**/ are presented to listeners with the same F0, then /**i**/ sounds slightly lower in pitch. The interpretation was that listeners are able to parse F0 of vowels into a prosodic component actively controlled by the speaker and a component that is the automatic consequence of articulating different vowels. Put the other way round, if /**i**/ and /**a**/ have the same F0, then the speaker must have prosodically intended a higher pitch on /a/, and this is what the listener hears. While quite convincing evidence was found that this basic mechanism exists, not all details of the results could be completely explained. In particular, the magnitude of the perceptual effect was rather weak compared to the physical magnitude of IF0, at least for spoken as opposed to sung vowels. In other words, for spoken vowels the intrinsic pitch difference between /**i**/ and /**a**/ amounted to over 10 Hz but the amount by which /**i**/ could be higher than /**a**/ in F0 and sound the same in pitch was less than 2 Hz. Perceptual aspects of intrinsic pitch are considered further in the concluding discussion (Chapter 6).

2.7 German as a test-case: The problem of tense vs. lax vowels

On balance, the above discussion leads to the conclusion that tongue-pull, or tongue-root advancement, makes a significant contribution to IF0. This leads in turn to a puzzle with respect to German. The puzzle can be simply stated: tense-lax vowel pairs have very similar F0 but differ very clearly on tongue height. Fischer-Jørgensen (1990) was the first to highlight this problem and our debt to her comprehensive analysis is enormous. In order to set the scene we give some illustrations from our own earlier work on German vowel articulation.

The next figure shows for the vowels /**iː**, **ɪ**, and **eː**) average tongue position (acquired by EMA) averaged over seven speakers and 5 repetitions, and next to it an extract from our more recent work on tongue EMG, showing activity of posterior genioglossus for the same vowels (one speaker, averaged over 18 repetitions).

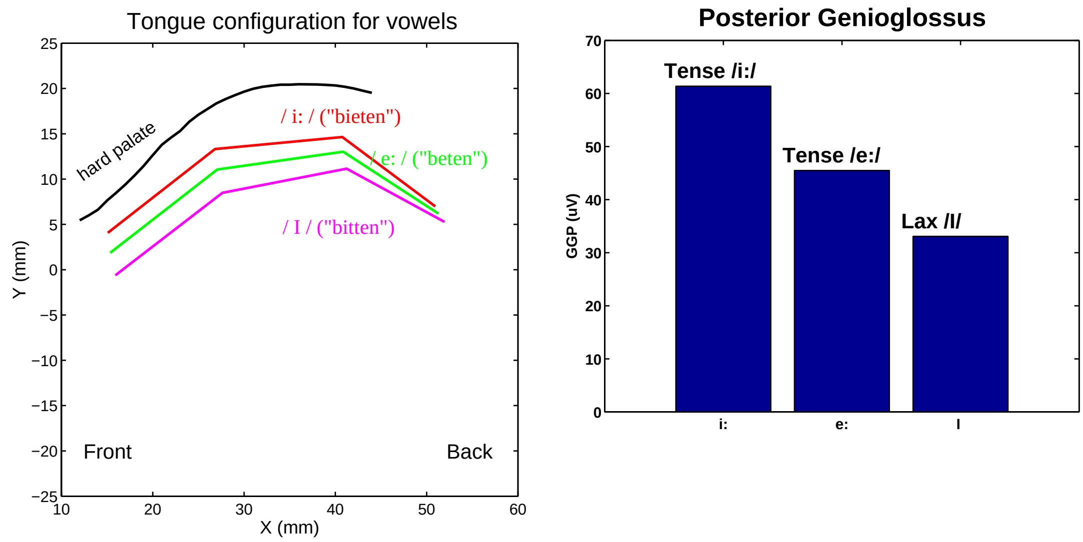


Fig. 2.5: Tongue position (left) and genioglossus posterior activity (right) for the 3 German vowels /**i**ː/, /**e**ː/ and /**ɪ**/

The substantially lower tongue position of /**ɪ**/ than /**e**ː/ is readily apparent, and the EMG data indicate concomitant differences in tongue-root advancement. Although these figures are merely illustrative, the relationships can be assumed to be robust (see Hoole & Mooshammer (2002), Hoole (1999) for more details of the EMA investigations). In fact, this basic finding has been known since Meyer, 1910, (who used a so-called plastographic method), and is also clearly apparent in the classic radiographic study of German by Chiba & Kajiyama (1941/1958), which formed part of Wood's (1982) extensive discussion of the tense-lax distinction. An articulatory characterization of the complete (except diphthongs) German vowel system is given in the next figure (based on the PARAFAC factor analysis presented in Hoole, 1999; same speakers as in left panel of previous figure).

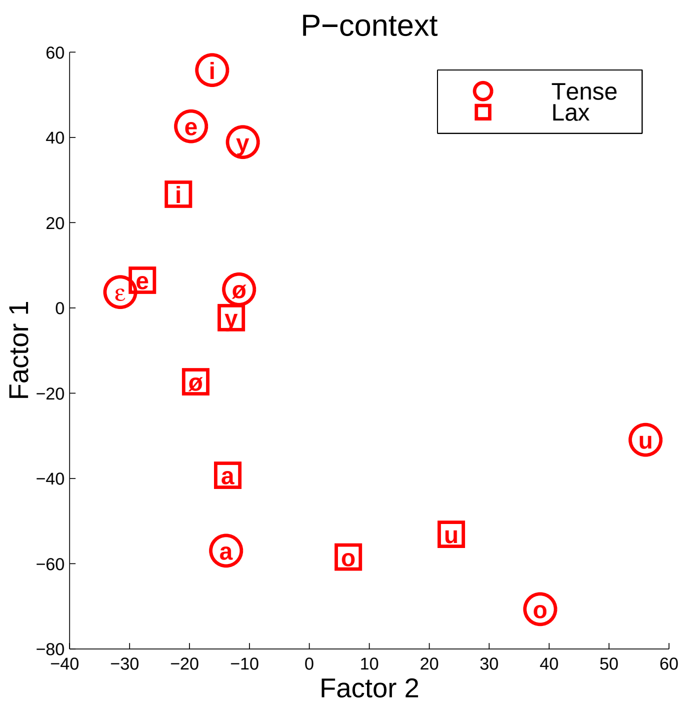


Fig. 2.6: An articulatory representation of the German vowel system based on PARAFAC factor analysis of tongue position in /**p**V**p**/ context for 7 speakers

What are typical findings for intrinsic pitch in German? The next figure shows F0 in each vowel averaged over the 6 male speakers who participated in the EMA experiment on which the above figures are based. Results are shown for two different speech rates. At the normal speech rate there are several cases where the lax vowel is actually slightly higher than the tense counterpart (e.g /**iː**/ vs. /**ɪ**/); at the fast rate the lax vowels are generally somewhat lower (perhaps a kind of undershoot of the F0 peak in the short vowels?) . However, the crucial observation is that there is absolutely no tendency for lax /**ɪ**/ to be lower than tense /**eː**/, lax /**ʏ**/ lower than tense /**øː**/, etc.

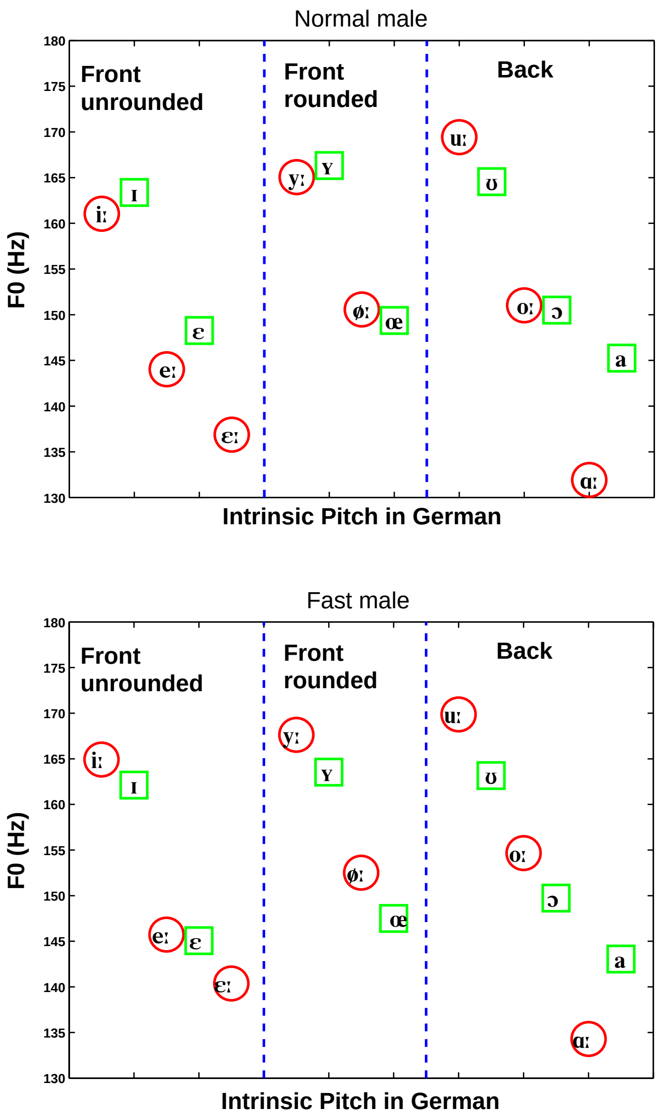


Fig. 2.7: Intrinsic pitch of tense vowels (red circles) and lax vowels (green squares) in German for 6 male speakers at two different speech rates. Within the groups 'front unrounded', 'front rounded' and 'back' vowels are arranged from left to right by decreasing vowel height, with tense-lax pairs adjacent to each other

Again, this may be regarded as a robust effect: Fischer-Jørgensen summarizes the German data available prior to 1990 (assembled in her Figs. 1-4). The lax vowels are sometimes slightly higher and sometimes slightly lower than the tense cognates, depending on vowel category and investigation, but ***never*** as low as would be predicted from tongue height.

Fischer-Jørgensen discusses a whole range of possible explanations for these findings. The one we would like to focus on here is whether IF0 is more closely related to jaw position than tongue position. Her own interest in this topic was reawakened by a finding along these lines made by Zawadzski & Gilbert (1989) for American English - reawakened because she reports first becoming alerted to the problem by data for English reported in Alfonso et al. (1982) covering tongue and larynx EMG together with formants and F0.[10]

She also makes some very apposite remarks (p.104) about why it makes more sense to look at this question in detail for German rather than English:

> *" ... German was chosen as a more appropriate language than English, because the often very pronounced diphthongization of English [e:] and [o:] complicates the comparison with [I] and [U]. These vowels have therefore often been left out in studies of intrinsic F0 in English, which is one of the reasons why the problem of tense and lax vowels in relation to intrinsic F0 has not been noticed."*

2.8 Is jaw position a relevant factor in IF0?

Fischer-Jørgensen investigated jaw height herself using video filming for five German speakers and found that there was indeed a close relationship between jaw height and F0 - in effect because tense-lax pairs tend to have very similar jaw position. The expected correlation between tongue-height and F0 is indeed found when the tense and lax vowel series are considered separately, but *not* when they are combined (this result can be suspected from the way we have arranged the intrinsic pitch data in the figure above).

She is, however, very careful not to posit a causal relationship (p. 125): *"It is, however, not easy to explain the correlation between jaw opening and F0. It is highly improbable that a smaller jaw opening can cause a higher F0 directly."* The need for caution - completely justified in our view - is borne out by our own analysis of the relationship between tongue-height, jaw-height, and F0. This is illustrated in the following set of three figures. Each figure compares the correlation coefficent for F0 vs. jaw-height and F0 vs. tongue-height, but groups the data in different ways.

The first figure (each panel corresponding to one subject) has two basic groups: front unrounded, consisting of /**iː**, **ɪ**, **eː**, **ɛ**/ (labelled with I), and front rounded consisting of /**yː**, **ʏ**, **ø**, **œ**/ (labelled with Y). Each panel of the figure includes a diagonal from (-1, -1) to (1,1). If data points fall below this line, it means that the correlation of F0 with jaw is stronger than that with tongue. By and large, this is the case, replicating Fischer-Jørgensen's basic finding. In the second figure (at the bottom of the page) the correlations are calculated over vowels grouped into pairs contrasting only in phonological height (e.g /**iː**, **eː**/). Here the correlations are overall stronger, and there is no clear preference for jaw or tongue to show the stronger correlations. This is to be expected as this simply replicates the traditional intrinsic pitch finding without the "interfering" effect of the tense-lax distinction. The more important point emerges from the third figure (top of page following the first two) in which the correlations are calculated over tense-lax pairs. Here the correlations are substantially weaker, and in many cases are even strongly negative. For the jaw vs. F0 correlations the reason for this is that in our data there was a slight tendency for the jaw to be lower in the lax vowels, while, as we have seen, cases can readily be found where F0 is higher in the lax vowels. Thus at a more fine-grained level of analysis the relationship between jaw-height and F0 breaks down, suggesting that a causal relationship is unlikely.

---

[10] As she notes, the F0 data was generally not shown in later more accessible publications, at least not until Honda & Fujimura (1991) - i.e after her own paper - if I have reconstructed the geneology of this dataset correctly..

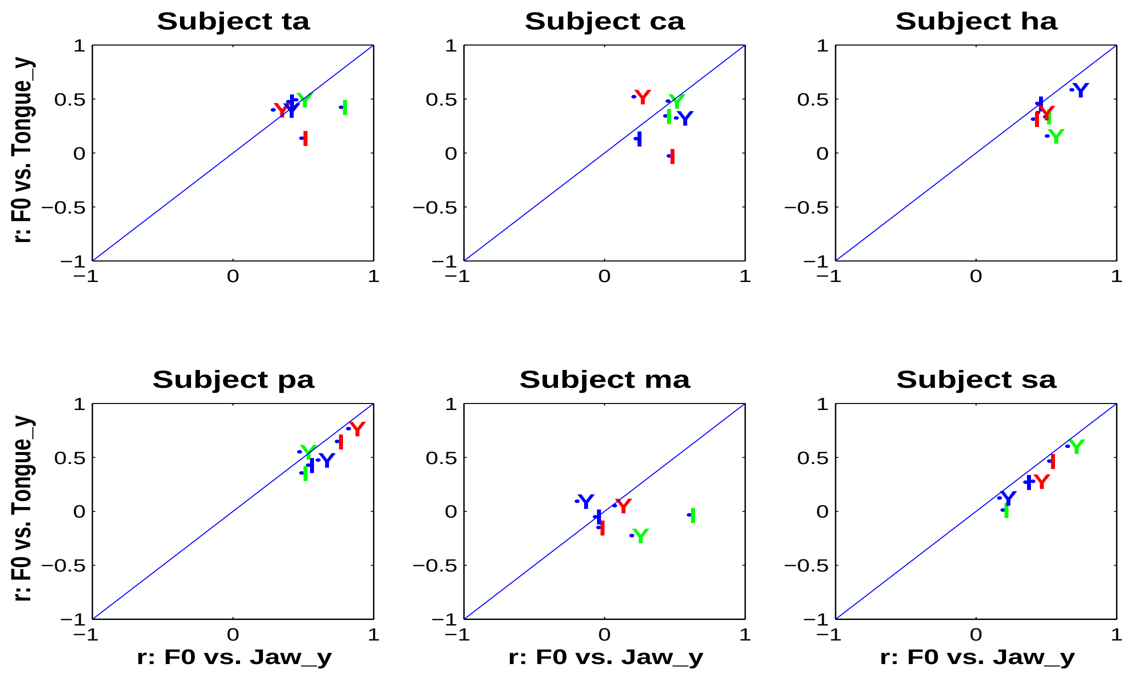


Fig. 2.8: Correlation between F0 and jaw height vs. correlation between F0 and tongue height over major front vowel groups. I = Front Unrounded, Y = Front Rounded. Consonant context p (red), t (green), k (blue)

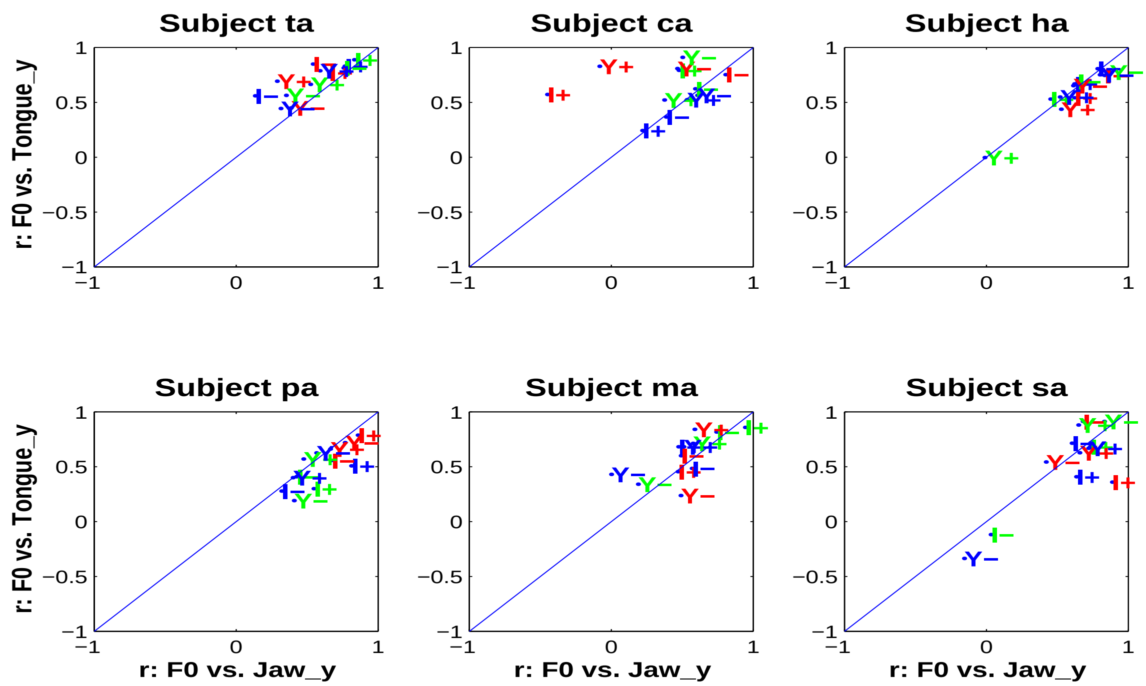


Fig. 2.9: Correlations as above. Vowels now grouped into pairs contrasting by height. I = / **i**, **e** /, Y = / **y**, **ø** /. Tense marked by ‘+’, Lax marked by ‘−’. Consonant context as above.

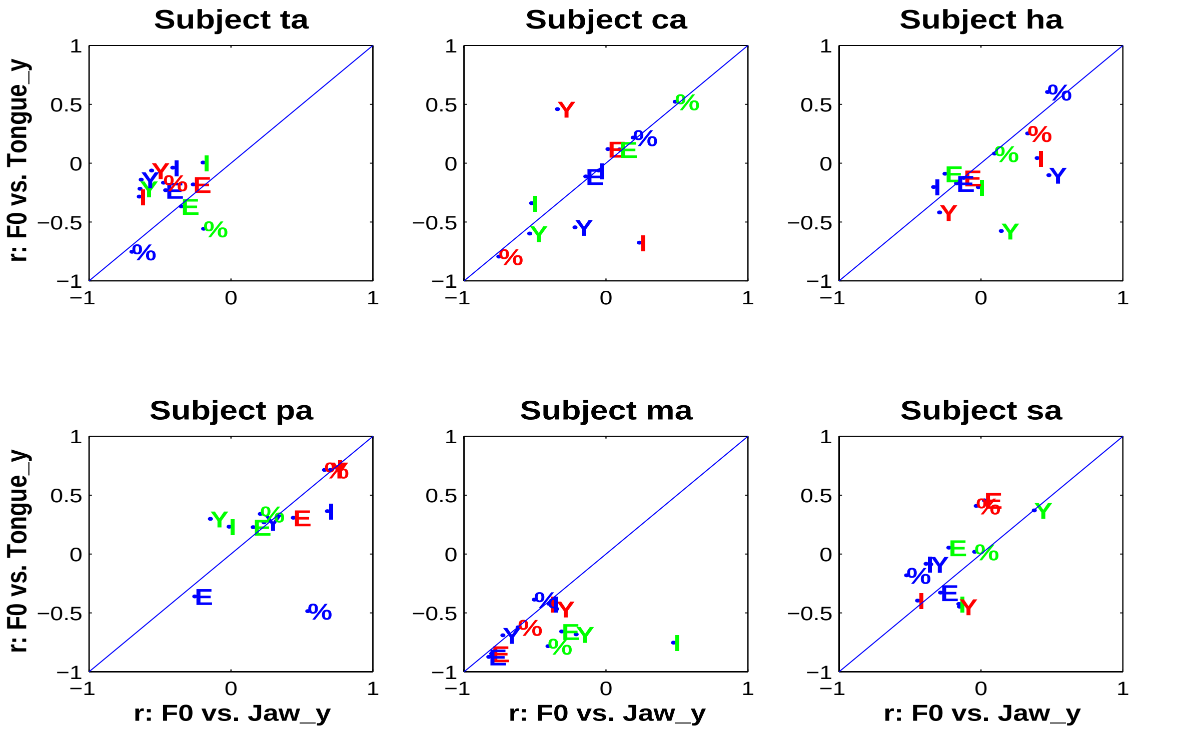


Fig. 2.10: Correlations as above. Vowels now grouped into tense-lax pairs. (% stands for the pair / **ø**ː, **œ** /). Consonant context as above.

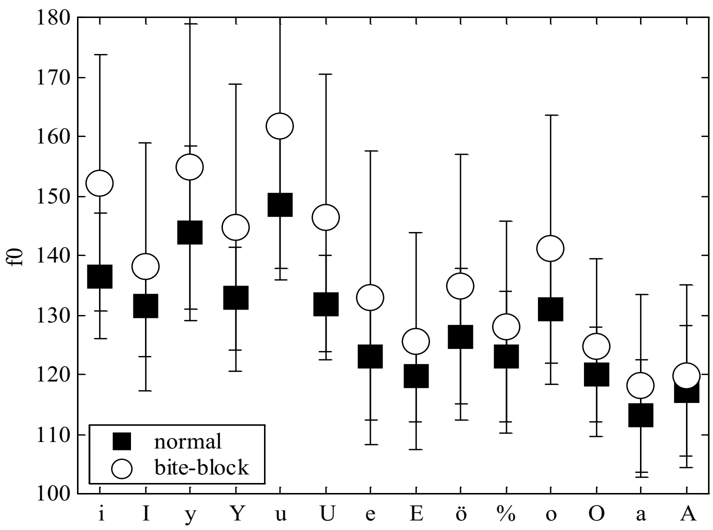


Fig. 2.11: Comparison of intrinsic pitch of German vowels in a bite-block and normal condition. Abscissa labelling: lower case for tense, upper case for lax ('%' stands for lax /**œ**/).

This interpretation is reinforced by a separate investigation in which the German vowels were spoken in a bite-block condition, thus in effect replicating the bite-block experiment of Ohala & Eukel presented above. The results, shown in the figure immediately above, clearly indicate that bite-block speech does not result in a levelling of F0[11]. In fact, (although we have not tested it

[11]These recordings were carried out at the phonetics lab of ZAS, Berlin, and presented in 2001 at an ASA meeting under the title: Mooshammer, C., Hoole, P., Alfonso, P. & Fuchs, S, "Intrinsic pitch in German: A puzzle?".

statistically) there is a tendency for the IF0 effects to be enhanced by bite-block speech, as found by Ohala & Eukel. We are not quite sure why this condition resulted overall in a rise in F0, but the high vowels appear to rise somewhat more than the low vowels.[12]

2.9 The necessity for EMG data on German

Fischer-Jørgensen's wide-ranging study make it clear that the puzzle of IF0 in German tense and lax vowels can most directly be resolved by investigating laryngeal EMG activity. This provided the motivation for the present investigation. In terms of the experimental outcome it is potentially a win-win situation: If no EMG differences are found between tense and lax vowels then at least we have gained the useful insight that there must be a missing element in biomechanical models of intrinsic pitch, especially those based on tongue-root advancement. If, however, differences are found between the two vowel classes then this means that biomechanical explanations remain valid, but are effectively overlaid by a prosodic difference between the two vowel classes.

We will not review here our earlier work on the articulatory substrate of the tense-lax opposition and its relation to phonological theories throughout the 20th century (Sievers (1901); Trubetzkoy (1938/1939; *Anschlusskorrelation*), Vennemann (1991, 2000; *Silbenschnitt*), except to mention the basic finding, which was that lax vowels show a tighter cohesion between the CV and VC movements (Hoole & Mooshammer (2002), Hoole et al. (1994), Kroos et al. (1997); further phonological background in Mooshammer (1998)), and is thus localized within the prosodic organization of syllables. In this sense, any further findings that are not narrowly segmental in nature would be of great interest for our understanding of the phonetic implementation of this phonological distinction.

If an EMG difference between tense and lax vowels were found, then the implications for perception would also be intriguing: Assuming e.g higher EMG activity in the lax vowels, but very similar F0 then it would be conceivable in the spirit of motor theory or direct perception that listeners ***hear*** a higher pitch in the lax vowels despite acoustic similarity. An alternative explanation in the line of enhancement theories would be that higher F0 on lax vowels helps to distinguish such pairs as (/**ɪ**/, /**eː**/) where F1 may be quite similar. A third possibility, entertained by Fischer-Jørgensen, is that speakers simply try to mark tense-lax pairs as belonging together by giving them similiar F0. Given the physiological emphasis of the experiments presented in this monograph a thorough treatment of these possibilities would be outside its scope. However, in the final chapter of general discussion we will look briefly at the results and implications of an ongoing series of experiments on the perception of intrinsic pitch being carried out by Pape et al. in Berlin.

---

[12]It will be observed in this corpus that the lax vowels are generally somewhat lower than the tense vowels, but - as in all other investigations - never as low as the next lower tense vowel. In fact, there can quite easily be a methodological bias towards higher F0 in the tense vowel: When the vowels form part of a pre-tonic rise, then often completion of the rise does not come until after the end of the lax vowel, but may fall within the time-course of the tense vowel. Thus, if the F0 contour of the vowel is reduced to a single figure by averaging over the whole vowel, then effectively the intonation contour may bias the long vowels towards a higher value. This effect will be relevant for interpretation of the results on the tense-lax distinction (p. 69ff; see there for further references).

3 Cricothyroid activity and the relationship between F0 and the control of consonant voicing

3.1 Introduction

In this section we review what evidence has accumulated to date that cricothyroid muscle activity is involved in the realization of the voicing distinction in consonants, and consider, in that case, what its precise role could be.

It is, of course, very well-known that the cricothyroid is the laryngeal muscle with probably the least ambiguous role in raising fundamental frequency (the details of fundamental frequency regulation involve a complex interplay between the cricothyroid, the thyroarytenoid, further intrinsic and extrinsic laryngeal muscles, and subglottal pressure (or the transglottal pressure difference), see e.g Atkinson (1978), Titze (1994), Honda (2004)). Thus consideration of cricothyroid involvement in consonant production might appear to involve pursuit of minor physiological details. There are, however, at least two reasons for further close consideration:

First of all, the voicing distinction itself is such a key linguistic opposition that full understanding of its physiological substrate is phonetically crucial. The two most obvious elements involved in the devoicing of consonants are the abduction of the vocal folds (the topic of part II of this work; see there for detailed background), and the increase of intraoral airpressure. However, K. Stevens has also emphasized the importance of regulating the tension in the vocal folds when controlling the conditions under which vocal fold vibration can occur. Increasing the tension will help to suppress vibration. Halle & Stevens (1971) is the frequently quoted source at the origin of this idea, and we will consider it in more detail at the end of this literature review. In any case, increased tension could plausibly involve the cricothyroid, and leads to the second reason for the considerable linguistic ramifications of this subject.

This second reason is the well-known fact that the fundamental frequency tends to be higher following voiceless consonants than following voiced ones. Following the seminal work of Lisker and Abramson (1964) on Voice Onset Time , attention soon turned to the cue value of pitch differences in the perception of the voicing distinction. Abramson & Lisker (1985) tended to play down its significance, but see also Silverman (1986), and for further contexts Kohler (1985) for postvocalic stops, and Schiefer (1986) for Hindi. The linguistic impact of this phenomenon is even more striking when one considers that fundamental frequency differences related to the voicing status of the consonant are generally considered to play a key role in tonogenesis (including tone split), high and low tones developing out of voiceless and voiced contexts, respectively (and it is worth recalling that most humans alive today speak a tone language). Moreover, in a synchronic perspective, too, the voicing status of the consonant has long been known to play a huge role in tonal rules for extant tone languages (e.g Schuh (1978), Ansre (1961)).

Thus even though additional physiological information (to be discussed below) has since become available, a very good source for initial consideration of possible reasons behind the influence of voicing status of the consonant on adjacent vowels is the book “Tone, a linguistic survey” (Fromkin (ed.), 1978), particularly the chapters by Hombert (“Consonant types, vowel quality, and tone”) and Ohala (“Production of tone”).

3.2 Voiceless consonants and raised F0: Early accounts

Here I cannot resist re-quoting the somewhat facetious quote from a paper by Matisoff (1973) on tonogenesis in southeast Asia with which Hombert kicks off his chapter:

> *And Change said “Let the consonants guarding the vowel to the left and the right contribute some of their phonetic features to the vowel in the name of selfless intersegmental love, even if the consonants thereby be themselves diminished and lose some of their own substance. For their decay or loss will be the sacrifice through which Tone will be brought into the world, that linguists in some future time may rejoice” (Hombert p.77, from Matisoff, p. 73).*

The reasons considered by Hombert and Ohala in 1978 for higher F0 following voiceless consonants can be considered in terms of the following three hypotheses.

3.2.1 Airflow hypothesis

One of the first hypotheses appealed to the effect of differences in airflow following the different consonant categories. This hypothesis was at first sight attractive because F0 of vowels seems to be much more clearly affected by preceding rather than following consonants. The supposition was that the high rate of airflow following voiceless aspirated consonants in particular would increase the Bernouilli effect in the closing phase and lead to an overall faster glottal cycle (conversely perhaps that the low rate of airflow through the glottis in voiced consonants would lead to a lowering of F0). In practice, it seems, however that the temporal extent of any airflow differences is too short to explain the often very long duration of F0 differences in the vowel. The temporal extent of F0 differences in the vowel is an issue to which we will return when it comes to interpreting our own results: Can differences plausibly be attributed to differences in muscle activity, and can they be seen as a perturbation directly related to the requirements of consonant articulation, or do speakers actively enhance on the vowel itself effects originally related to the consonant?

A problem with the aerodynamic hypothesis was that, according to measurements and modelling of Ohala, subglottal pressure - which can certainly have an effect on F0 - may actually be lower at onset of voicing following an aspirated stop, precisely because of the immediately preceding high rate of airflow through the vocal tract during the aspiration phase (this ties in neatly with our results for Icelandic given in the Appendix: intraoral pressure was lower in the closure phase of ***pre***-aspirated stops).

3.2.2 Vertical laryngeal tension hypothesis

The second possible explanation relates to the possibility of different vertical larynx position for voiceless vs. voiced consonants. There is some evidence that the larynx tends to be higher following voiceless consonants, and some evidence that vertical position affects vocal fold tension (higher position leading to higher tension; cf. also discussion of explanations for intrinsic pitch in Chapter 2). A potential problem with this approach is that one would normally assume that the active adjustment is to lower the larynx to support voicing in voiced consonants by slowing the rise of intraoral pressure. This would predict that F0 is lowered in voiced consonants versus sonorants and voiceless consonants, whereas in fact voiced consonants tend to pattern with sonorants in terms of F0, and the active consonantal perturbation appears more likely to be a raising of F0 on voiceless consonants rather than a lowering on voiced. Our own experimental data do not allow us to address this hypothesis explicitly (but see Hoole & Kroos, 1998), yet it remains worth bearing in mind for the discussion in later sections.

3.2.3 Horizontal laryngeal tension hypothesis

The third explanation considered by Hombert and Ohala was the horizontal tension hypothesis put forward in the Halle & Stevens paper just mentioned, and which Ohala referred to at that time as “interesting and innovative”. The main drawback at that time was simply that supportive EMG results were not available. Hombert quotes papers by Hirose, Lisker & Abramson (1973) and Hirose & Gay (1972) as providing no evidence for differences in laryngeal tension related to the voiced-voiceless distinction. It will be the task of the next section to consider to what extent the balance of evidence has shifted since then. A further apparent drawback of the horizontal tension hypothesis is that it does not immediately explain why consonants tend to affect following vowels more than preceding ones. Again, this is an issue to which we will return when considering the timing of EMG activity. One advantage of the hypothesis over the aerodynamic hypothesis relates to aspirated and unaspirated voiceless consonants. These appear to have fairly similar effects on F0 and do not appear to have played different roles in tonogenesis (see also Ohde, 1984). They obviously share the necessity for a devoicing gesture, but differ markedly in their aerodynamic properties.

3.3 A specific contribution of the cricothyroid: More recent evidence

We now consider those studies having appeared since 1978 that look specifically at cricothyroid activity in consonants.

3.3.1 Collier et al. (1979)

The first paper, Collier et al. (1979), examined several laryngeal muscles for a single speaker of Dutch, the speech material consisting of voiced and voiceless consonants and fricatives, as well as combinations of two consonants. Regarding CT, the authors simply note that it “appeared not to be relevant to the consonant distinctions under investigation” and show no further results for it. This paper does give a considerable amount of information on vocalis activity, however. This is the other principal element involved in regulating horizontal tension in the vocal folds, and generally appears to *reduce* its activity in voiceless consonants. Further discussion of this will be deferred to later, but this discussion is necessary since Collier et al. interpret their results as speaking against the Halle/Stevens hypothesis. In view of the fact that the investigations to be considered next generally did find voicing-related differences in CT activity, it is worth spending a moment to consider whether there is any explanation for the negative finding in Collier et al., other than the traditional but unhelpful cop-out in physiological experiments with single subjects of appealing to inevitable subject differences. A possible explanation could be that in this investigation the target consonants were in post-stress position (medial consonant in pseudowords with stress on the first syllable), and thus in a phonetically weak position. It would have been interesting to know whether in this weak position the expected differences in fundamental frequency were also absent, but this is not reported by the authors.

3.3.2 Dixit & MacNeilage (1980)

The next study to be considered is also a single-subject study, this time on Hindi, which is a particularly interesting language in this context (Dixit & MacNeilage, 1980). The interest resides of course in the fact that Hindi contrasts voiceless unaspirated, voiceless aspirated, voiced unaspirated and voiced aspirated consonants. These contrasts were recorded for various places of articulation in pre-stress position. The basic finding was that the voiceless consonants had higher levels of CT activity than the voiced consonants, and the CT results matched the accompanying differences in F0 quite well. The voiced aspirated consonant appeared to have more CT suppression than the voiced unaspirated consonant, and also had the lowest F0. The result for the voiced aspirate is particularly illuminating since it indicates that increased CT activity is not some kind of automatic accompaniment of glottal abduction; at the same time it provides support for the idea that regulation of horizontal tension can be important in consonant articulation: maintaining voicing while the glottis is abducted is aerodynamically a somewhat delicate matter (and in fact it is well-known that the aspiration phase of these sounds is not always continually voiced), and if indeed horizontal tension is a relevant factor then it makes perfect sense that it should be kept at a particularly low level in these sounds.

(Schiefer (1986) has shown that the particularly low F0 in voiced aspirates may be important in perception in distinguishing these sounds not only from the voiceless sounds but even from the voiced unaspirated cognate.)

The following figures show the results presented graphically by Dixit & MacNeilage for EMG and F0, namely for the bilabial stops and postalveolar affricates.

The EMG results are by and large clearer for the affricates, perhaps because these sounds are overall somewhat longer. A point to which we will be returning repeatedly is the timing of the EMG activity, and the delay between electrical activity and its mechanical effect. If we assume that in this case the EMG should be shifted to the right about one subdivision (40ms) on the displayed time axes, then the main differences between the consonants are clearly located squarely on the consonantal closure phase.

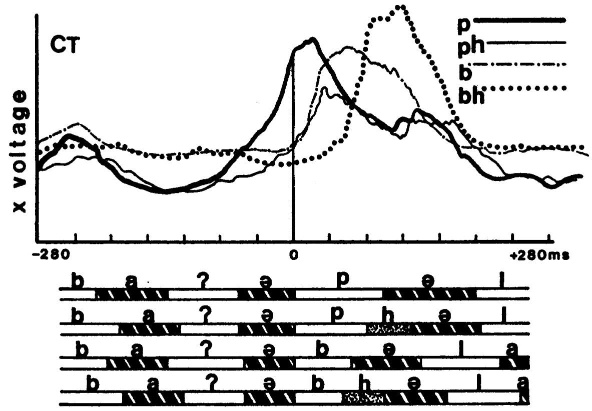


Fig. 1. Averaged EMG records of cricothyroid (CT) activity during /əpəl/, /əphəl/, /əbəl/, and /əbhəl/. The records are lined up at the point of articulatory closure for the stops indicated by zero on the abscissa.

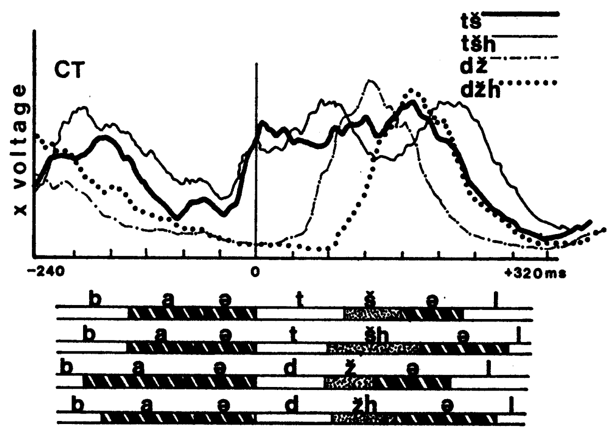


Fig. 2. Averaged EMG records of cricothyroid (CT) activity during /ətšəl/, /ətšhəl/, /ədžəl/, and /ədžhəl/. The records are lined up at the point of articulatory closure for the affricates indicated by zero on the abscissa.

Fig. 3.1: Reproduction of Fig. 1 and 2 from Dixit & MacNeilage (1980) showing CT activity in plosives (top) and affricates (bottom), for voiceless unaspirated, voiceless aspirated, voiced, and voiced aspirated.

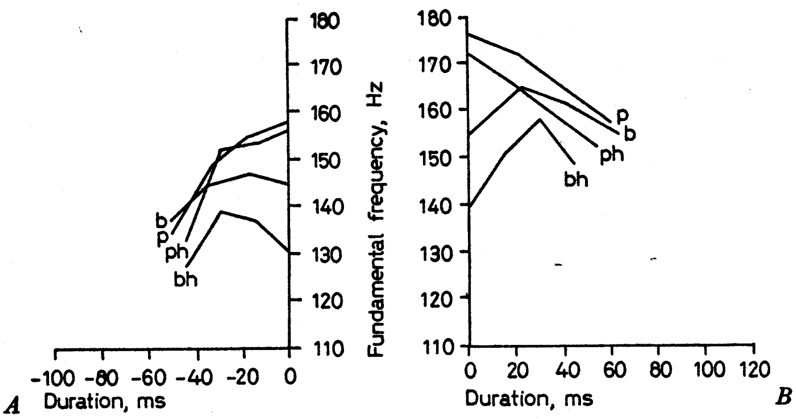


Fig. 3. Averaged F_0 values of vowels preceding and following stops in the same words as in figure 1. The zero on the abscissa in *A* represents the point of articulatory closure, whereas in *B* it represents the point of vowel onset with respect to articulatory release, which occurred later in real time in voiceless and voiced aspirated stops.

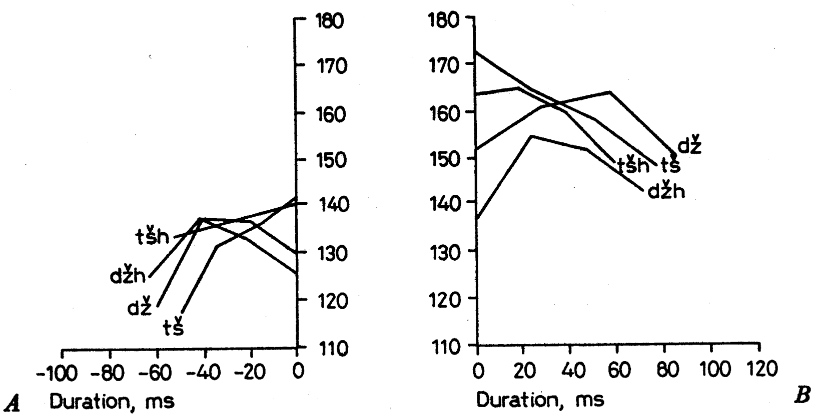


Fig. 4. Averaged F_0 values of the vowels preceding and following the affricates in the same words as in figure 2. The zero on the abscissa in *A* represents the point of articulatory closure, whereas in *B* it represents the point of vowel onset with respect to articulatory release, which occurred later in real time in voiceless and voiced aspirated affricates.

Fig. 3.2: Reproduction of Fig. 3 and 4 from Dixit & MacNeilage (1980) showing F0 for the consonant categories shown in the previous figure.

3.3.3 Hutters (1985)

In 1985 Hutters published an extensive investigation of laryngeal behaviour in Danish stops, based on both transillumination and fiberoptics as well as EMG. A total of 9 speakers was involved, but not all measurements were carried out on all speakers. For CT 2 two speakers provided usable signals, but unfortunately only results for one of them are shown in the paper. It is not quite clear whether the comments made on the signals are intended to apply equally to both speakers. The results are reproduced below.

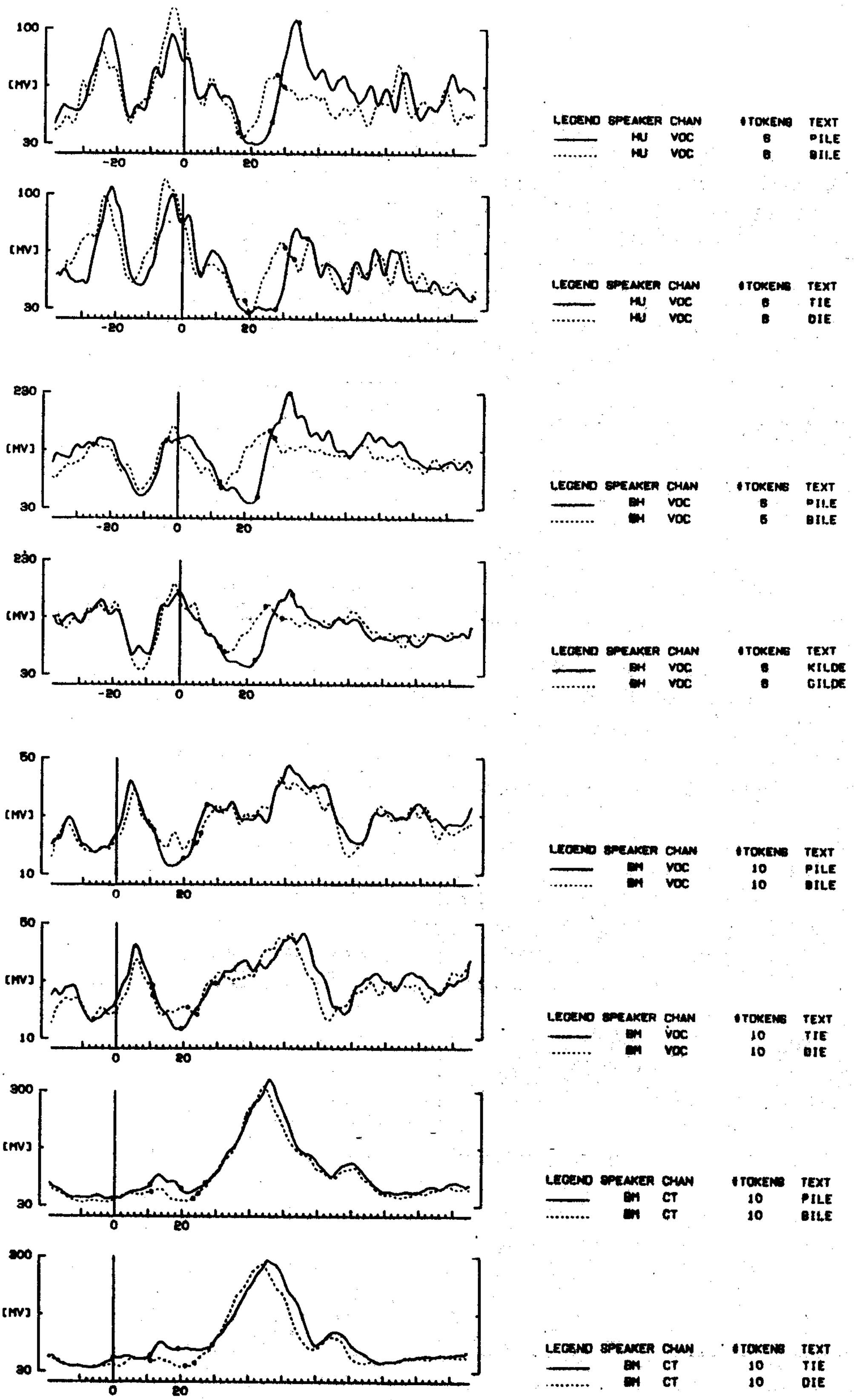


Figure 4 (cont.)

Fig. 3.3: Reproduction of Fig. 4 from Hutters, 1985, showing CT and VOC activity in selected voice-voiceless pairs. Top 6 panels VOV (total of 3 speakers); bottom 2 panels CT (1 speaker)

For both voiced-voiceless pairs investigated, somewhat higher CT activity is found in the voiceless cognate. Relative to the very strong rise in CT activity to raise F0 for the following accented vowel, the difference is quite subtle, and Hutters admits that her data is fairly limited. Differences are observable roughly from the time of onset of consonantal closure up to a time corresponding to voice onset after the aspirated plosives (again, the question of a time-shift of the EMG is not raised in this paper, so in terms of effect the EMG should probably be shifted slightly to the right). Hutters makes a good methodological point regarding comparison of F0 following aspirated versus voiced (or unaspirated) consonants: It is apparent from the figures that the timing of the strong CT peak for the following vowel is very similar for voiced and voiceless consonants . Thus if the test syllables are spoken on a rising intonation contour (i.e with overall increasing CT) then it is almost inevitable that aspirated consonants will have higher F0, if F0 traces are aligned relative to voice onset (voice onset for the aspirated case is in effect later in the intonation contour). It is often not clear in the literature whether this potential confound has been taken into account. It is thus methodologically preferable to compare F0 contours relative to some independent articulatory landmark such as consonantal release.

Hutters' results for vocalis, for which more speakers were available, are also of some interest. Four speakers were recorded, and results for three of them are shown in the paper and in the figure reproduced here. During the consonantal closure phase there is a consistent (and expected) effect that vocalis activity is suppressed, with the suppression generally stronger for the aspirated consonant. The interesting point is that the vocalis activity increases particularly strongly at the consonant-vowel transition for the aspirated consonants so that it is actually at a higher level than for the unaspirated consonants around the time of voice onset (admittedly, for speaker BM this is not too clear, especially for p vs. b). Since the timing here refers to unshifted EMG, it would seem that the higher vocalis activity could extend quite substantially into the vowel, whereas CT differences between voiced and voiceless - unfortunately only available for BM - may disappear earlier.

3.3.4 Löfqvist et al. (1989)

We now turn to the paper that has now become probably the most quoted one on this question. Löfqvist et al. (1989) investigated cricothyroid activity in two speakers of American English, and one speaker of Dutch. The speech material consisted of the voiced and voiceless plosives and fricatives of these languages (plus affricates for English). The consonants were uttered in CV syllables (V=/**i**, **a**, **u**/) in a reiterant speech phrase modelled on "The man went to market". Sentence stress fell on "man"; the consonant in the corresponding reiterant speech syllable was the one that was analyzed (thus as in Hutters (1985) a pre-stress position of the consonant).

The main findings were quite consistent over the different speakers, and are reproduced in the following figures: there was higher CT-activity for the voiceless consonants over a period starting roughly where the sound amplitude started to decrease at the transition from vowel to consonant, and extending up to the release of the consonant. The maximum difference between curves is generally located at about the point where full occlusion is achieved. This is the first paper of which I am aware where the differences were also tested statistically; there were highly significant differences between voiced and voiceless consonants in all cases except the affricates of speaker TB (where also little difference is observable in the ensemble averages). Interestingly, this was precisely the case where F0 differences in the post-consonantal vowel were least salient.

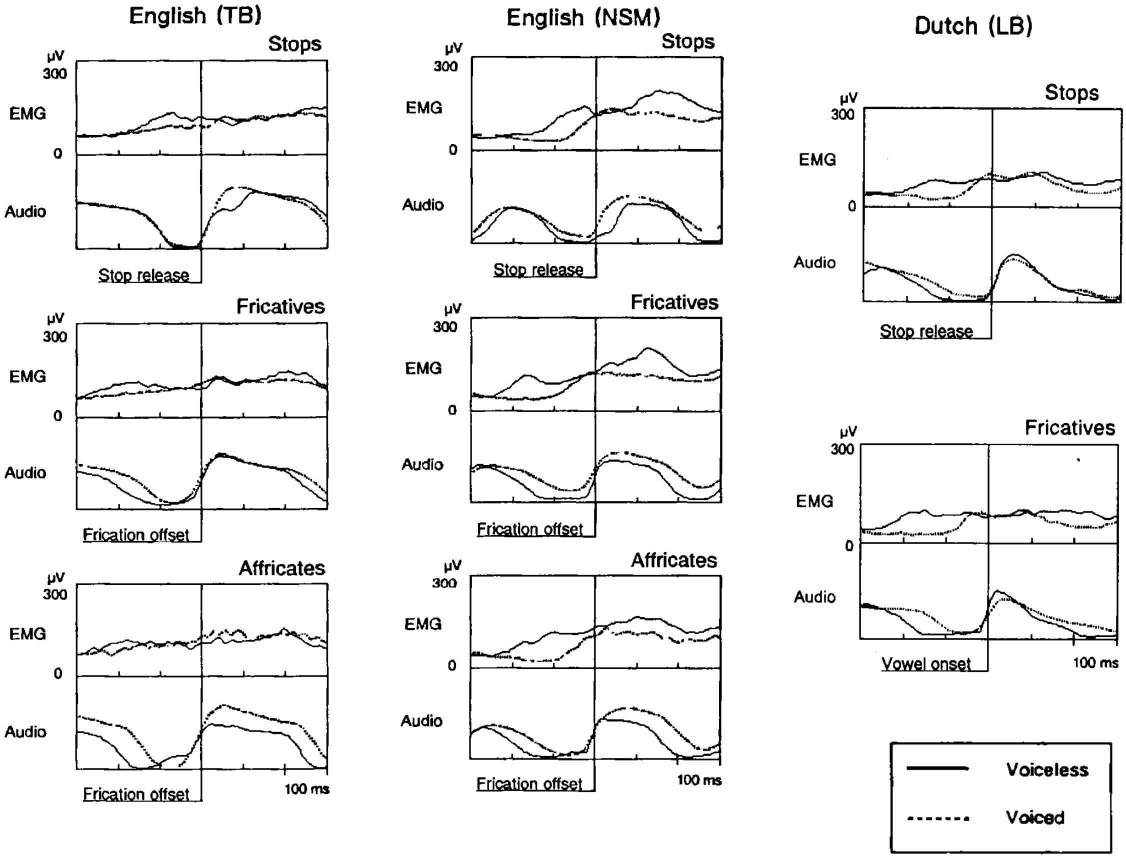


Fig. 3.4: Slightly rearranged reproduction of Figs 1, 2 and 3 from Löfqvist et al., 1989, showing averaged CT activity and audio signals in voiced and voiceless stops, fricatives and affricates for two speakers of English and one speaker of Dutch (TB: n=47 for stops and fricatives, n=15 for affricates; NSM: n=89 for stops, n=120 for fricatives, n=28 for affricates; LB: n=45 for stops, n=29 for fricatives)

In their discussion Löfqvist et al. consider whether the purpose of the higher cricothyroid activity is indeed to support devoicing , or whether it is conceivable that the speakers' primary intention is control of F0 in the following vowel. Taking the timing of the activity into account, they find the former view more plausible. If F0 differences in the vowel were the primary aim, then the strongest EMG differences are probably located too early. This remains true even if one takes the delay between EMG and effect on F0 into account. Assuming a figure of about 50ms, this would amount to shifting the EMG curves half a subdivision to the right in the plots. This would locate the portion of the curves where difference between voiced and voiceless is minimal around the amplitude maximum of the vowel, so the authors' reasoning appears quite plausible (the fact that the EMG curves - after meeting around the time of consonantal release - then separate again can be attributed to the effect of the following consonant). The authors feel that the higher F0 following voiceless consonants is still explained by their results, but the reasoning behind this is less clear. The F0 differences in the vowel are - as in many other investigations - extremely robust: statistically highly significant effects extending up to at least the 10th pitch period following voice onset. While the authors do find a reasonable positive correlation between CT activity and F0 at voice onset, and while it is revealing that in the case of TB's affricates the weakest F0 effects were linked to the weakest EMG effects, it is still not clear what the mechanism is that causes F0 effects to persist in time to a point where CT differences appear to be small. One physiologically relevant point raised by them, but not discussed in detail, is that relaxation times for CT may be longer than contraction times. This is qualitatively the kind of effect that is required for a full account of the data; whether the quantitative details are appropriate is less clear. In any case, this serves to make clear that interpretation of findings in this area can hinge rather finely on details of timing and on assumptions about delays between electrical activity in the muscle and effects on acoustic output.

The results for the Dutch subject, both for EMG and F0, were substantially the same as for the two American subjects. Since the Dutch voiceless plosives are unaspirated, this reinforces previous findings that aspiration is not a crucial variable in this connection. (No explanation is offered by the authors for the difference between these findings and those of Collier et al.)

3.4 A recent note on F0 and the perception of consonant voicing: Whalen et al. (1993)

Before we conclude this section with the physical reasoning behind the Halle/Stevens suggestion of stiff vocal folds in voiceless consonant production we will give brief consideration to one of the more recent papers to examine the contribution of F0 to perception of the voicing distinction (Whalen et al., 1993). Even though perception is not at the focus of our attention it turns out to provide a convenient summary of some of the issues that have been raised as background to our own experimental results.

While earlier experiments showed that the perception of stimuli that are ambiguous with respect to VOT can be affected by F0 contour, the tendency was to play down the importance of F0 in perception since in most natural speech VOT values are not so ambiguous, and in such cases the influence on listeners judgements, or their behaviour in normal speech situations, was assumed to be negligible (e.g Abramson & Lisker, 1985, quoted above). In the Löfqvist et al. paper just discussed this was used as an argument for the primary association of CT activity with devoicing rather than F0 control. The present study of Whalen et al. relativized this view by introducing a paradigm whereby voicing decisions had to be made under time pressure, and where reaction times were measured. This allowed the authors to show that even unambiguous VOTs, when associated with inappropriate F0 contours, can exhibit an increase in processing time. In other words, the F0 information is clearly reaching the listeners' speech perception system, and not simply being automatically discarded as phonologically redundant, and irrelevant. As the authors point out, this provides interesting support for the traditional tonogenesis argument. It is difficult to understand how the consonantal voicing distinction can lead to tone if the concomitant F0 perturbations are simply not impinging on listeners' perceptual systems.

The second point made by the authors related to tonogenesis will be quoted in full, since it emphasizes that our knowledge of CT activity in consonants is still rather rudimentary, and raises the question of enhancement of physiologically triggered effects - these points being an important part of the motivation for our own experiments:

> *A second assumption is that the F0 effect of voicing must be enhanced before it can begin to be used distinctively. Otherwise, the loss of the voicing distinction would, of necessity, mean the loss of the F0 difference. This, of course, assumes that the configuration for the presence versus absence of voicing are (sic) directly responsible for the F0 perturbations. While such a view seems to hold true at*

> *present (Löfqvist et al., 1989), the number of languages that has been examined to date is too small to reach any firm conclusion about whether the relationship is a necessary one and/or how widespread enhancement of the perturbation might be. (Whalen et al., 1993, p. 2158)*

3.5 Controlling the conditions under which vocal fold vibration can occur: The contribution of vocal fold tension

Since features such as +/- Tense Vocal Folds have obtained wide currency in the phonetics and phonological literature it is worth explicitly reproducing a figure from the seminal article of Halle & Stevens (1971) showing the interplay between transglottal pressure difference, glottal width and vocal-fold stiffness in determining the conditions under which vocal fold vibration can occur.

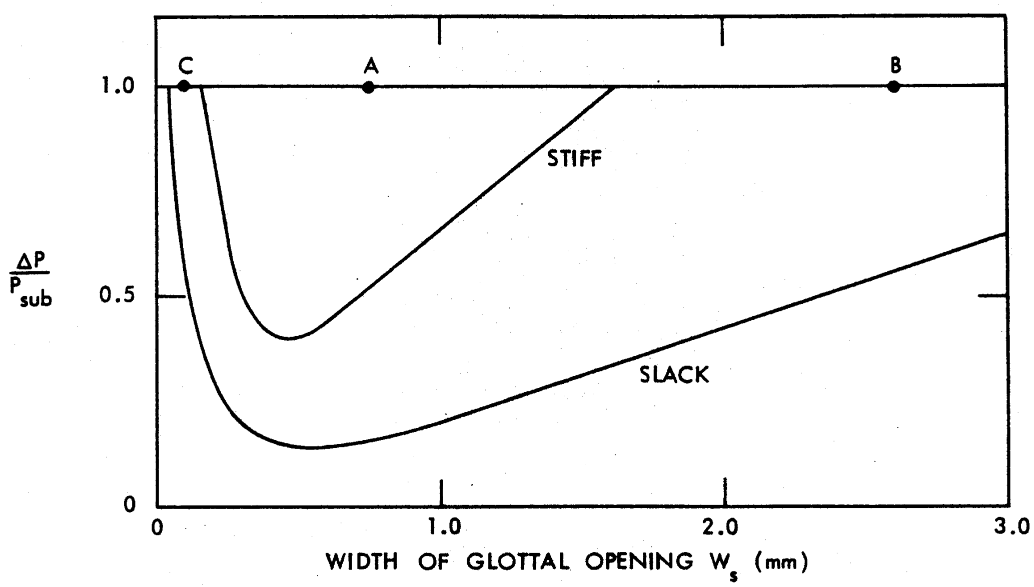


Fig. 3.5: Fig. IX-15 of Halle & Stevens (1971). Original legend: *"Sketch showing approximate ranges of conditions under which vocal-cord vibration occurs. ΔP is the pressure across the glottis, P_{sub} is the subglottal pressure, and w_s is the static width that would be assumed by the glottis if there were no vibrations. If the values of ΔP and w_s give rise to a point above the curve labeled "slack," then vocal-cord vibration is initiated when the vocal-cord stiffness is small. Likewise the curve labeled "stiff" represents the boundary of vocal-cord oscillation for relatively stiff vocal cords. Below these lines, the vocal cords remain in a static position with no oscillations. Points A, B, and C represent glottal widths that lie within regions of "normal" glottal vibration, spread glottis and constricted glottis for nonobstruents, for which $\Delta P = P_{sub}$. The portion of the chart corresponding to obstruent configurations is well below the line $\Delta P / P_{sub} = 1$. The regions are based on an assumed subglottal pressure of ~8cm H_2O. The shapes of the curves for $w_s > 0.5$ mm are derived from theoretical analysis of a two-mass model of the vocal cords (Ishizaka and Matsudaira (1968) and Stevens). For smaller values of w_s the curves are estimated."*

Additional background can be found, for example, in Stevens (1977). Briefly, increased stiffness can be seen as reducing the energy transferred from the aerodynamic to the mechanical system. If this transfer becomes too small, losses in the mechanical system are not compensated for, and vibration is no longer possible (op.cit., especially Fig. 4 (based in turn on Baer (1975), and discussion on pp.269-271).

We will reserve more detailed discussion of this area for the concluding chapter, in particular in the light of a more recent paper by Hanson & Stevens (2002) which raises the possibility that increased stiffness is not so much being used to ensure suppression of voicing for a voiceless consonant, but rather to regulate when voicing re-starts after it.

In terms of rounding off the present section the points to make are that the physiological evidence that speakers actually use this mechanism is still rather slight, but that if they do, the the cricothyroid is the place to look.

The other muscle clearly involved in regulating the biomechanical state of the vocal folds is the thyroarytenoid (the interplay between cricothyroid and thyroarytenoid is an extremely complex matter that is discussed at length by Titze, 1994; see also Lowell & Story, 2006), however we have seen above that there is fairly consistent evidence for suppression of the thyroarytenoid in voiceless consonant production (perhaps leaving aside special cases like the fortis stops of Korean). This makes sense because it is well-known that thyroarytenoid activation incorporates an adductory component, whereas the essential task for the speaker in voiceless fricatives and plosives (the latter especially if clearly aspirated) is to abduct the vocal folds. There is no evidence that the cricothyroid would in any way impede glottal abduction. Thus, if suppression of voicing requires not only abductory movement but also increased vocal fold tensio, then the cricothyroid must be the candidate of choice.

4 Experimental Procedures

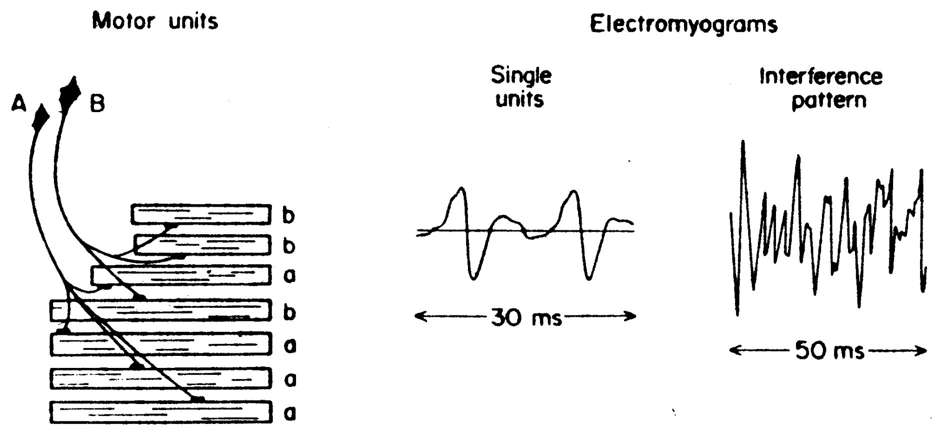
The seven sections of this chapter will cover the following areas:

(1) General background to EMG
(2) Specific background to the electrodes used
(3) Anatomic background for insertions into the cricothyroid muscle (CT)
(4) Other hardware components in the experimental setup
(5) For each subject, details of the insertions carried out, and the speech material recorded. Also notes on handling of artefacts in EMG signals
(6) Processing of EMG signals, including discussion of time-alignment with the audio speech signal
(7) A brief outline of the procedure used to calculate fundamental frequency (F0)

4.1 The principle of electromyography

In this section we will outline, with appropriate background, the EMG procedures on which the present experiments were based.

First of all, what does EMG measure? It records the muscle action potentials arising in muscle fibres when a muscle is activated. What, in turn, is a muscle action potential? Here it is convenient to introduce the term “motor unit”. This has been referred to by MacNeilage as the “basic neurophysiological unit of movement control” (1973, p.56). It consists of a motoneuron together with the muscle fibres innervated by it (refer to the left part of the following figure from MacNeilage, 1973). The cell bodies of such motoneurons (capital “A” and “B” in the figure) are located (for the speech muscles) in the cranial nerve nucleii in the brain-stem. When a motoneuron fires, its action potential propagates along its axon to the neuromuscular junction where it triggers similar action potentials in the nerve fibres to which it is connected (thus motoneuron “A” is connected to all muscle fibres “a” in the figure). The sum of the action potentials in all fibres of a motor unit is known as the muscle action potential. Thus although muscle action potentials are not electrically identical to the action potential in the nerves (transmission at the neuromuscular junction also involves chemical processes) the attraction of the electromyogram for analyzing motor control processes resides in the fact that there is a one-to-one relationship between the MAP and the activity of the corresponding neuron in the central nervous system.



Schematic illustration of; left, motor units; center, the muscle action potential of a single motor unit; right, an interference pattern.

Fig. 4.1: Reproduction of Fig. 1 from MacNeilage, 1973

There is, of course, a great deal more that could be said about motor unit physiology (for example, the significance of the number of muscle fibres per neuron, order of recruitment of motor units, etc.). However, since this would not impinge directly on the interpretation of our data, the reader is referred to standard physiology and motor control texts for further details (e.g Brooks (1986), Rosenbaum (1991), and the volume edited by Brooks in the Handbook of Physiology (Geiger, 1981)). The remaining point that should be made here is that with the kind of experimental setup used in our experiments the electrical signal we record is usually referred to as an "interference pattern", defined by MacNeilage as the "summed electrical effect of many single motor units firing simultaneously" (compare the voltage patterns in the right and middle panels of the figure; note that there is a difference in the time axis for interference pattern vs. single unit). While it can be interesting for basic physiological studies to isolate the activity of single motor units (usually with a paradigm involving subject training[13]; see MacNeilage et al., 1979, Basmajian, 1967, McClean & Clay, 1995) this would actually be a disadvantage in our case since we are interested in overall level of activity in the muscle. Typical firing rates of motor units for speech could be about 20/s (though estimating this is not straightforward, see MacNeilage, 1973), which is rather infrequent as a basis for a smooth estimate of muscular activity.

4.2 EMG electrodes and insertion techniques

4.2.1 Background to use of hooked-wire electrodes

Clearly, the simplest electrodes to use are surface electrodes. While miniaturized surface electrodes have proved extremely useful for analysis of speech muscles such as orbicularis oris they have some equally obvious disadvantages. Apart from the problem of being unsuitable for

[13]also easier with needle, rather than hooked-wire electrodes (see below), as electrode position is easier to vary

poorly accessible muscles, there is the major problem of lack of selectivity of which muscle is being recorded from. This is always a problem when the muscles are arranged in layers, as is frequently the case with the speech muscles. Thus many situations call for the use of intramuscular electrodes. For basic physiological studies concentric needle electrodes are popular (concentric in order to allow bipolar connection) but are often too much of a hindrance or too uncomfortable for the subjects when the aim is to analyze movement. For analysis of the speech musculature the development of hooked-wire electrodes was thus a major breakthrough. The basic design idea can be traced back to Basmajian & Stecko (1962). To my knowledge, the first major speech article in which their use is described is from Hirano & Ohala (1969). The basic construction principle can be observed in the following illustration: two fine insulated wires are threaded through a hypodermic needle, bent over at the ends to form hooks, and then the insulation is removed under the microscope for a distance of about 0.5 to 0.7mm. The resulting two bare wires act as the two parts of a bipolar electrode. The purpose of the staggered arrangement of the hooks is to reduce the chance of short-circuit between the two leads. The length of the hook is varied slightly in practice, depending on the thickness of the muscle to be investigated.

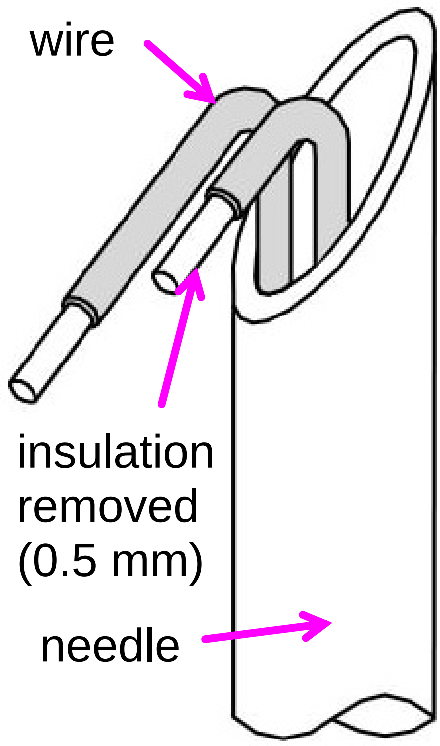


Fig. 4.2: Design of hooked-wire electrodes (sketch kindly provided by Kiyoshi Honda)

Fig. 4.3: Use of needle to insert hooked-wire electrodes into a laryngeal muscle (neck of Christian Kroos, hands of Kiyoshi Honda)

The needle is used to insert the wires into the muscle (see photo). Because of the hook arrangement the wires remain in the muscle when the needle is withdrawn. After this has been done the other ends of the wires can be connected to the amplifier input. The great advantage of the hooked-wire procedure is that once the needle has been withdrawn discomfort for the subject and impediment to articulation is negligible, at least for the laryngeal and lingual muscles we have examined to date (this may admittedly not be true for less accessible laryngeal muscles such as posterior cricoarytenoid that require a peroral approach). In any case, subjects generally reported finding the procedure less demanding than laryngeal fiberscopy, for example. At the end of the experiment the hooked-wire electrodes can be removed painlessly with a slight tug. The introduction of the technique led to a spate of EMG-experiments in the speech field; in particular, the fruitful collaboration between the Research Institute of Logopedics and Phoniatrics, Univ. Tokyo, and Haskins Labs., New Haven, resulted in many advances in our understanding of muscle activity in speech (for an early review of techniques for the main speech muscles see Hirose, 1971).

4.2.2 Electrode specifications for the current experiments

The technical details for our electrodes were as follows:

Wire: Platinum(90%)-Iridium(10%) alloy. Teflon coating. Diameter 0.002" bare and 0.0045" with coating. Manufacturer: A-M Systems.

Needle: Diameter 25G

To decrease electrode impedance and capacitance the electrodes were treated electrolytically by immersing an electrode pair in salt water and connecting to a 9V battery for 30 sec, then repeating after reversing polarity.

The insertion site on the neck was treated with topical anaesthetic for 30min. prior to insertion (Lidocaine tape).

4.3 Anatomical background

We will now outline the anatomical background relevant to electrode insertion into the cricothyroid muscle. Choosing the correct angle and depth of insertion requires considerable expertise on the part of the investigator. A basic view of CT anatomy is shown in the following figure.

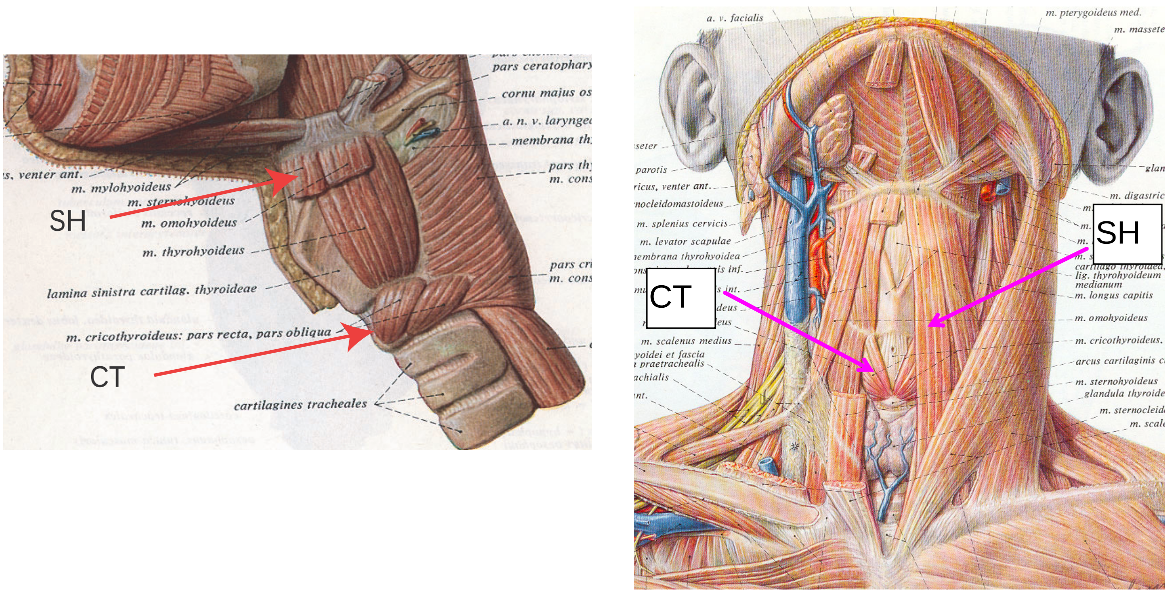


Fig. 4.4: Basic anatomy of cricothyroid (CT) and sternohyoid (SH; cut away in left panel). From Sobotta, Atlas der Anatomie des Menschen, 18th ed., 1982 (left panel: excerpt from Fig. 308; right panel: Fig. 475)

The cricothyoid is generally considered to be divided into two compartments, the pars recta and the pars obliqua (the more anterior and posterior part, respectively; clearly visible in the left panel). The possible functional significance of this subdivision is discussed in a short section of the results at the end of Chapter 5 (p. 97ff) in which those of our results that were relevant to this issue are also presented. In our recordings we generally aimed to obtain an insertion into the pars obliqua, and in the cases outlined below a further insertion into pars recta was attempted. Since it is in practice not possible to be sure whether the insertion is fully in the intended part, we will refer to insertions as posterior vs. anterior when a distinction needs to be made.

Insertion towards the pars obliqua can be illustrated by the left part of the next figure, which reproduces Figs. 3 and 4 from Hirano & Ohala (1969).

Insertion is made just above the cricoid ring about 1cm from the midline, then directed posterolaterally and slightly upwards aiming at the lower edge of the thyroid lamina posterior to the inferior tuberculum. The depth of the insertion is of the order of 2-2.5cm (less depth required if aiming for more anterior location in the muscle). The right-hand panel of the anatomical illustrations from Sobotta above shows the main difficulty for cricothyroid insertion: at the normal insertion site the cricothyroid is covered by a more superficial muscle, namely the sternohyoid (in the left panel the sternohyoid has been cut away). It can thus very easily occur that recording will be made from this muscle if the insertion is not deep enough. This is illustrated more schematically in the sketch in the right part of the figure from Hirano & Ohala. This also shows the counterpart: if the insertion is too deep, recording may be made from the lateral cricoarytenoid.

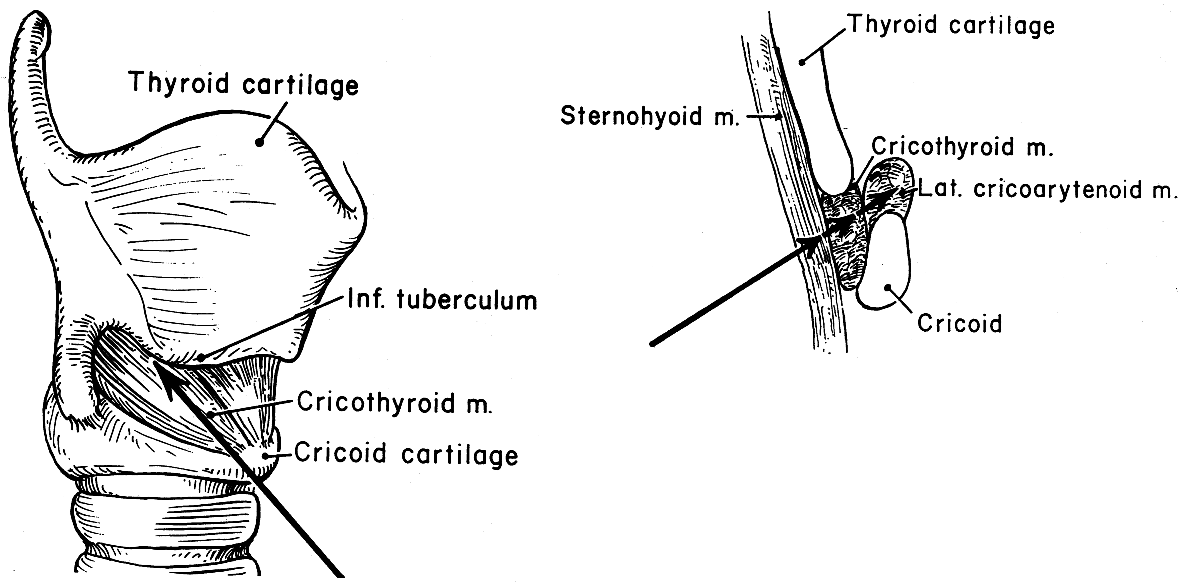


Fig. 4.5: Reproduction of Fig. 3 (left) and Fig. 4 (right) from Hirano & Ohala, 1969, showing insertion into pars obliqua of cricothyroid (left panel), and, in the right panel, the relationship between insertion depth and muscular layer along the line of the arrow in the left panel.

In addition to completely missing the intended muscle, a frequent problem is obtaining a signal not unambiguously related to just one muscle, e.g one of the two hooks may be very close to or in a neighbouring muscle resulting in a contaminated signal (or one hook may not be in muscle tissue at all, giving a weak signal). In fact, the sternohyoid is also an interesting muscle, since it is generally considered to have two rather different functions: firstly, it has often been implicated in lowering of F0 (thus the opposite of CT), and secondly it may play a role in jaw opening, probably by stabilizing the hyoid to support the action of e.g the anterior belly of the digastric (Atkinson & Erickson, 1977; Erickson et al., 1995). Accordingly, in order to have a counterpart to CT we aimed in these experiments to always also have one good SH insertion (this succeeded, but the data has not yet been analyzed in detail). In practice, this was often possible without an additional burden on the subject, since if a good SH insertion was obtained while looking for a CT signal, then this was retained and a further search for CT was made.

It will be readily understandable from the above discussion that tasks for verification of insertion position have always played an important part in EMG experiments. For our experiments the following tasks were used. They were repeated at regular intervals throughout the experiment because there is always the possibility of a shift in electrode position:

CT: active for pitch rise; no activity for jaw opening against load; no activity for glottal closure (the last task to distinguish from lateral cricoarytenoid)

SH: active at lowest part of F0 range; active for opening jaw against load.

4.4 Basic hardware setup for the present experiments

We now summarize further details of the experimental setup that were common to all recordings. Final details of the insertions are given individually for each subject below.

EMG amplifier: 8 Channel Grass Model 15LT system with Model 15A54 amplifier models

Amplifier setting: High pass filter 30Hz, Low pass filter 3kHz, 50 Hz notch filter.

Amplifier gains: Typically 10000 (EMG signals are in the range of about 100uV)

EMG and microphone signals were recorded on a multichannel DAT instrumentation recorder (Sony PC208) at a samplerate of 24kHz.

To monitor signal quality during insertions and during the course of the experiment the data could be monitored oscillographically in realtime using the computer interface of the DAT recorder. EMG channels could also be monitored individually via loudspeakers or headphones. Since EMG signals lie comfortably within the audio range, audio control is usually the most convenient souce of feedback to the investigator during the insertion procedure.

All recordings reported here were carried out in the sensorimotor lab of the Max-Planck-Institut für Psychologische Forschung, Munich. Two pilot experiments with speaker CK had previously taken place at ATR labs, Kyoto, Japan. Insertions were carried out either by Kiyoshi Honda or Emi Murano.

4.5 Subjects: Insertions and speech material

It is convenient to present details of the insertions performed on each subject in conjunction with an outline of the speech material recorded. There are differences between subjects in both respects, and the stability of the signals over the course of the experimental sessions also had an impact on exactly the nature and quality of the speech material finally available for analysis. The subjects had in common that they were phonetically trained an in their thirties at the time of the experiments.

4.5.1 Subject CK

Male, born and raised near Munich. No marked regional characteristics in his speech.

EMG insertions

- CT (right). A fairly posterior location in the muscle, estimated to be between pars recta and pars obliqua.
- SH (left). This insertion resulted from the first attempt to obtain CT.

The signals for both muscles were unambiguous and strong, and the quality did not change noticeably over the course of the experiment.

Speech Material

The target items consisted of pseudo-words containing either tense or lax vowels in either a voiced or voiceless context.

The vowels were /**iː, yː, uː, ɑː**/ (tense), and /**ɪ, ʏ, ʊ, a**/ (lax)[14]

Two voiced and voiceless contexts were used:

/**gəlVbə**/ and /**gəbVlə**/ (voiced) vs. /**gəfVpə**/ and /**gəpVfə**/

Each utterance consisted of a carrier phrase containing either both voiced or both voiceless items, in the order just given:

“Ich habe WORD1 nicht WORD2 gesagt” (= “I said WORD1 not WORD2").

10 randomized repetitions of every sentence were recorded

The stimuli were presented to the subject under computer control on a prompt screen (the computer also generated a synchronization pulse that was recorded on the DAT tape with the other signals). Only the first target word in each sentence was presented on the screen; the subject was required to generate the second target word “on-line” by reversing the order of the target consonants. This was done as a means of keeping the subject alert, since EMG experiments ideally require large numbers of repetitions of very simple speech material. The disadvantage is, of course, that the target words occupy different positions in the sentence intonation contour, so this procedure was not used for the following subjects. In fact, to date only the first target word in each sentence has been analyzed, i.e /**gəlVbə**/ and /**gəfVpə**/ for the voiced and voiceless contexts respectively.

With regard to CT involvement in the voicing distinction, since this was, to our knowledge, the first experiment carried out with a German subject, and since our interpretation of the literature

[14]Throughout this work it will often be necessary to refer to tense-lax pairs together as a single category. This will be done by using the basic Roman letter without length mark, e.g /**i**/ refers to /**iː, ɪ**/. For the a-vowels this system is ambiguous, so the intended reference will be disambiguated verbally, if necessary.

made the outcome seem pretty open, we decided to simply contrast a completely voiced and voiceless context rather than looking in detail at a wide range of voiced and voiceless consonants.

4.5.2 Subject CG

Male, born and raised near Lake Constance. No marked regional characteristics in his speech.

EMG insertions

A very interesting feature of this subject was that three usable CT insertions were obtained:

- One posterior insertion (probably pars obliqua), left side, made along the edge of the cricoid ring to the attachment of the CT to the thyroid cartilage (depth of insertion 2.7cm)
- Two anterior insertions (probably pars recta), right side.

  The first of these two insertions was estimated to be in the middle of the pars recta fibers, between its attachments on the thyroid and cricoid cartilage.

  The second of the two insertions aimed to sample SH activity, but was found to be clearly CT. The insertion was made more obliquely (upward towards the thyroid cartilage).

  Depth of both these insertions was about 2cm.

This is the only recording of which we are aware in which 3 samples of CT activity have been recorded simultaneously, and there are only isolated reports allowing anterior and posterior locations to be compared. Possible differences in activation patterns for speech tasks will be discussed below (p. 97ff). One difference in a non-speech task can be noted here: The posterior insertion showed some activation during swallowing performed as a control task, but neither of the anterior insertions showed any activity at all. Nevertheless, none of the insertions gave any evidence of sampling from the Thyroarytenoid or the Lateral Cricoarytenoid (e.g. as indicated by activation for glottal closure)[15].

Speech Material

For this subject the recording was divided into two parts. In each part, only a single target word was recorded in each utterance.

In both parts, the target vowels were /**iː, yː, eː, uː, ɑː**/ (tense), and /**ɪ, ʏ, ɛ, ʊ, a**/ (lax), i.e one more tense-lax pair than in the corpus for CK.

In the first part the voiced and voiceless contexts were /**gəlVbə**/ and /**gəfVpə**/, respectively (i.e the contexts used for WORD1 in the corpus of speaker CK).

In the second part, the vowels were embedded in the symmetric voiced and voiceless contexts /**gəbVbə**/ and /**gəpVpə**/. In the strong consonantal position at the start of the target syllable we

---

[15]This recording of CG was a repeat of a session done 4 days earlier in which clean CT data was not obtained (the CT signal was considerably contaminated with SH activity). Since this earlier session also gave a clean SH insertion, it was decided not to perform further insertions in the search for SH in the later session.

thus have /**b**/ in Corpus 2 instead of /**l**/ in Corpus 1, and, correspondingly, (aspirated) /**p**/ instead of /**f**/.

For both parts the carrier sentence was "habe ____ besucht" (= "I visited ____").

Once again, 10 randomized repetitions of every utterance were recorded.

4.5.3 Subject SF

Female, born and raised in Eastern Germany. Some regional colouring in her speech

EMG insertions

- SH (left side). No problems encountered.

As with subject CG, two different CT insertions were aimed for, but some difficulties were encountered.

- (1) CT, posterior insertion (left side)

  This insertion was basically successful but movement artefacts tended to occur (including microphonic effects at the voice frequency) especially towards the end of utterances. Possibly, changes in overall larynx height or changes in the angle of the thyroid relative to the cricoid over the course of the utterance may have been disturbing the position of the wires (or they may have been disturbed by vibration in the thyroid).

  The overall quality of the signal was improved considerably (despite a marked decrease in amplitude) by high-pass filtering at 500Hz[16].

- (2) CT, anterior insertion (right side).

  While showing clear CT activity it also had contamination from the sternohyoid.

  Since we also had a good SH signal for this subject (see above), it might conceivably be possible to remove the contamination by estimating the SH contribution to the CT signal. However, as a reliable method for doing this has not yet been identified, this signal will not be discussed further here.

Speech Material

For this subject the same corpora were planned as for subject CG. However, while the second part was being recorded (vowel contexts /**gəbVbə**/ and /**gəpVpə**/) the above-mentioned artefacts in the posterior CT insertion became more pronounced in the central part of the utterances (i.e in the vicinity of the target word). After about 6 repetitions an attempt was made to adjust the position of the wires, but shortly afterwards the signal disappeared completely. As the second part of the corpus thus has fewer repetitions, and possibly at lower quality, it will not be considered in the statistical analyses, but has been included in the sets of figures based on ensemble averages.

---

[16]See Hutters & Rischel (1980) for discussion of choice of high-pass filter cutoff frequency when dealing with EMG artefacts (with special reference to laryngeal EMG and specific problems such as microphonics)

Examples of artefacts

The next figure uses one utterance of speaker SF to illustrate the problems encountered in the EMG signal. The top part of the figure shows the complete utterance, with a microphonic episode in the left-side CT insertion marked by a black arrow. The signal labelled CT_L is the raw signal, CT_LF the high-passed filtered signal. The bottom part of the figure zooms in on about 0.2s of the signal at the arrow location, clearly showing the microphonic effect in the raw signal, and its absence in the filtered signal. The remaining two signals in the upper part of the figure are sternohyoid (SH) and the right-side CT insertion (CT_R). It is quite easy to see that CT_R has activity in parallel both with SH and CT_L, i.e the insertion is not picking up activity just from a single muscle.

Further considerations in the processing of EMG signals are the topic of the immediately following section.

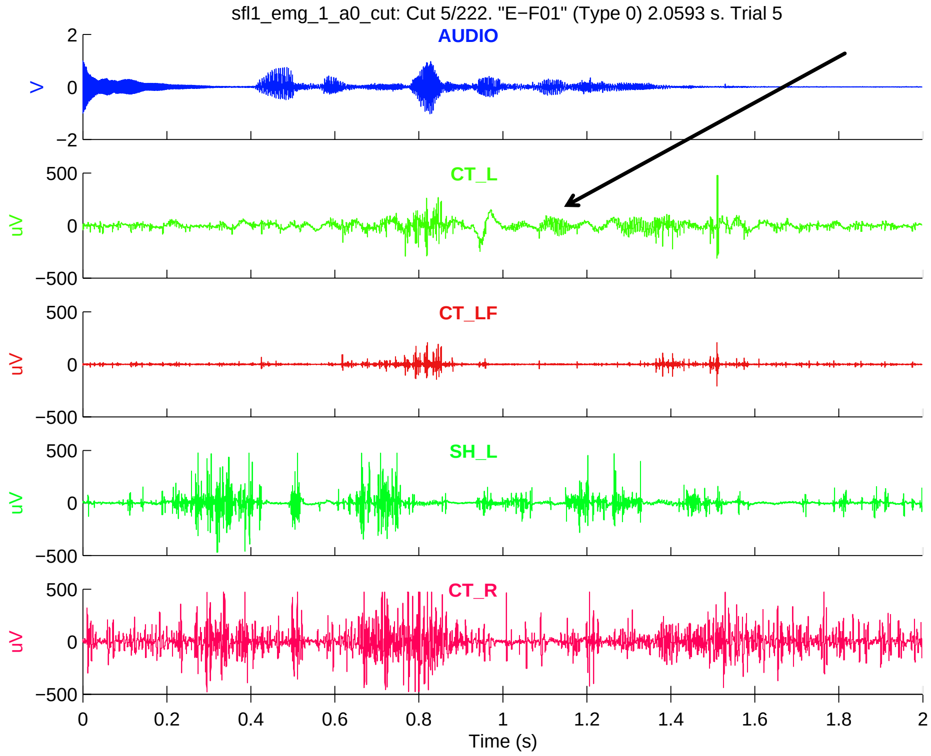


Fig. 4.6: Illustration of disturbances in EMG signal. Arrow marks location of high time-resolution excerpt in the figure below (see text for further details)

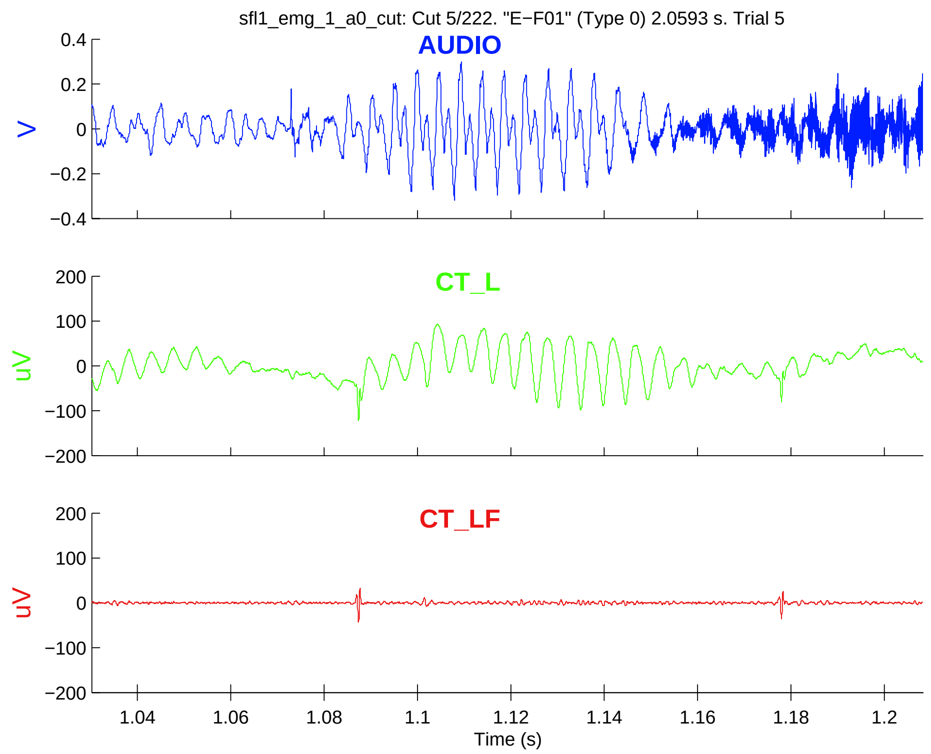


Fig. 4.7: Detailed view of microphonic disturbance

4.6 Processing of the EMG data

4.6.1 Estimating strength of muscle activation

In order to be amenable for analysis the raw EMG signals require processing. In effect, the high-frequency interference pattern resulting from the overlap of a multitude of MAPs (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 20Hz). Analogously to the calculation of an amplitude envelope from a raw audio signal the most basic way of processing the raw EMG signal is to calculate the RMS amplitude over a sliding window. This is undoubtedly the most popular procedure to be found - in numerous guises - in the speech literature. We followed this procedure, too, but also incorporated some variants suggested by the literature. Firstly, Rischel & Hutters (1980) noted that high-frequency emphasis of the EMG signal may provide better differentiation of speech related patterns. Moreover, as already noted above, movement artefacts tend to have low-frequency content, providing additional motivation for emphasizing the higher frequencies (up to the 3kHz bandwidth determined by the preamplifier setting, cf. above). Rischel & Hutters looked specifically at the filtering of signals from the intrinsic laryngeal muscles and concluded that the high-frequency content of these muscles is strong enough that high-pass filtering at settings considerably above typical settings found in the literature can be beneficial for removing movement and microphonic artefacts in these signals. 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. These derived signals will be referred to as **RMS** and **RMSD** respectively. The details of the calculations were:

Window length 40ms
Window shifted in steps of 2.5ms (giving an output sample rate of 400Hz)
Additional smoothing of the 400Hz signal with a Kaiser FIR filter, with Fc=15Hz.

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. These derived signals will be referred to as **ZEROX** and **ZEROXD**, respectively[17].

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)). In a recent simulation study - albeit on the basis of surface electrodes - Zhou & Rymer (2004) investigated how well standard processing parameters such as amplitude and zero-crossing measures can be used to estimate motor unit global firing rate. For amplitude measures the relationship was fairly linear, but depended somewhat on the motor-unit firing scheme and the relationship between electrical and force

[17]In order to avoid spuriously high zero-crossing rates in weak portions of the signal close to the noise level, a noise-band was determined by eye in each raw signal over a stretch of signal where the muscle was inactive. Only those zero-crossings were counted where the signal also crossed this noise-band about zero (cf. E.g Zhou & Rymer, 2004).

output assumed in the simulations. For zero-crossing rate, the relationship was less dependent on the simulation parameters, but was only linear at low firing rates, probably because of the large amount of superposition of individual motor-units at higher rates (though note that we are certainly a long way from maximum muscle contraction in the kind of experiments carried out here). This study thus provides some justification for considering alternative parameterizations of the raw EMG signal, since in practice for any given insertion we are largely ignorant about the precise motor-unit composition of the signal we are recording.

Let us just mention in passing that 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 (and gave similar results to zero-crossing in Zhou & Rymer's study)[18].

The next figure shows an example of all four measures for capturing the EMG activity. This was chosen as an example where there are some characteristic differences between the measures; in many other cases the four measures were very similar indeed.

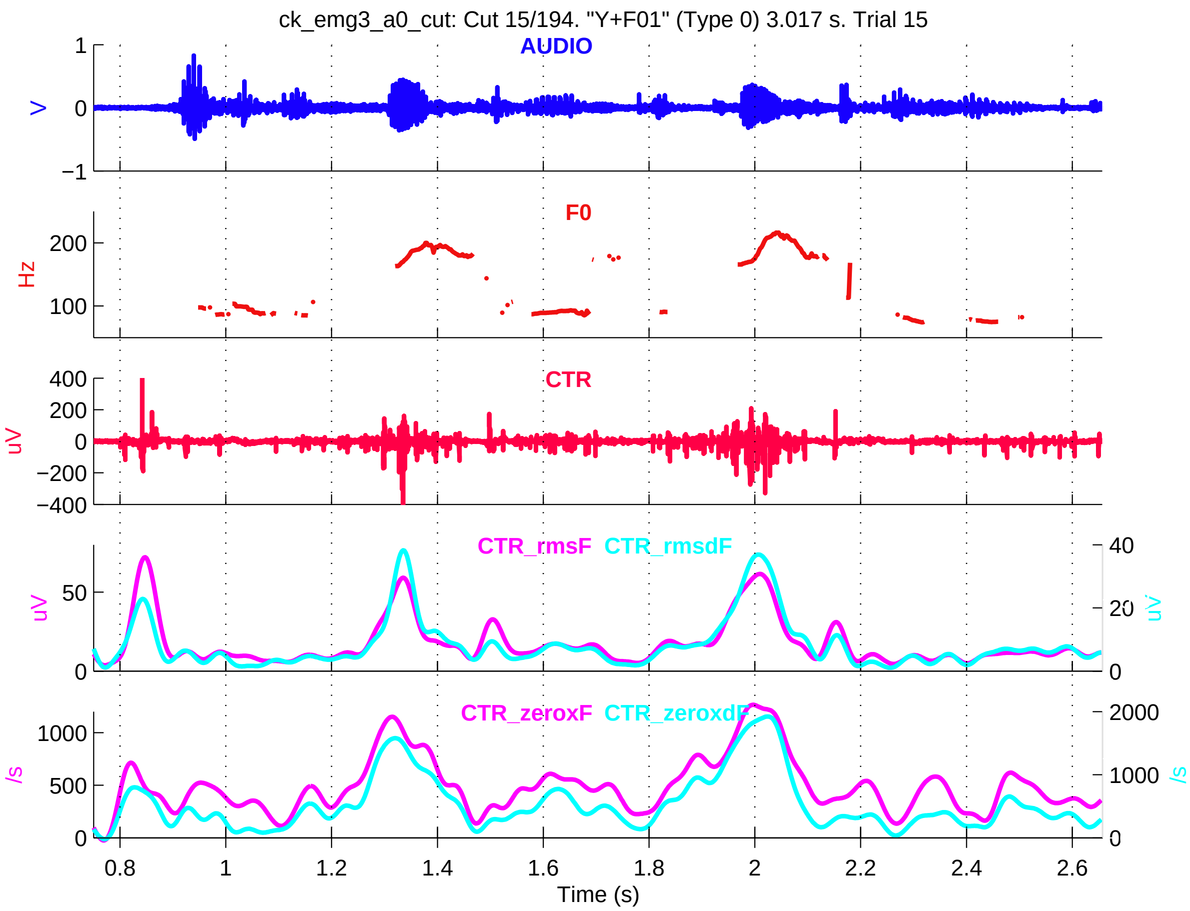


Fig. 4.8: Example of estimation of the level of EMG activity using the four derived signals RMS, RMSD, ZEROX and ZEROXD. The bottom panel shows the zero-crossing based signals, the bottom panel but one the RMS-based signals. In each of these two panels the magenta-coloured signal is based on the undifferentiated raw signal and the cyan one on the differentiated version. The utterance is "ich habe gefüpe und nicht gepüfe gesagt"

[18]First differencing can also be seen as a primitive way of whitening the EMG power spectrum, which has been suggested as a pre-processing step for increasing the reliability of EMG amplitude estimation (Clancy et al., 2002).

In this example, there are several occurrances of spikes in the raw signal that have much higher amplitude than the surrounding signal, and are probably not representative of the overall level of the signal (for example just after 0.8s, just before 2.2s, at about 1.5s on the time axis). Not surprisingly, such events have a much stronger effect on the RMS than on the zero-crossing signal, and a stronger effect on the undifferentiated than on the differentiated signal. Thus the small peaks at 1.5s and just before 2.2s in the RMS signals are not visible at all in the zero-crossing, and the peak just after 0.8s is much larger in the RMS than the zero-crossing. It seems plausible that the amplitude of all these peaks is overweighted in the RMS signal. There is also an indication in both the RMS and zero-crossing signals that the respective differentiated version shows less unsystematic fluctuation, and allows the peaks associated with the two clear F0 prominences to emerge more saliently.

Considering the recordings as a whole it was by and large the case 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 this version of the EMG parameterization 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 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. The relevant comparisons are given in Table 4.1 below. These correlations were computed from the average EMG activity in the vowel, and the median F0 (the EMG was shifted in time as explained below). Because of their longer duration it was expected that tense vowels would give more robust results, so the results in the table are based on these vowels (lax vowels were also examined and did in fact almost invariably give lower correlations)

For CK and SF there was a clear difference in favour of zero-crossing rate. For all insertions of speaker CG we chose RMS amplitude on the basis of the stronger correlations for the posterior and first anterior insertion (for the second anterior insertion, where correlations are overall weaker, there is little to chose between the two parameters).

A further point that emerges from this table is that it is not possible to make a distinction between posterior and anterior insertions regarding strength of the correlation with F0. The weakest correlations in the whole table are found for CG's second anterior insertion, but his first anterior insertion is extremely similar to his posterior insertion. We will, however, see that a slightly different picture emerges when we consider the effect of the experimental variables at the end of Chapter 5 (p. 97ff).

	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
RMSD	0.68	0.56	0.74	0.74	0.5	0.6	0.66	0.34
ZEROXD	0.85	0.73	0.68	0.39	0.48	0.47	0.42	0.37

Table. 4.1: Comparison of RMS amplitude and zero-crossing rate (both based on differentiated signal) for correlations of EMG activity with F0 (Pearson's r). The row labelled "Ins." indicates the location of the CT insertion: P=Posterior and A=Anterior

4.6.2 Time alignment of EMG 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.

In his influential study Atkinson (1978) arrived at a delay of 40ms for CT, calculated by looking for the lag in the the cross-correlation function between CT activity and F0 giving the strongest correlation. As he notes, this figure agreed quite well with muscle contraction times reported in the literature for non-human but anatomically comparable species. Using a completely different paradigm involving prolonged steady phonation and special averaging techniques Baer (1981) estimated that the F0 perturbation caused by a single firing of a single motor unit in CT peaked at 70-80ms. Larson et al. (1987) found Baer's figure unexpectedly large, and in effect repeated his experiment with more subjects, finding a mean CT latency of about 22ms (with a range of 6-75ms). Using a completely different paradigm again, Sapir et al. (1984) examined CT-F0 latency when speakers were required to modulate F0 sinusoidally at a rate of 1Hz. They found a delay of about 150ms. They note that this is, of course, much larger than Atkinson's figure but considered it similar to findings of Erickson et al. (1981). I am not completely convinced that this is the case. These authors unfortunately do not quote specific numbers (they were interested in relative contribution of CT suppression and SH activation in F0 falls, rather than CT latency as such), but I would estimate from their figures that CT latencies were in most cases between 50 and 100ms. This agrees quite well with the figures given by Dyhr (1990) for the Copenhagen laryngeal EMG investigations. This in turn agrees with figures given by Sawashima et al. (1982) for latencies between CT increase and F0 increase (averages for each utterance category examined were in the range 50-100ms). An interesting feature of this paper is that they explicitly distinguish latencies for F0 rise following CT increase and latencies for F0 fall following CT decrease. The latter case typically has longer (and more variable) latencies.

Finally, in a relatively recent study Herman et al., (1999) consider the problem of delay in some detail. Unfortunately - since the study is carefully done - they arrive at the high figure of 150ms for CT (and more for SH). The analysis of CT was restricted to single rising segments of the intonation contour. This is certainly basically a good idea. It is a pity that they did not do a separate analysis for CT with falling contours (only SH was analyzed in this case). They do not really discuss why they arrive at a figure about 100ms longer than Atkinson. In view of these troubling discrepancies in the literature one wonders whether the influence of window length on time alignment of data (this can be relevant both for EMG and F0) may have been handled differently in the different studies.

Can any clear message be derived from this mosaic of results?

As Sapir et al. point out, it is physiologically plausible that latencies for movements involving multiple motor units are longer than those for single motor units. Also, contraction of CT is still mechanically at a couple of removes from change in the tension of the vocal folds, so pure muscle contraction times probably define a lower limit, rather than what is behaviourally relevant. A complicating factor for correlations based on complete intonation contours is that CT is not of course the only factor affecting F0, and, as we have just seen, latencies could differ over rising and falling portions of the contour. The fact remains that we are fairly ignorant of how different laryngeal tasks impact on the observed latencies.

In the attempt to get an empirically supported value for our own dataset we first tried cross-correlation analyses of F0 and CT for each utterance individually as in Atkinson's approach, but found it difficult to get stable results, probably because many utterances had intervening voiceless segments, and probably also because of the complications caused by different relationships for rising and falling portions of the contour. As an alternative we then adopted the following procedure based only on F0 in the target vowel:

Preliminary inspection of EMG and F0 contours suggested that EMG led F0 by a time in the range of 30 to 120ms. We then shifted the EMG in steps of 10ms over this range, and at each time step determined the average EMG and median F0 for each target vowel and calculated the correlation coefficient between these two parameters. 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. Table 4.1 above is based on this time-shift value. Often, correlations were already quite strong at the shortest lag of 30ms but had declined noticeably by 120ms. We consider the final figure of 60ms to be quite well-justified from the literature: it is at the lower end of the 50-100ms range often encountered, which seems appropriate as the target region of our utterances is generally located on a rising F0 contour, and we are also interested in potential increased CT activation for voiceless consonants. In other words, the latency should be more appropriate to CT activation than to CT suppression.

A representative example of ensemble-averaged (see below) EMG and F0 aligned in this way is shown in the following figure. It is indeed noticeable that EMG activity falls away faster after the activity peak than does the F0 contour.

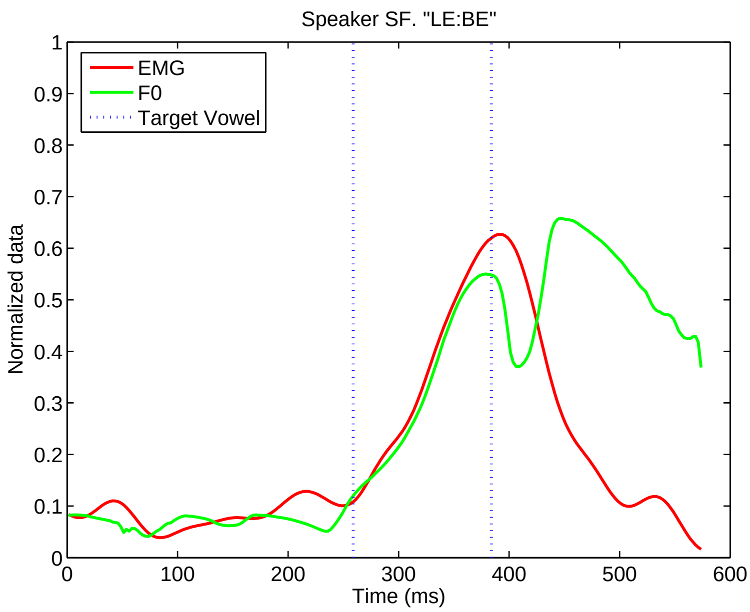


Fig. 4.9: Example of ensemble-averaged EMG and F0 (speaker SF)

4.6.3 Further pre-processing (1): Ensemble Averages

In view of the inherent variability of the raw EMG signal, one of the main approaches to identifying the salient patterns in the data has been to calculate ensemble averages over all repetitions of a given speech utterance (in our experiments we generally aimed for 10 repetitions per utterance, which is about the lower limit for what has typically been used in the literature). In order to be able to perform the averaging a line-up point is required that can be identified independent of the EMG signal. Typically some clear event in the audio signal, such as release burst of a consonant is used. If, however, the temporal structure of the utterances is not very similar, then the averaging tends also to result in a temporal smearing effect as distance from the line-up point increases. To counteract this effect, various procedures for time-warping each utterance before averaging have been suggested (e.g Strik & Boves, 1991). For our purposes, where the utterances are quite short, a middle way between use of single line-up point and a continuous timewarping function appeared adequate. The following procedure was adopted:

The speech wave was segmented into C1, V and C2 segments[19]. Then for each utterance category the token with the most typical duration for these 3 segments was determined (minimum summed deviation from the mean durations). This "typical" token was taken as a template, and the durations of the three segments in all other tokens were warped to the target durations given by the template. After warping the signals in this way they were averaged over all tokens (in addition, to include somewhat more context around the target segments, additional "dummy" segments with a fixed duration of 150ms and 100ms were incorporated in the averaging procedure before C1 and after C2, respectively. The procedure was carried out interactively, so that tokens with noticeable remaining artefacts, or very strongly deviant temporal structure could be skipped. The following figure compares for one utterance type over a time interval corresponding to that from onset of C1 to offset of C2 the ensemble that would result with and without time-warping (the non-time-warped ensemble is effectively lined-up at the onset of C1). Note the higher and more sharply shaped peak in the time-warped version. In other words, without time-warping some smearing of the shape of the EMG activity pattern probably occurs.

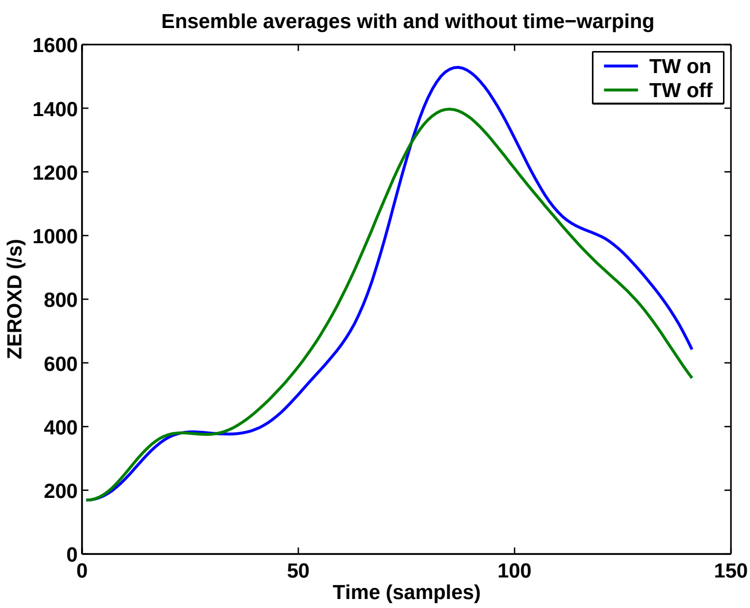


Fig. 4.10: Example of ensemble-average computed with and without segment-wise time-warping. See text for details

[19]Much of the segmentation and labelling work was carried out by the team of the phonetics lab at ZAS, Berlin, to whom I am most grateful.

Conceivably, in addition to this temporal adjustment one could also normalize signal amplitude over all utterance categories for each block of repetitions, in order to reduce the effect of overall changes in signal amplitude over the course of the experiment (for example, caused by shifts in electrode position). However, the signals appeared stable enough to make this unnecessary.

4.6.4 Further pre-processing (2): Data reduction for statistical analysis

In many EMG studies of speech, presentation of the results is often confined to qualitative discussion of ensemble averages. We were encouraged by the approach in the investigation of Löfqvist et al. (1989) to envisage carrying out classical statistical analysis such as ANOVA in which each token constitutes an observation. The variability in individual EMG signals still made it seem unwise, however, to follow the procedure typically used in many acoustic or kinematic investigations of extracting the signal amplitude at a single point in time, so instead we extracted *average* signal amplitude for each of the three acoustically-defined segments just mentioned, namely C1, V and C2 (the procedure of Löfqvist et al. was slightly different, as they computed averages over fixed-length windows).

4.7 Acoustic processing: Calculation of fundamental frequency

The acoustic parameter of key interest in this study is F0. This was extracted using the the YIN algorithm of de Cheveigné & Kawahara (2002; available as a MATLAB toolbox). The algorithm appeared quite well able to cope with the somewhat less than ideal conditions for acoustic recordings during the experimental sessions.

The basic processing paramters were: Window length 33ms, window shift 2.5ms (giving a 400Hz samplerate for the F0 contours to match the EMG data. Minimum and maximum F0 search range set by hand for each speaker

5 Results

5.1 Overview of the presentation of the results

The structure of the experiment involves three independent variables, namely voicing of the consonant context, vowel tenseness, and vowel category (4 vowel categories for CK, five for CG and SF). These three factors will be referred to as VOICE, TENSE, and VOWEL, respectively, and the results will be presented in that order in three main sections corresponding to these factors. Note that under the heading of the factor VOWEL we are in effect looking at the traditional intrinsic pitch question, but as the question of vowel tenseness is of more central concern here this will be presented before it, and at greater length (in fact some aspects of the analysis of VOWEL also give a good summary of the tense vs. lax effects discussed immediately beforehand). In each of these sections the time course of EMG activity will be presented using ensemble averages, and the results of statistical analysis carried out for each of the segments C1, V and C2 will be discussed. The relevant F0 patterns will also be presented in a similar way (with ensemble averages and statistical analysis), and will be related to the EMG findings. Presentation of the results in these three sections will result in a certain amount of overlap (since the underlying dataset remains the same), but seemed preferable because of the very different nature of the hypotheses attached to each independent variable. On the other hand, one unexpected finding (with interesting linguistic and methodological implications) was that of fairly frequent interactions between the VOICE and TENSE effects. This will be discussed in detail at the appropriate point in the exposition of the results, but effects of this kind made it seem advisable to give a complete breakdown of the statistical analysis (carried out with the GLM procedure of SPSS) here at the beginning of the results section, so that reference can be made back to it whenever necessary.

5.2 Tabulation of statistical results

The first and probably most useful table in this section attempts to compress an overview of all main effects and two-way interactions into a single table (there were no significant three-way interactions). The following typographical conventions have been used:

Significant main effects in the expected direction: x ($p<0.05$), X ($p<0.01$), **X** ($p<0.001$)

Significant main effects not in the expected direction: o ($p<0.05$), O ($p<0.01$), **O** ($p<0.001$)

Interaction effects are given in italics after the main effect (if any), with the following letters indicating the source of the interaction:

V = **V**OICE, *T* = **T**ENSE, *W* = VO**W**EL.

The significance level follows the same convention as the main effects:

lowercase ($p<0.05$), *UPPERCASE* ($p<0.01$), ***UPPERCASE BOLD*** ($p<0.001$)

Example: The top left cell of the first table (main effect of VOICE) contains an italic *T*, indicating a VOICE x TENSE interaction that is significant at $p<0.01$. The same interaction term appears at the same location in the following table (main effect of TENSE), but now indicated reciprocally with *V*. In order to give an overview of the interactions without this typographical ambiguity, all three possible two-way interactions are tabulated in the second table of this section (using the x/X/**X** notation to indicate significance).

Regarding the division of the main effects into an expected and unexpected direction the following choices were made for the expected direction (basically corresponding to the hypotheses we aim to test):

VOICE:	Voiceless stronger
TENSE:	Lax stronger
VOWEL:	High vowels stronger

This division should be considered as a rough guideline only. Whenever strong interactions occur a clear division may not be possible, and is also not so straightforward for the factor VOWEL, which unlike VOICE and TENSE has more than two levels.

One point that emerges from a first view of these tables is that for the independent variables of primary interest, namely VOICE and TENSE, effects are weaker for the anterior insertion locations for subject CG. In the three main results sections below we will accordingly concentrate on results from the posterior location, and reserve comparison of results from the different insertion locations for a separate brief section at the end (p. 97ff).

As a further aid to orientation in the following table we have accordingly outlined in bold those cells that correspond to the insertions and segments which are most central to the following analyses: for the segments these are Consonant 1 (C1) for consideration of VOICE, and Vowel (V) for consideration of TENSE and VOWEL.

FACTOR VOICE								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1	**X** *T*	**X** *W*	**X** *tW*	*t*		**X**		X
V	X *T*	**X** *tw*				X		
C2	**X**	**X** ***T***				X		

FACTOR TENSE								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1	X *V*	X *W*	**X** *v*	*v*		x		
V	X *V*	O *v* ***W***	**X**	x		X *W*		x ***W***
C2	**X**	**X** ***V***	**X**			**X**	X	X *W*

FACTOR VOWEL								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1	**X**	X *VT*	**O** *V*	x	o	**O**	X	O
V	**X**	**X** *v* ***T***		x	**O**	**O** *T*	**X**	**O** ***T***
C2	X	X		**O**	**O**	**O**	**O**	**O** *T*

Table 5.1: Overview of main effects and two-way interactions each segment. The row labelled "Ins." indicates the location of the CT insertion: P = Posterior, A = Anterior. See text for further explanation.

VOICE x TENSE INTERACTION								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1	X		x	x				
V	X	x						
C2		**X**						

VOICE x VOWEL INTERACTION								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1		X	X					
V		x						
C2								

TENSE x VOWEL INTERACTION								
	Corpus 1					Corpus 2		
Subj.	CK	SF	CG			CG		
Ins.	P	P	P	A1	A2	P	A1	A2
C1		X						
V		**X**				X		**X**
C2								X

Table 5.2: Overview of all two-way interactions for each segment. Other details as for Table 5.1.

5.3 Results for Consonant Voicing

First of all the results of EMG analysis will be presented, then the results for F0, and then the relationship between EMG and F0 patterns will be discussed.

5.3.1 EMG results

Ensemble averages

Results will first be presented by juxtaposing the EMG ensemble averages for corresponding voiced and voiceless consonantal contexts. The curves for each voiced-voiceless pair are lined-up at the release of C1 (indicated by dashed vertical line at zero on the x-axis, and labelled “C1-Offset”). The circle symbol on each curve to the left of the line-up point indicates the start of C1, and consecutive circle symbols on each curve to the right of the line-up point indicate offset of the target vowel, and release of C2 respectively. Figures are arranged on facing pages so that utterance categories with tense target vowels are on the left and lax on the right. For speakers CG and SF the two corpora are shown separately (/lVb/ vs. /fVp/ and /bVb/ vs. /pVp/, respectively), with Corpus 1 on the left half of the page and Corpus 2 on the right (recall that corpus 2 for Speaker SF had fewer repetitions and has been omitted from the statistical analysis)[20].

[20] When ensemble averages are calculated it is also possible to compute the variance at each time-point. This could be used in the figures to give additional information on the separation between the curves in the different conditions. However, we have preferred to omit this here to keep the figures clearer. The variability is taken into account by performing the statistical analyses.

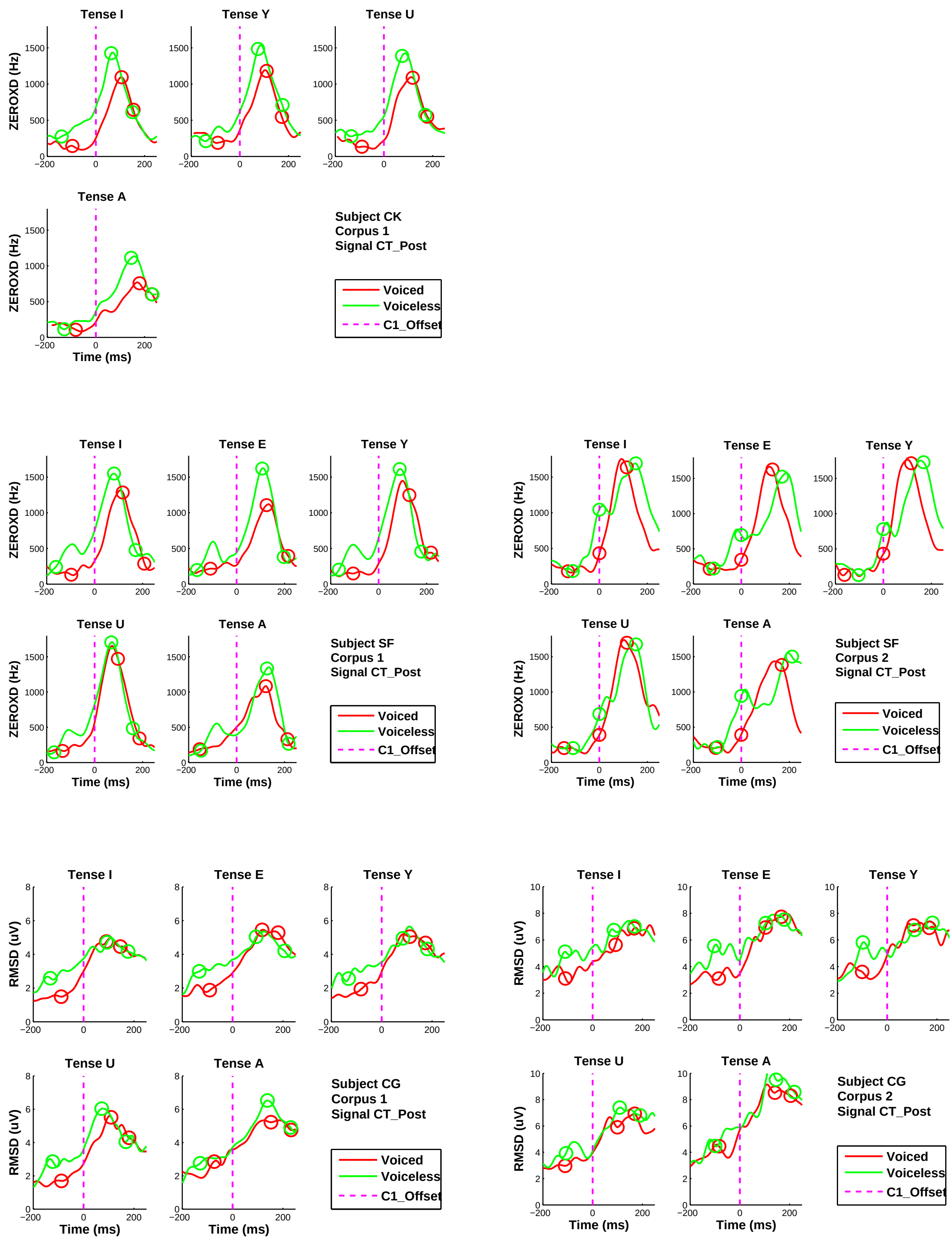


Fig. 5.1: Ensemble averages of CT activity in target words with tense vowels. Each panel contrasts corresponding voiced (red) and voiceless (green) consonantal contexts. Corpus 1 on left side, corpus 2 (for speakers SF and CG) on right side. Line-up point is release of C1. Other segment boundaries indicated by circles. Each ensemble normally based on average of 10 repetitions.

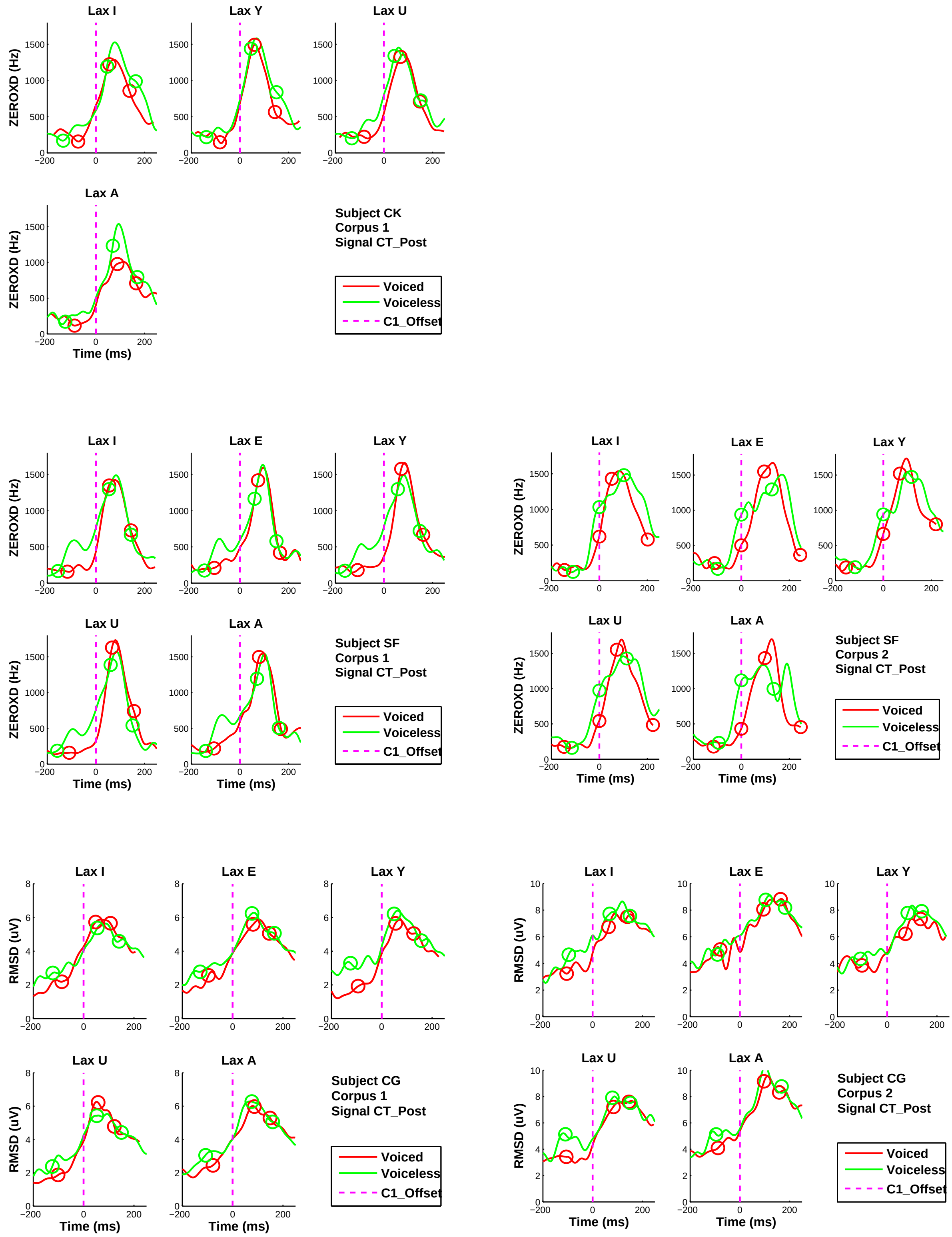


Fig. 5.2: Ensemble averages of CT activity in target words with lax vowels. Other details as for figure with tense vowels on opposite page

Taking the material as a whole, it is clear that there is a very consistent effect for more CT activity in the voiceless consonants. A possibly distracting effect in the figures is that the main peak in CT activity (usually about 100ms after the line-up point) is simply related to the rising intonation contour over the target syllable of the utterance. The most important location to look initially is just to the left of the line-up point (between the leftmost circle symbol and the line-up point), i.e at the time segment corresponding to the pre-stress initial consonant C1, since this is where most vigorous devoicing activity is to be expected. Here, almost invariably, higher activity for the voiceless consonants is indeed to be observed, regardless of speaker, corpus and vocalic context. For Speaker SF, in particular, a rather clear pulse of activity is often to be observed in the consonant, i.e a local activity maximum within the consonant, with a slight decline in activity towards the end of the consonant. There is more diversity in the results immediately following the line-up point, i.e in the target vowel. Speaker CG shows no consistent differences between voiced and voiceless contexts after the line-up point (his curves are also rather more 'bumpy' than those of the other two speakers, indicating that a larger number of repetitions might have been useful). For CK and SF, however, there are several cases where higher CT activity persists right through the vowel - especially for tense vowel contexts from Corpus 1. On the other hand, since this is not invariably the case it will be of interest to examine below to what extent the different strength of the voicing effect in the vowel is observable in the F0 contours. Regarding the post-stress consonant C2 (/**p**/ vs. /**b**/ in both corpora; between the first and second circle symbols after the line-up point), this is the segment where it is most difficult to see any clear pattern in the data. This is partly here for the purely graphical reason that with the lineup point set at release of C1 the location of C2 is not at the same point in the voiced and voiceless contours. For CK and SF in the tense vowel contexts of Corpus 1 there are indeed some clear cases where CT activity is higher at the start of C2, but the design of the corpus unfortunately make it difficult to decide whether this is really the result of increased activity for C2, or rather a carry-over effect from the vowel (and C1). Overall, this initial view suggests that voiceless C2 leads less consistently to higher CT activity than does C1, but for a balanced view it will be necessary to refer to the statistical analysis below. Finally, regarding Corpus 1 vs. Corpus 2 (i.e pre-stress [**f**] vs. [$\mathbf{p}^h$]) , it would appear that both voiceless consonants lead to a similar increase in CT activity. However, for speaker SF there is an interesting difference in the timing. For corpus 2, with aspirated [$\mathbf{p}^h$] the whole time-course of the CT activity appears shifted to the right. Note that the line-up refers to the release of the consonant, not to the onset of voicing. It is thus perhaps not surprising that the hypothetical "pulse" of activity related to the consonant itself gets shifted to the right. One could even imagine that it is linked quite closely to the time-course of the glottal abduction-adduction gesture. This could support an interpretation, to which we will return later, that the purpose of the CT activity is not just to stop voicing at the onset of voiceless consonants, but also to prevent voicing restarting too readily in the aspiration phase (cf. Hanson & Stevens, 2002). This shift in the timing does lead, however, to a rather complicated pattern in the vowel: For corpus 2, the level of CT activity is actually ***lower*** for much of the vowel, apparently because the rise up to the overall CT peak starts later. Again, it will be interesting to see to what extent these complications are visible in the F0 contours. The preceding remarks apply only to SF; for speaker CG there are no apparent differences in timing between the two corpora. One reason for this could be the relative length of the aspiration phase for the two speakers. Overall average VOT for CG was approx. 40ms, for SF approx. 73ms.

Statistical results for each segment

On the following pages we present the results in summary form, based on the average EMG activity for each of the segments C1, V, and C2, in turn, linking these observations with the statistical results that were tabulated at the start of the results section.

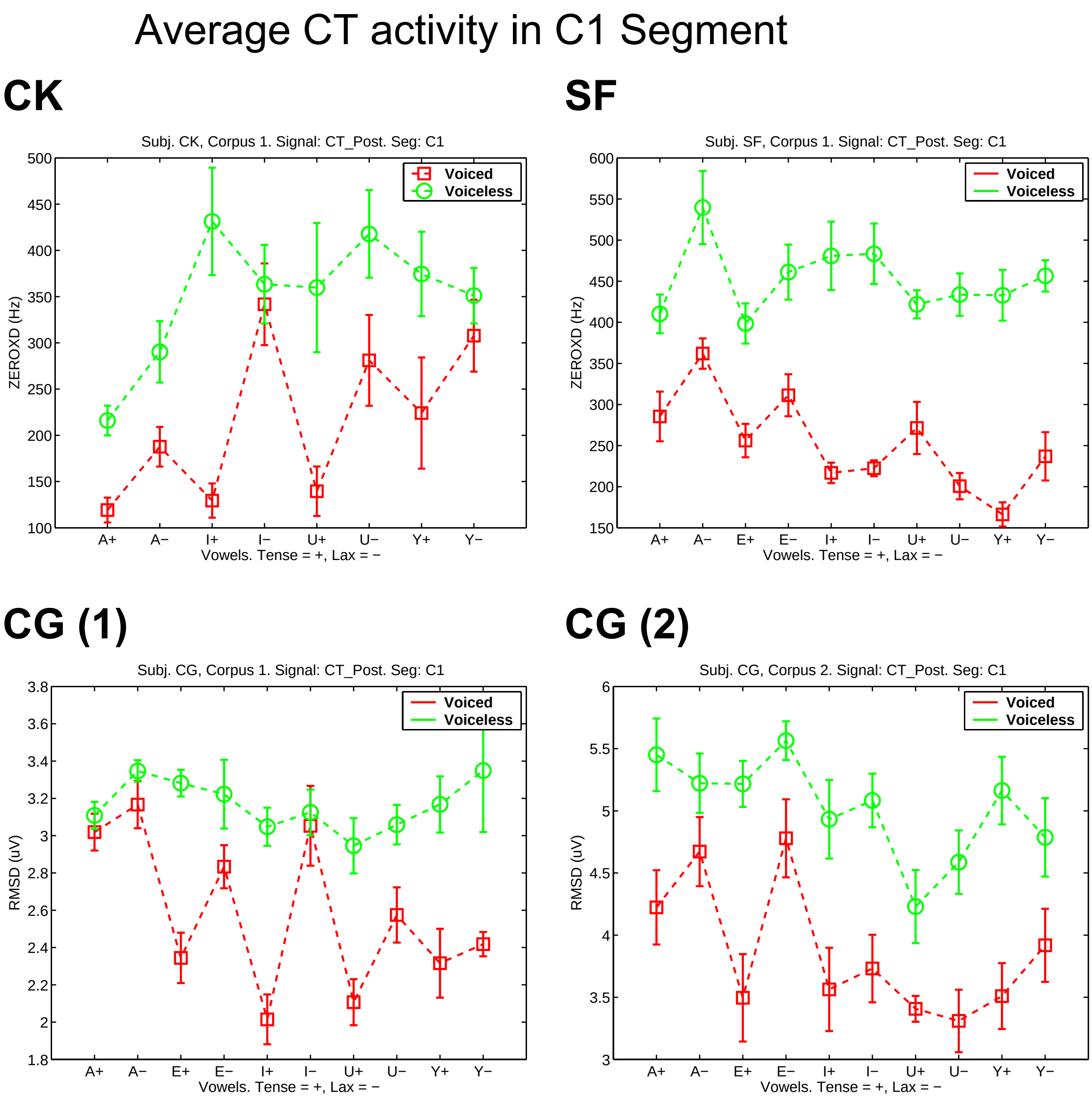


Fig. 5.3: Mean (over 10 repetitions) and standard error of average CT activity in each C1 segment, broken down by following vowel. Separate panels for each speaker (and separate panels for each corpus for speaker CG). Vowel category indicated on abscissa by upper-case letters, with vowel tenseness indicated by '+' for tense and '**-**' for lax. The identity of C1 was /**p**/ and /**b**/ for CG(2), /**f**/ and /**l**/ in all other cases.

<u>C1 Segment</u>

This figure shows the mean (and standard error of the mean) of EMG activity in C1 for each vowel category. Note that tense-lax pairs occupy adjacent positions on the x-axis. Each speaker is shown in a separate panel (for SF only Corpus 1; as mentioned above Corpus 2 was excluded from the statistical analyses because of the lower number of repetitions).

The most obvious point is that the figure simply reinforces the impression already gained from the ensemble averages that the voicing contrast in the consonant has very robust effects. For all speakers (and both corpora for CG) the main effect of VOICE was significant at $p<0.001$. The figure also reveals a point that is less easy to derive from the ensemble averages, namely that there is a tendency for the effect of VOICE to be clearer when C1 is followed by a tense vowel. This becomes apparent from the rather zig-zag course of the lines in the above figure (particularly for the voiced consonant), since tense-lax pairs are adjacent to each other. The VOICE * TENSE interaction was in fact significant for CK ($p<0.01$) and CG in Corpus 1 ($p<0.05$). We will encounter this interaction in various guises throughout the results section.

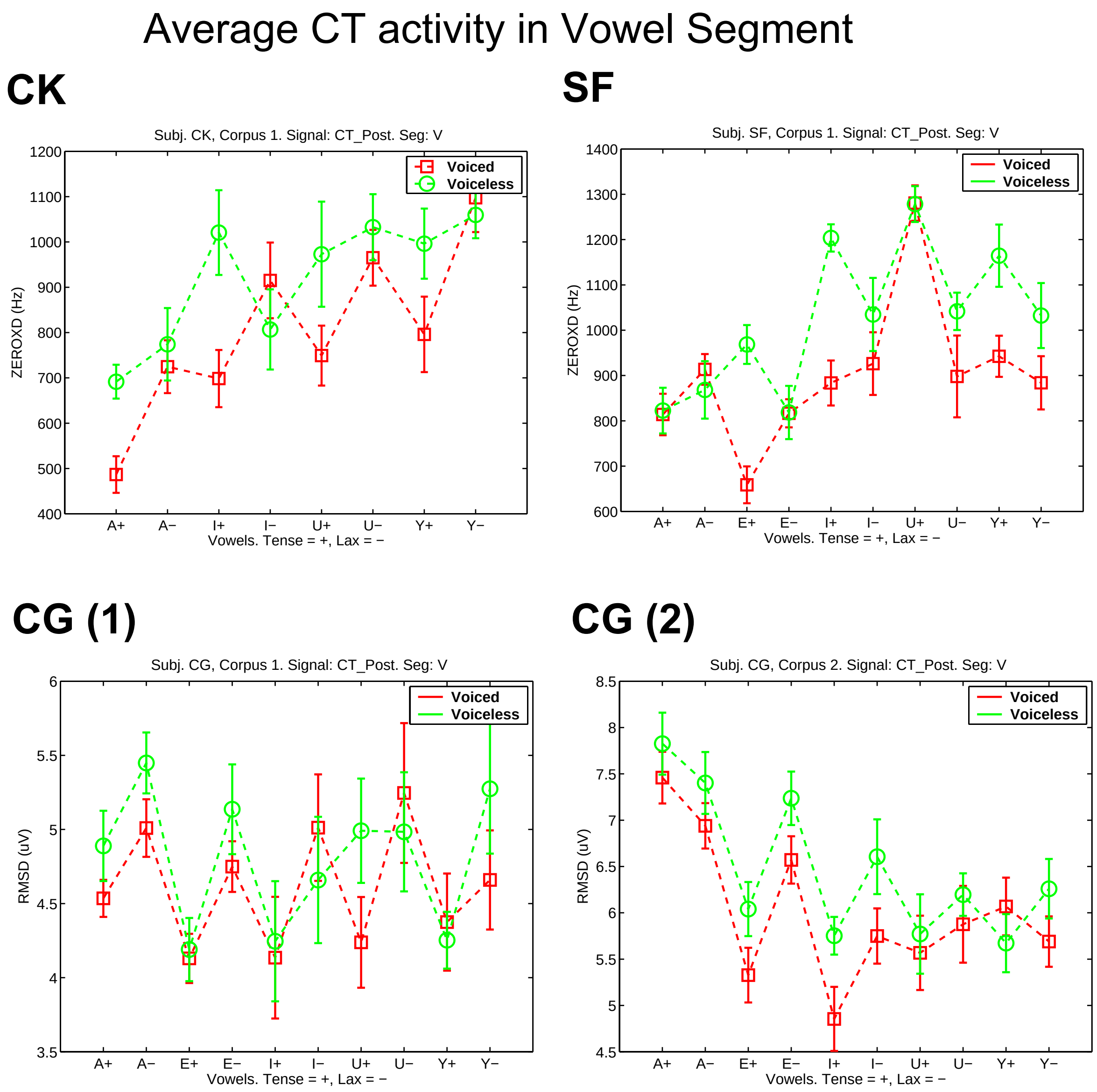


Fig. 5.4: Mean and standard error of CT activity in vowel segment. Other details as in previous figure

<u>Vowel Segment</u>

Since the main expected acoustic consequence of higher CT activity, namely higher F0 (to be looked at immediately below), will be located in the vowel segment, it is, of course, interesting to examine the relative strength of voicing effects in the vowel compared to the primary consonantal segment itself. The summary results in the figure above indicate, of course, that the direction of the effects remain the same, i.e higher CT activity in voiceless contexts, but the separation between voiced and voiceless is clearly less sharp than in the consonantal segment itself. Except in the case of CG Corpus 1 the main effect of VOICE is still significant: CK $p<0.01$, SF $p<0.001$, CG Corpus 2 $p<0.01$. However, the results were again complicated by the VOICE * TENSE interaction. The interaction was significant for CK ($p<0.01$) and SF ($p<0.05$): for these two speakers the VOICE effect was really only apparent in the tense vowels, being essentially negligible in the lax vowels.

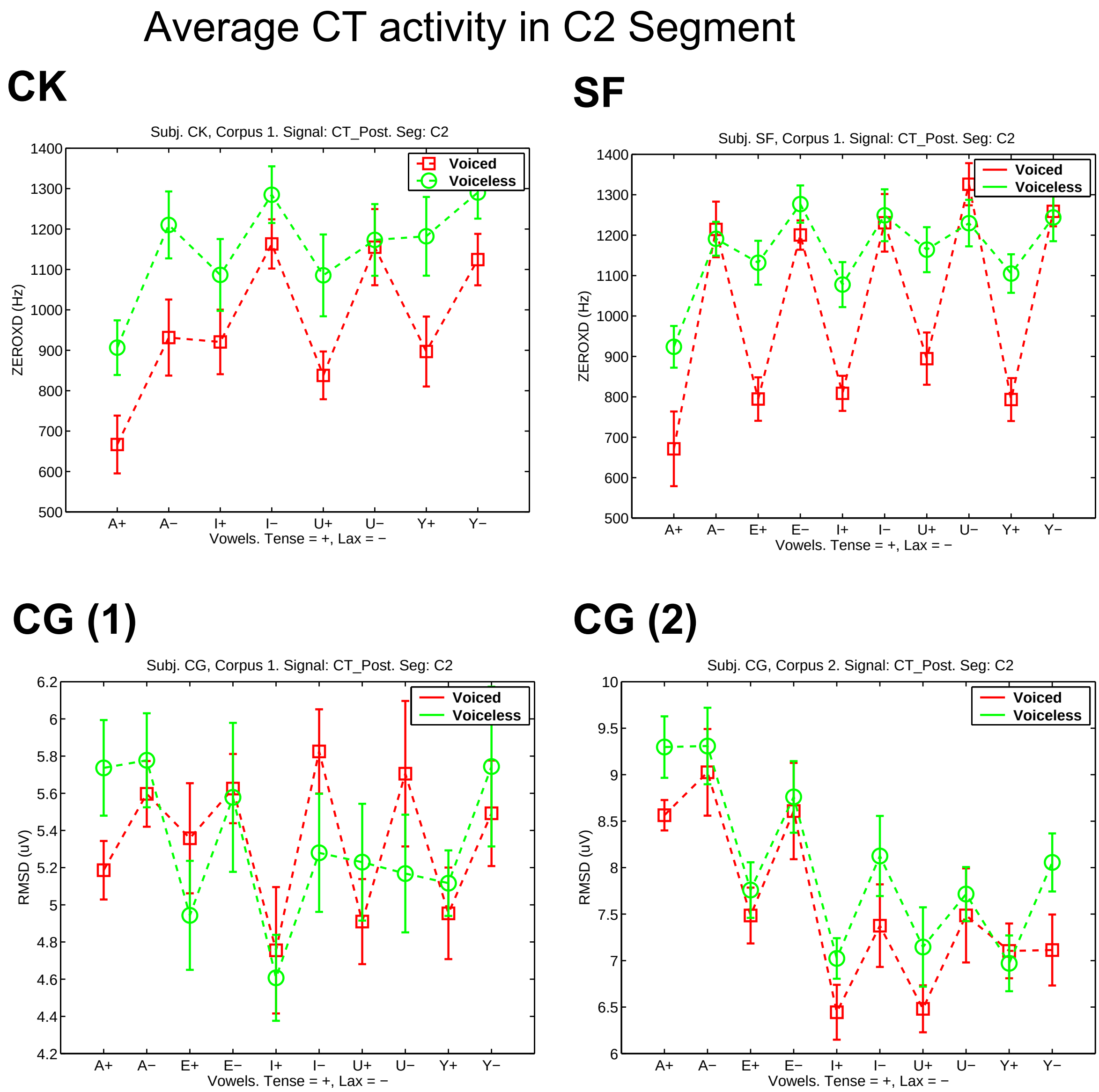


Fig. 5.5: Mean and standard error of CT activity in C2 segment. Other details as for previous figure.

C2 Segment

The most important point to make for this final target segment is that speaker-specific patterns are much more in evidence than for C1: For CK the effect of VOICE remains very clear ($p<0.001$). Note that the scaling of Fig. 5.5 is not the same as for Fig. 5.3 (C1) so the possible visual impression of a stronger effect for C1 is misleading. For SF the main effect of VOICE is also significant at p<0.001 but the already encountered VOICE x TENSE interaction becomes particularly strong (p<0.001), owing to the effect of VOICE being very large for tense vowels but negligible for lax ones. CG is the only speaker for whom VOICE effects are consistently weaker for C2 than C1: n.s for Corpus 1, and p<0.01 for Corpus 2.

One could have tested formally for differences in VOICE effects over the different segments by expanding the ANOVA to include segment as a factor, and then looking for significant interactions. However, because of the blatant speaker differences this would not have been very helpful. Instead we will simply summarize the differences by means of a figure in the next section.

Summary of statistical results

To sum up the results presented hitherto, we assemble in a single figure the voiced-voiceless differences for all speakers and all three segments (arranged in temporal order from left to right on the abscissa).

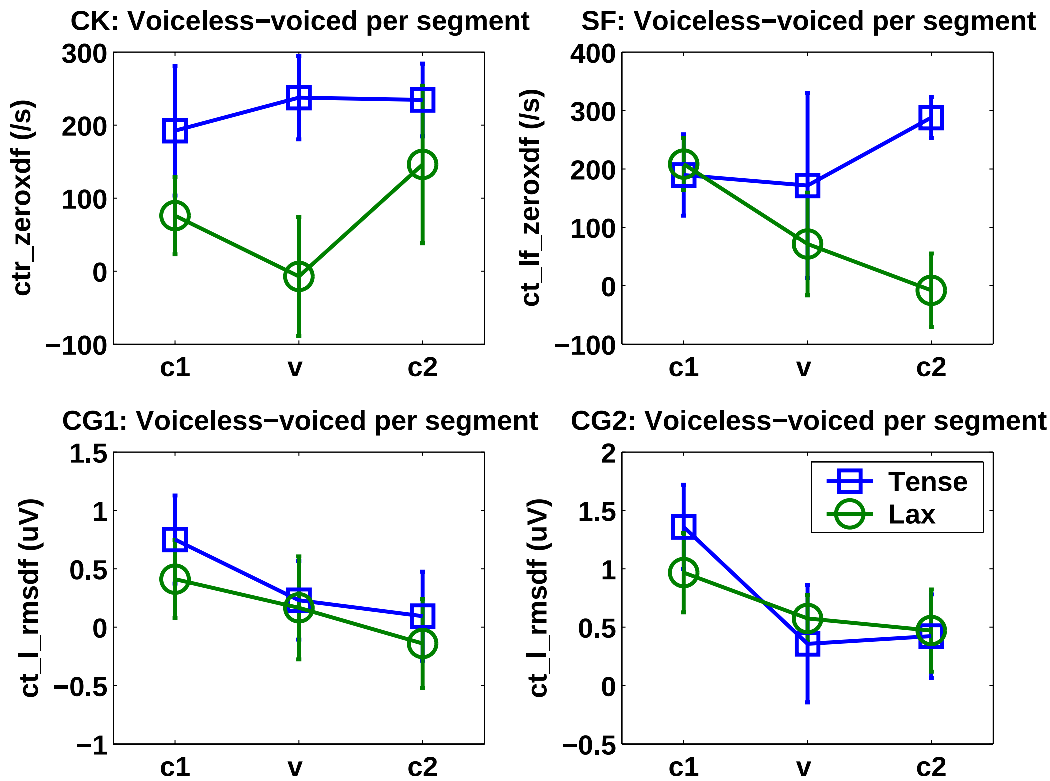


Fig. 5.6: Difference in CT activity between corresponding segments in voiced and voiceless contexts. Positive values indicate more activity in voiceless context. Calculated separately for tense and lax vowel contexts. First average EMG activity for each word form was calculated. Then each data point was calculated as the average difference between word pairs contrasting only with respect to voicing category. Error bars are the standard deviation over the corresponding number of word pairs: 4 for CK, 5 for all other speakers.

Summarizing the results for the three segments, we observe an extremely clear VOICE effect in C1, which then tends to become weaker in the following segments, especially for speaker CG. In several cases the basic pattern of stronger CT in voiceless contexts was overlaid by an interaction with TENSE, the voicing effect being clearer in tense vowel contexts. Further interpretation of the results now requires examination of the associated F0 patterns. Final interpretation of the somewhat mysterious VOICE * TENSE interaction will not be attempted until we have specifically examined the tense-lax contrast itself.

In order now to understand the functional relevance of the patterns of CT activity presented above, it is necessary to link up the muscular activity with associated F0 patterns. In the next section, we first present the F0 data in a very similar way to the CT data, and subsequently discuss the relationships between the two kinds of data.

5.3.2 F0 results

Ensemble averages

As with the EMG data we will first present the complete material in the experiments in terms of ensemble averages. Since the main point of linguistic interest is F0 in the target vowel - and since F0 is of course not defined in the flanking consonants in the voiceless contexts - we have expanded the time-scale somewhat compared with the EMG data, in order to make the patterns in the target vowel easier to identify.

Some methodological notes on ensemble averages with F0 data are necessary:

F0 is a much more readily available signal category than EMG, but calculation of ensemble averages is nevertheless a good deal trickier as the signal is undefined in voiceless passages, and poorly defined at voicing onsets and offset. In addition, the room in which the EMG recordings were made had a good deal of background noise and reverberation (for more recent recordings it has been possible to use a close-talking, noise-cancelling microphone). One particular feature of the YIN algorithm proved very useful in this situation: For every point in the signal an aperiodicity value is calculated, i.e parts of the audio signal with clear F0 structure have low aperiodicity values. This was used to weight the F0 values when computing the ensembles (higher weight for low aperiodicity values), which gave stabler results in the difficult regions. In addition, normal unweighted ensemble averages were computed of the aperiodicity values themselves. This was used as a graphic device, in such a way that the F0 contours are thinner and paler in colour in regions with higher aperiodicity. This helps to make visually more salient those parts of the pitch contour that are probably more important for the hearer (because of remaining problems with reverberation, the pitch contours have been set to missing data during the occlusion phase of the voiceless consonants).

Further prelimary notes on the ensemble average curves: Due to the expanded time-scale the start of C1 is never visible on the left. The end of C2 is never visible on the right for plots with long tense vowels, and not always visible for plots with lax vowels.

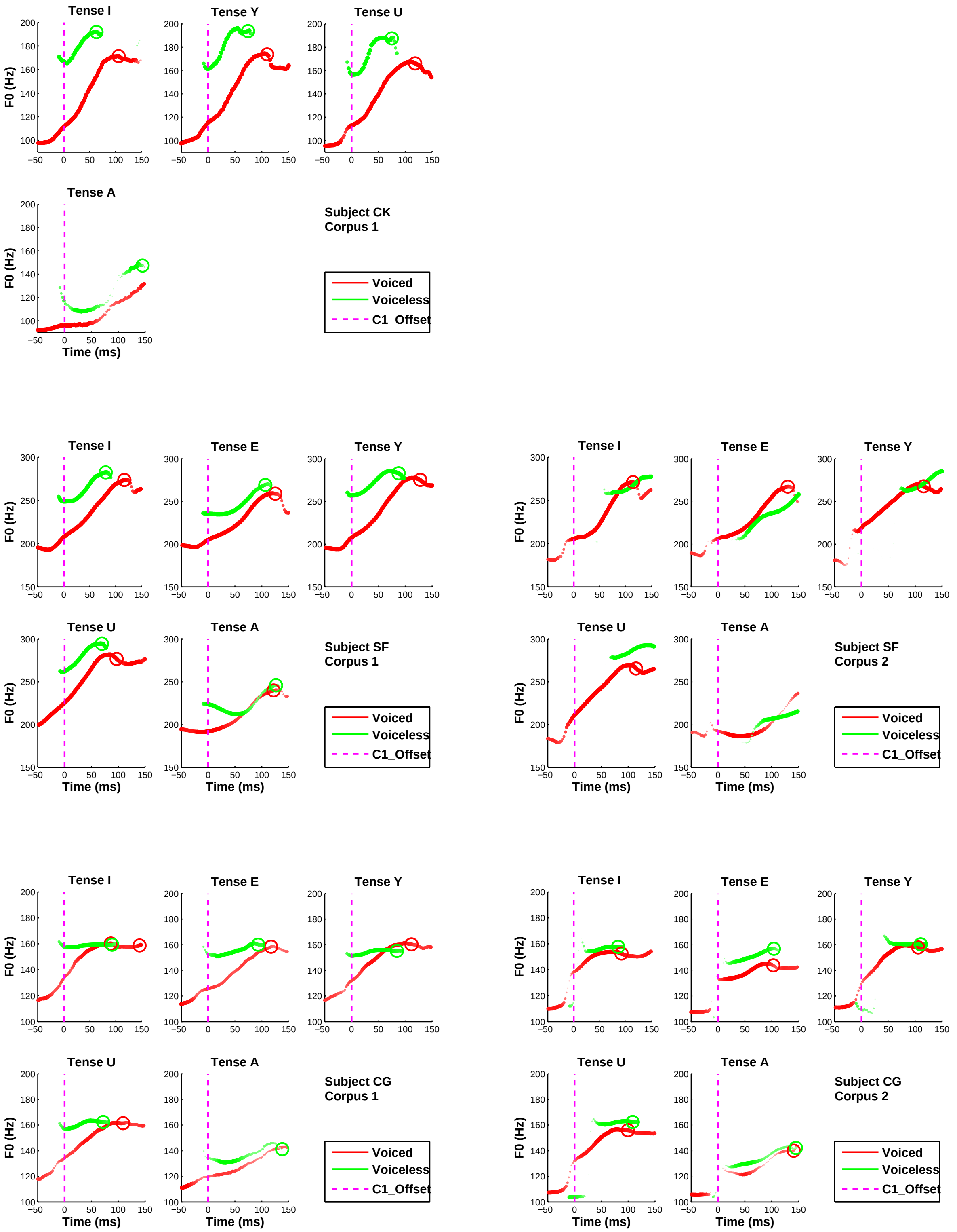


Fig. 5.7: Ensemble averages of F0 in target words with tense vowels. Each panel contrasts corresponding voiced (red) and voiceless (green) consonantal contexts. Corpus 1 on left side, corpus 2 (for speakers SF and CG) on right side. Line-up point is release of C1. Other segment boundaries indicated by circles. Each ensemble normally based on average of 10 repetitions. For Corpus 2 material note that because of the aspiration phase for /**p**/ the F0 contour does not start immediately at the line-up point.

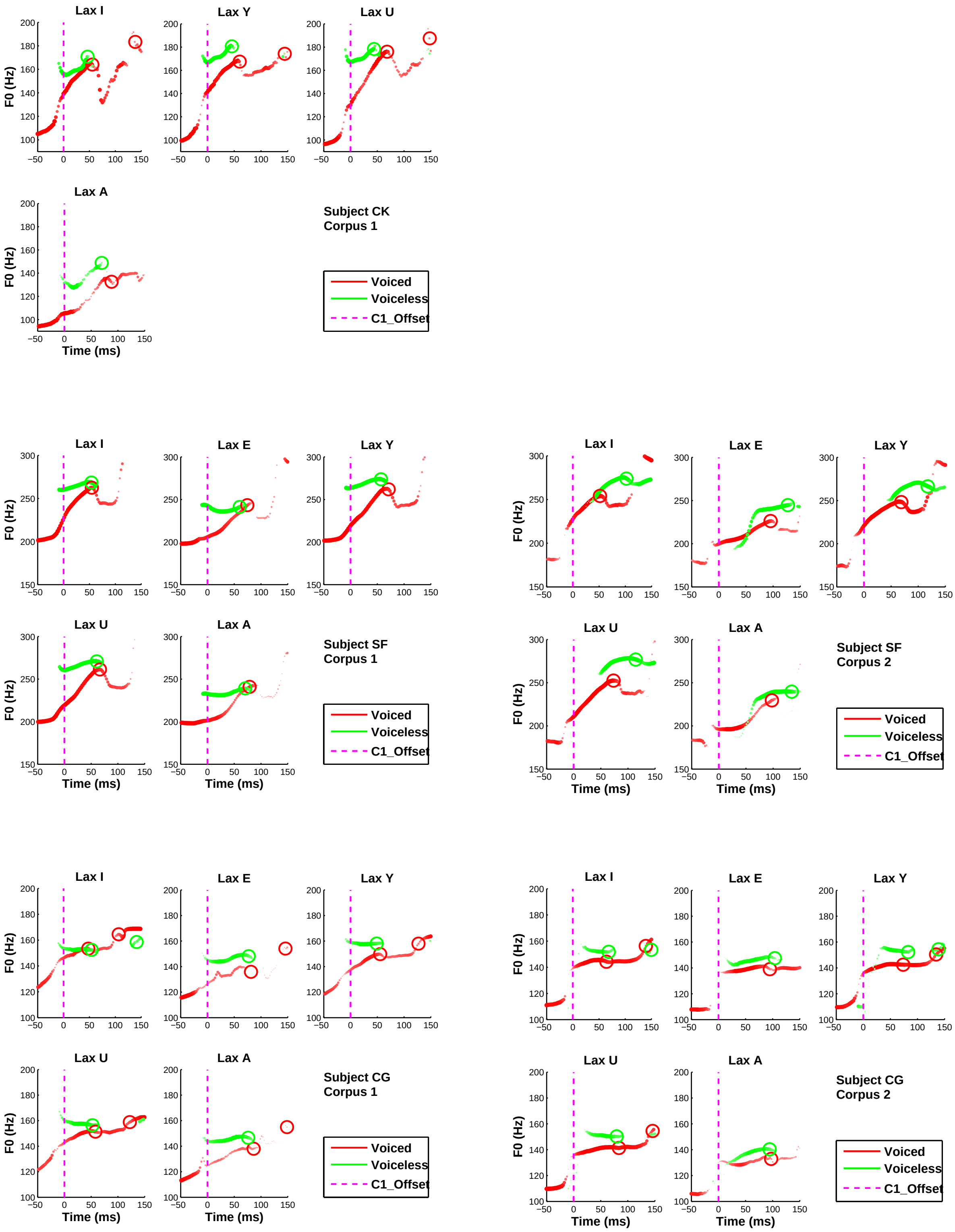


Fig. 5.8: Ensemble averages of F0 in target words with lax vowels. Other details as for figure with tense vowels on opposite page

Comments on ensemble averages

The most obvious feature in these results is also, of course, completely expected: consistently higher F0 following the voiceless consonants. Of more interest with a view to linking the results with EMG activity below is the time course of the differences. Virtually without exception the traces for voiced and voiceless contexts do not merge until the end of the vowel (and in several cases, especially tense vowels for CK and SF (Corpus 1) not even then). The pattern for Corpus 2 (CG and SF) is slightly less clear than for Corpus1. This is particularly the case for SF, fitting in with the EMG data (see EMG ensemble average data in Figs. 5.1 and 5.2). The reason for the less clear pattern is essentially that the voiceless consonant for Corpus 2 is aspirated. If F0 were to be compared at voice onset then one would of course find large differences between the voiced and voiceless contexts. When the data are aligned at consonantal release as here, then by the time voice onset occurs in the aspirated consonants the F0 contours have in fact almost joined up again. This of course does not detract from the fact that the nature of the pitch contour available to the listener differs radically between consonant contexts for Corpus 2 just as it does for Corpus 1.

Statistical results for the vowel segment

As with the EMG we can give some idea of the robustness of the effects by extracting a single value to characterize each vowel token, and using this in statistical analysis. The corresponding overview of means and standard errors for each linguistic category is given in the following figure. Because of the risk of occasional outliers at vowel onset and offset we used the median value of the vowel (rather than mean as in the EMG) as the characteristic value for each token. Since F0 contours are usually monotonically rising (or fairly flat) this 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 these are clearly greater towards vowel onset. The figures indeed indicate a very clear separation of the lines corresponding to voiced and voiceless contexts, and in a GLM-analysis (arranged as for the EMG statistical analysis) the main effect of VOICE was significant at $p<0.001$ in all cases[21].

Regarding more detailed effects in the data, the ensemble averages and the means show a tendency towards smaller voiced-voiceless differences in /**a**/ (presumably related to the lower pitch of /**a**/). In statistical terms this should correspond to an interaction between the VOICE and VOWEL effect, but this only turns out to be significant ($p<0.05$) for SF.

A perhaps more important case of an interaction was that between VOICE and TENSE. For CK (interaction significant at $p<0.01$) this parallels the EMG results rather precisely (see above Figs. 5.1, 5.2 (ensembles) and 5.4 (means); SF also had a weakly significant interaction for the EMG data but not here for F0). It is quite striking to observe in the F0 ensembles (comparing Figs. 5.7 and 5.8) how the onset of the lax vowel contour following voiced C1 is pulled towards the the contour for the voiceless consonant, when compared with the situation for tense vowels. This

[21]One could argue for CG Corpus 2 that there is a bias in favour of a significant result, since for voiceless contexts F0 is only defined for a portion rather late in an overall rising intonation contour. However, CG is actually the speaker with the flattest contours, so in practice this bias is unlikely to be strong.

interaction will crop up again in the next main section of the results when we consider the patterns from the point of view of the tense-lax distinction.

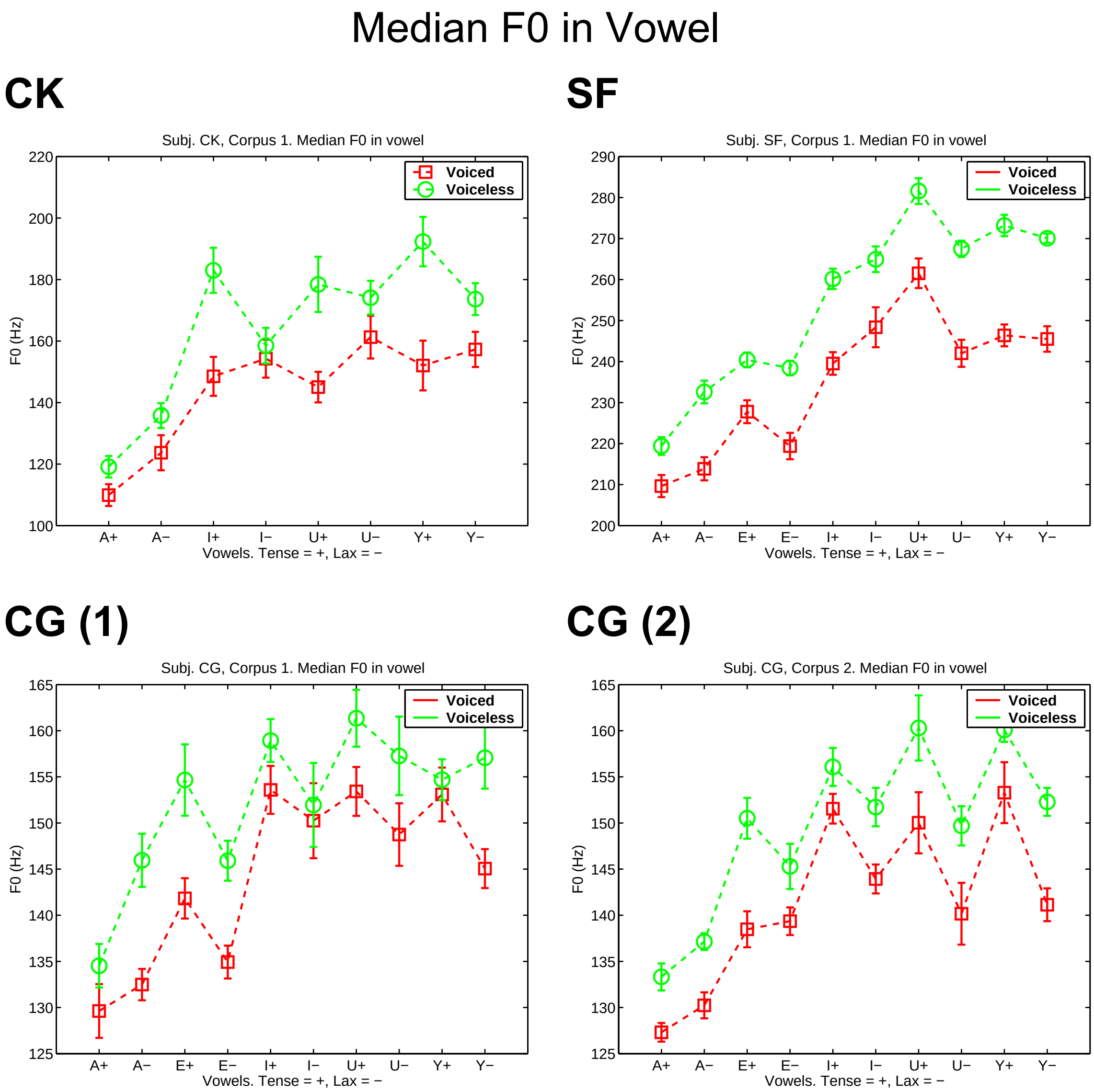


Fig. 5.9: Mean and standard error of median F0 in vowel segment. See corresponding EMG figures above for other details.

5.3.3 Linking EMG and F0

In the discussion of the EMG results a number of questions were raised that required parallel consideration of EMG and F0. The question of primary linguistic interest is: To what extent can the absolutely robust F0 differences on the vowel be explained as a simple side-effect of consonant production, or conversely, to what extent do speakers show evidence of active differences in muscular adjustments on the vowel itself? Clear evidence of a voiced-voiceless difference in activity on the vowel itself was found in two cases: the tense vowels of CK and SF. Thus this does appear to be a strategy that is potentially available to speakers. In the other cases, however, i.e all of CG’s material and the lax vowels of CK and SF, the differences were marginal to clearly negligible. Nevertheless, as already emphasized, F0 differences were robustly present throughout the material. (The only apparent exception to this comes from the lax vowels of CK, where we have just noted that voice-voiceless F0 differences reduce in saliency in parallel with his EMG results. However, although reduced, his F0 results do show a consistent tendency in the expected direction.) Thus, we appear to have a situation in which the F0 differences on the vowel are clearer than the EMG activity during the vowel would lead one to expect. We will give a more formal character to this observation in the course of the next sections on intrinsic pitch (both with regard to the specific question of the tense-lax distinction, as well as the more general question of high vs. low vowels) where explicit analysis of the properties of the regression equation linking EMG and F0 are crucial for the interpretation of the results (see discussion of Fig. 5.21 on p. 85f, and in more condensed form Fig. 5.25 on p. 94). This leads us back to the first question posed above, namely F0 differences on the vowel as a side-effect of consonant production. This investigation has certainly added significantly to the body of results suggesting robust increases in CT activity related to voiceless consonant production. But, of course, this does not in itself explain by what mechanism higher F0 on the following *vowel* then results. And in fact previous investigations in the literature have also not been very explicit on this point. After accumulating in the following sections the relevant information from the F0 vs. EMG regression analysis, we will come back to this point in the concluding discussion and provide some suggestions for the mechanisms that could be assumed here.

Another area where the regression analysis will also prove useful is with regard to the interaction between the effect of VOICE and TENSE which was in evidence for the EMG data of all of the speakers in at least one of the segments, making the VOICE effect more obvious in the context of tense vowels compared to lax vowels. This was an unexpected finding that has interesting implications for understanding the robustness of the main effect at issue, i.e here VOICE. However, as it will logically also play a role in the presentation of the results with respect to the next main effect, namely TENSE, we will again reserve further suggestions as to a possible source of the effect to the concluding discussion.

5.4 Results for the tense-lax opposition

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. This would speak firstly - on the linguistic side - for a prosodic difference between tense and lax vowels, and, secondly - on the biomechanical side - for the viability of tongue-pull explanations of intrinsic pitch. Substantiating the second point logically requires a second stage to the analysis, namely showing that there is a different relationship between CT and F0 for tense vs. lax vowels.

To show the results we will proceed in essentially the same way as for the consonant voicing distinction in the previous section, i.e set the scene with ensemble averages of each linguistic category, then move on to statistical analysis and the link with F0. In fact, we are, of course, showing exactly the same data a second time, but simply in a different arrangement. This may involve a certain amount of redundancy, but nevertheless seemed preferable for clarity of exposition with regard to the different linguistic oppositions investigated.

5.4.1 EMG results

Ensemble averages

Accordingly, the next double page of ensemble average illustrations shows a given tense-lax opposition in each panel, with the vowels in voiced consonant contexts on the first page, and voiceless context on the second page. Once again, the alignment point (dashed line at t=0 on the x-axis) corresponds to the release of C1 (i.e the onset of the target vowel). On each trace, the circle symbols to the left of the line-up point correspond to the onset of C1. The symbols on the tense and lax trace should normally be at virtually the same location because C1 is of course identical for the two cases. Following the line-up point the circle symbols are more spread out in time because the tense and lax vowels generally differ in length. The most frequent temporal order encountered is offset of lax vowel (green symbol), offset of tense vowel (red symbol), offset of C2 after lax vowel (green symbol), and offset of C2 after tense vowel (red symbol).

If one looks now at the first page of ensemble averages (vowels in voiced consonant context) and focusses in each panel on the region between the line-up point and the offset of the lax vowel a clear tendency for higher CT activity in the lax vowels will be observed. This is consistent and clear for CK and CG (1), less clear for CG (2), clear for SF (1) in 4 cases out of five but not present in one case (/**u**/), consistent but weak for SF (2). Looking at the voiceless contexts on the following page the same general tendency is apparent but is clearly weaker for all speakers (except perhaps SF (2)). This provides a counterpart to the interaction effect already discussed in the previous section: there we noted a tendency for clearer consonant voicing differences in the context of tense vowels, here we have a tendency for clearer tense-lax differences in the context of voiced consonants.

Before turning to the descriptive and test statistics for these patterns, a word on the temporal extent of the tense-lax differences: Any anticipation of the tense-lax difference in C1 (i.e to the left of the line-up point) would appear to be very weak (this will be tested below). Carry-over of tense-lax differences into C2 turns out to be an interesting special case which will be discussed in more detail below, but is difficult to judge from the ensemble averages, since due to the different vowel lengths for tense-lax pairs the onset of C2 is not time-aligned.

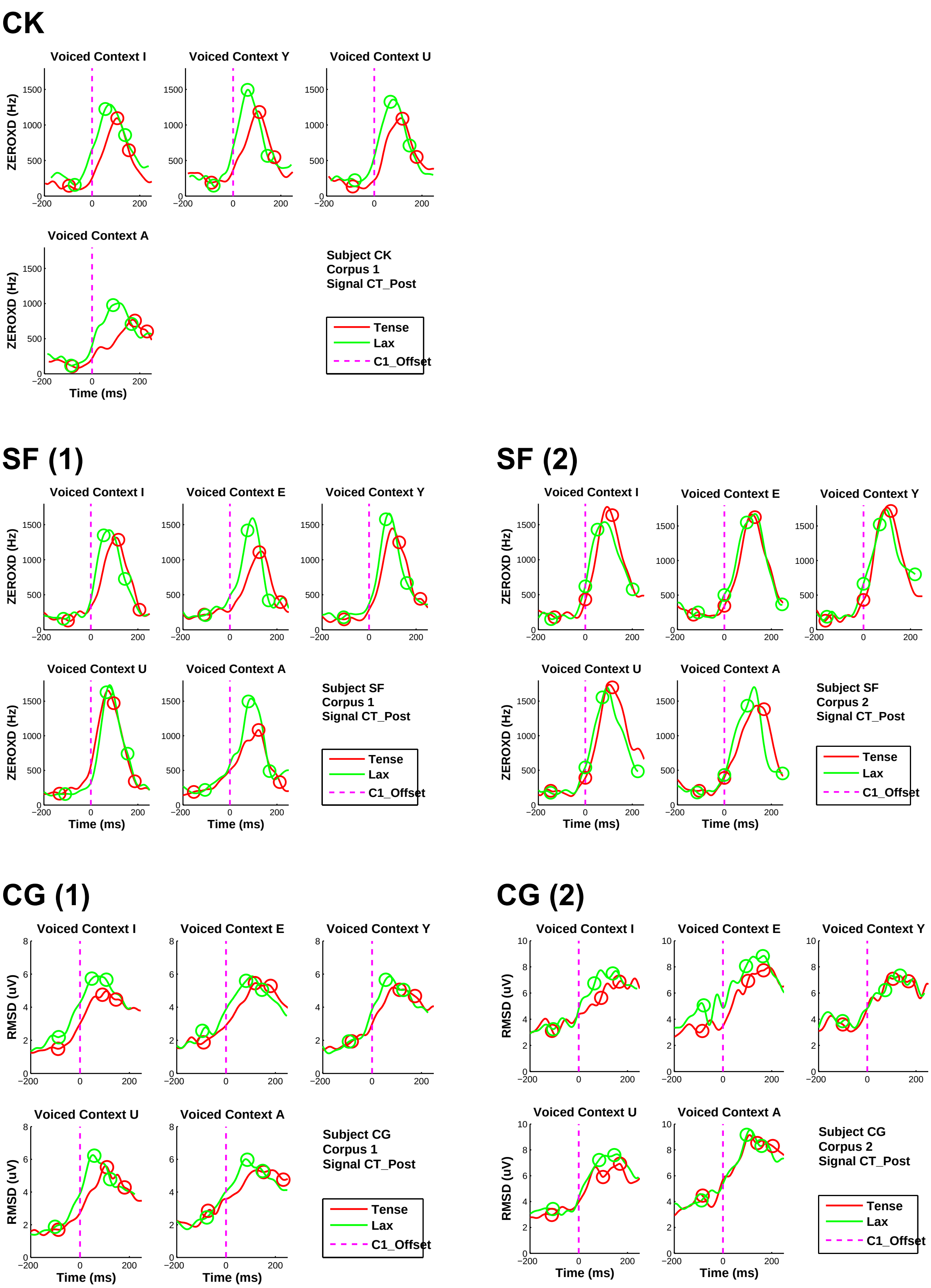


Fig. 5.10: Ensemble averages of CT activity in target words with voiced consonantal context. Each panel contrasts corresponding tense (red) and lax (green) vowels. Corpus 1 on left side, corpus 2 (for speakers SF and CG) on right side. Line-up point is release of C1. Other segment boundaries indicated by circles. Each ensemble normally based on average of 10 repetitions.

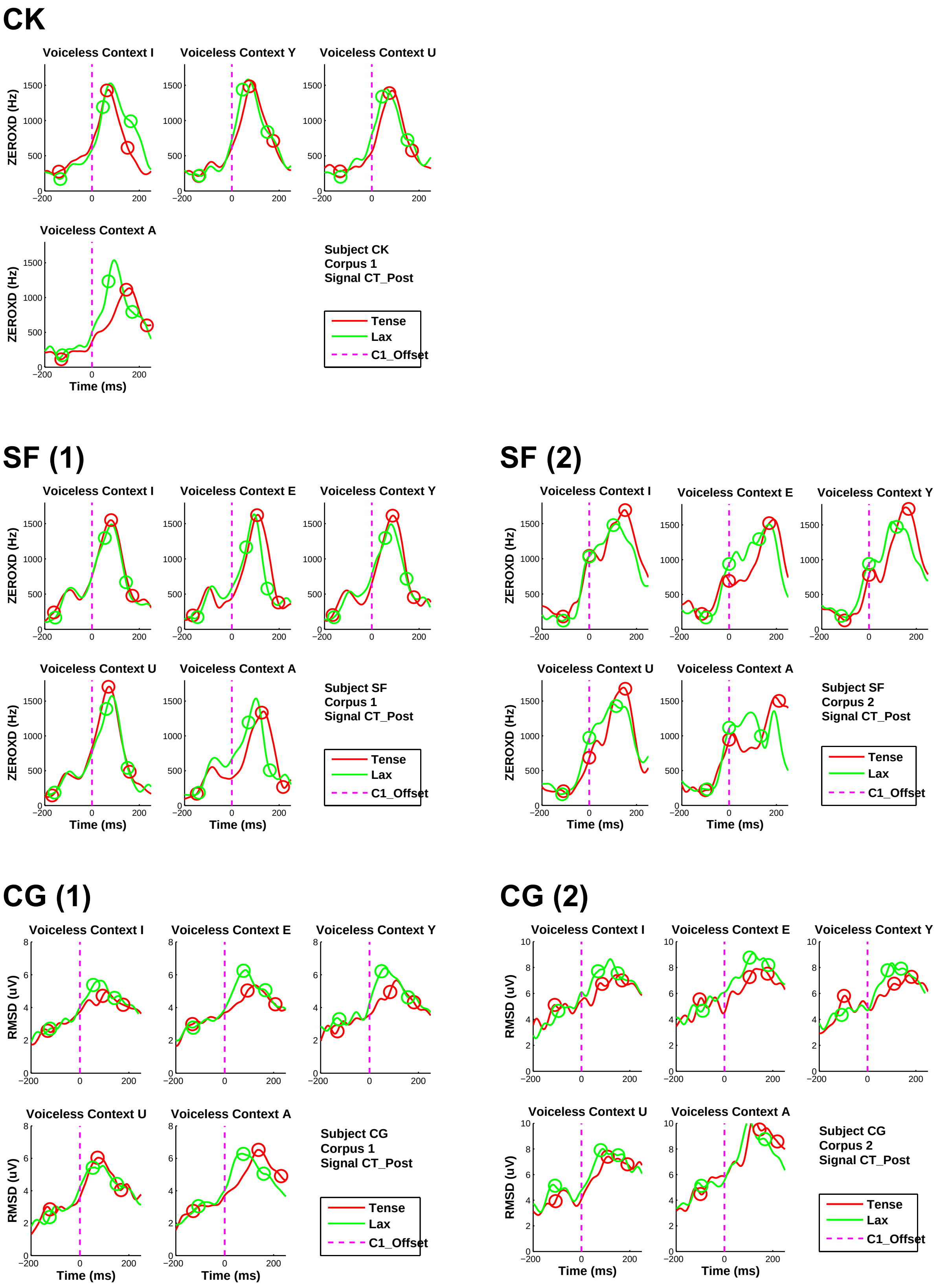


Fig. 5.11: Ensemble averages of CT activity in target words with voiceless consonantal context. Other details as for words with voiced consonantal context on opposite page.

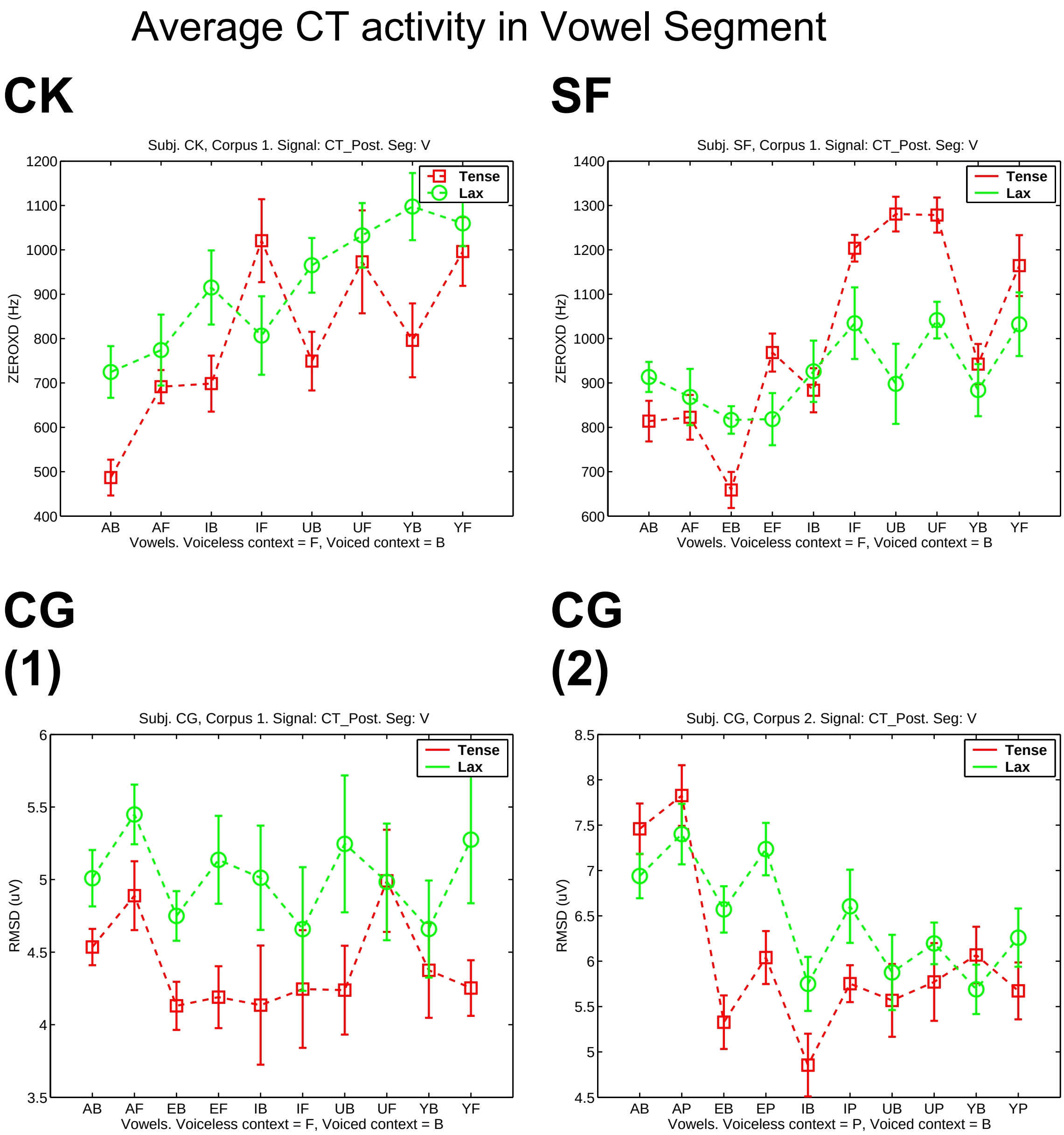


Fig. 5.12: Mean (over 10 repetitions) and standard error of average CT activity in each vowel segment. Separate panels for each speaker (and separate panels for each corpus for speaker CG). The two-letter upper-case codes on the abscissa indicate vowel category (first letter) and consonant context (second letter). This letter corresponds to the identity of C1. Thus ‘B’ always indicates voiced context, while ‘F’ and ‘P’ indicated voiceless context for corpus 1 and corpus 2, respectively.

Statistical results for each segment

<u>Vowel segment</u>

We start the presentation of the summary results with the vowel segment, as this is clearly the crucial segment for the tense-lax distinction. The above figure shows the mean (and standard error) of CT activity over the acoustically defined vowel segment (vowel onset defined as before as the timepoint of consonant release, not voice onset). The voiced and voiceless consonant contexts are located at adjacent positions on the x-axis. As with the corresponding results in the previous section on consonant voicing, a zig-zag pattern for one of the two lines is thus an optical indication of a VOICE*TENSE interaction.

These summary results confirm the trends derived from inspection of the ensemble averages.

For CK the main effect of TENSE was significant at $p<0.01$, but with the VOICE*TENSE interaction also significant at $p<0.01$ (the weaker nature of the tense-lax difference in the voiceless context is readily observable (and even one reversal of the general trend on /**i**/)).

For speaker CG, as before, we discuss here only the results for the posterior insertion.

For CG (1) the main effect of TENSE was significant at $p<0.001$. Although there is one case of a voiceless context (vowel /**u**/) where the tense-lax difference seems to disappear, the VOICE*TENSE interaction was not in fact significant. For CG (2), although the tense-lax difference seems less clear than in CG (1) the main effect of TENSE was still significant at $p<0.01$ (and there was no VOICE*TENSE interaction).

The least clear case is for speaker SF. There was actually a significant effect ($p<0.01$) contrary to the hypothesized direction, but above all a very strong VOWEL*TENSE interaction ($p<0.001$), i.e tense-lax effects in both directions can be found, depending on the vowel category. Referring back to the ensemble averages makes clear, however, that the results for this speaker are not as strongly counter to the general trend as might appear at first sight. Particularly in her case, the procedure of comparing average activity over the acoustically defined vowel segments in effect results in a bias in favour of higher activity in the tense vowels. This is because of the overall rising intonation pattern throughout the vowel. Because the lax vowels are shorter, the peak in CT activity is frequently not reached until after the end of the vowel, whereas the tense vowels tend to include more or less the whole of the rising flank of CT activity. Thus even if lax vowels have higher activity than tense when matched point for point over the course of the lax vowel, then it is still possible for higher average activity on the tense vowels to emerge (a figure illustrating this effect is given below in the section on the analysis of the C2 segment, where this effect is again of particular relevance for the interpretation of the observed patterns). While it might have been possible to reanalyze the data using for the averaging procedure a fixed window of appropriate length, we have preferred to leave the results as they stand, since they can then be assumed to be conservative with respect to the hypothesis of stronger muscle activity in the lax vowels.

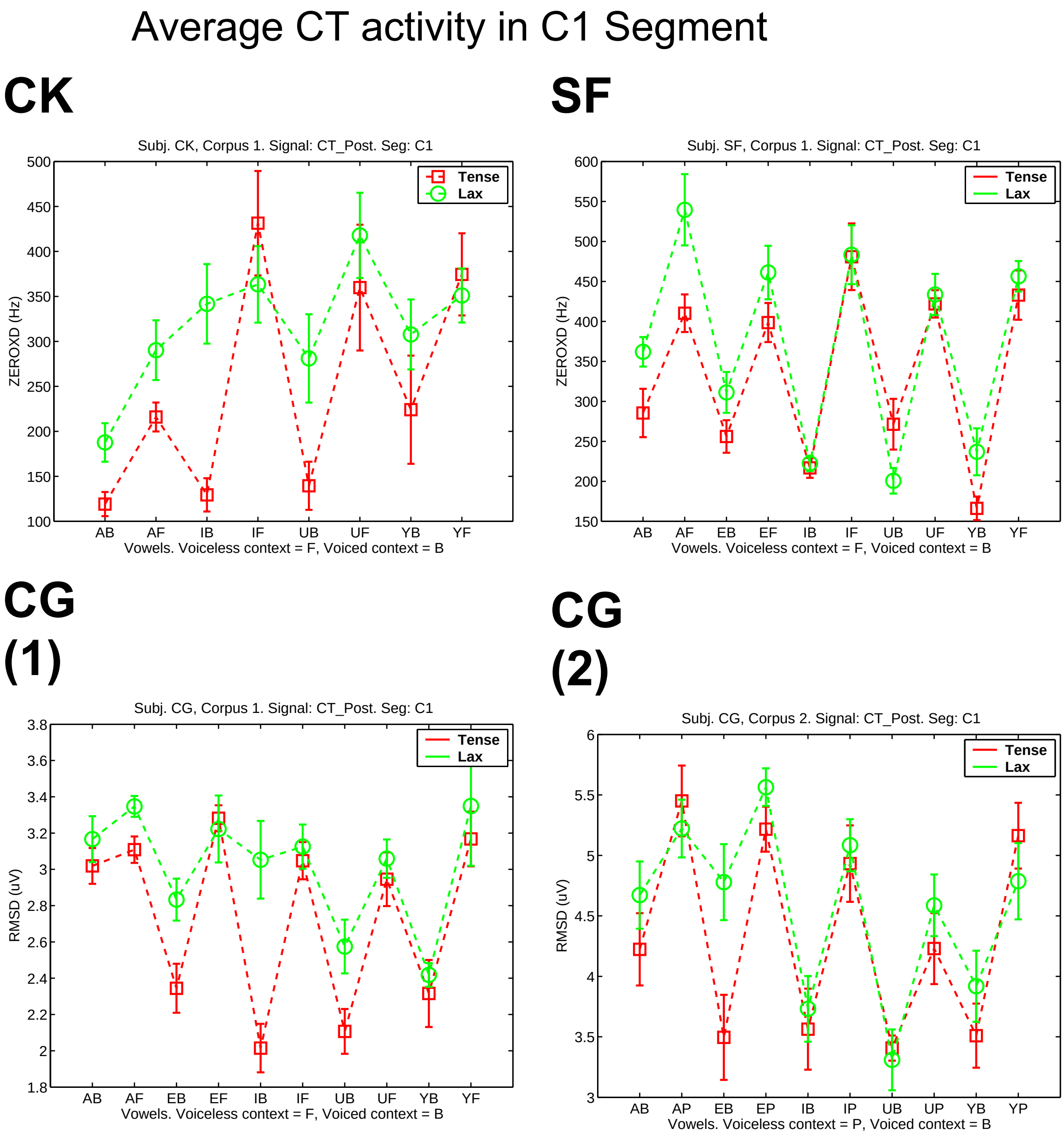


Fig. 5.13: Mean (over 10 repetitions) and standard error of average CT activity in each C1 segment. See previous figure (vowel segment) for further details.

<u>C1 segment</u>

Summary results for the C1 segment will be looked at only briefly.

It will be seen that to the extent that a trend is apparent it is in the direction of more activity in the lax vowel contexts. The very zig-zag nature of all the patterns indicate, not surprisingly, that consonant voicing has a stronger effect than vowel tenseness on this segment.

Nevertheless for CK the effect of TENSE is still significant at $p<0.01$, again with a significant VOICE*TENSE interaction at $p<0.01$. CG(1) again showed a main effect of TENSE that was significant at $p<0.001$ just as in the vowel, but now in the consonant the VOICE*TENSE interaction was significant ($p<0.05$). For CG (2) the main effect of TENSE was also still significant, though only at $p<0.05$ (no interactions). Interestingly, the result for SF was significant at $p<0.01$ in the hypothesized direction. It is not easy to see in the figure as the TENSE effect is swamped visually by the VOICE effect (the VOICE*TENSE interaction was not itself significant). The interpretability of the main effect of TENSE is mitigated somewhat by a significant VOWEL*TENSE interaction (p<0.01), but nevertheless I feel helps to support the interpretation that the apparently aberrant result for the vowel segment is due to the definition of the analysis window.

Overall, although not very salient in the ensemble averages, there is thus some indication that the tense-lax difference in CT activity is already starting to become evident in the prevocalic consonant.

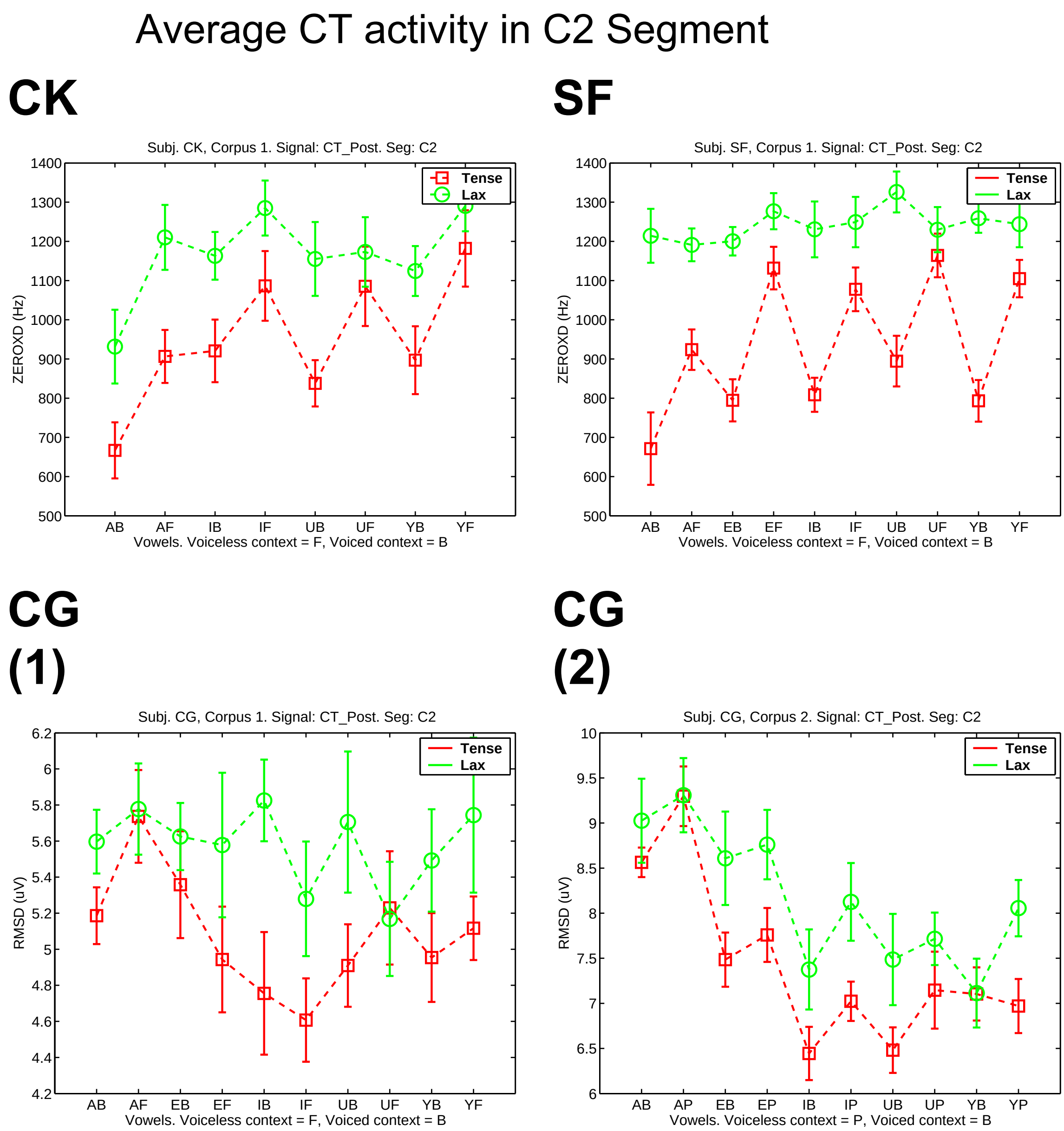


Fig. 5.14: Mean (over 10 repetitions) and standard error of average CT activity in each C2 segment. See corresponding figure for vowel segment for further details.

The C2 segment

This segment turned out to be a special case, in the sense that it enables an interesting link to be made back to the first section of the results where CT effects in consonant voicing were the focus.

We had no particular hypothesis for this segment; the intention was simply to find out to what extent tense-lax differences carry over to the following consonant.

It turns out that this was the segment that actually showed the clearest difference between tense and lax vowels. This is very obvious in the figure. The main effect of TENSE was significant at $p<0.001$ in all cases (the only interaction was a strong VOICE * TENSE effect ($p<0.001$) for SF, which is evident from the zig-zag nature of the line joining the tense vowels).

It would probably be misguided, however, to simply take this as further evidence for a tense-lax difference in CT activity. Most likely there is some carryover effect of the higher CT activity in the lax vowels, but this is only part of the story. An important component is undoubtedly the timing of the segments relative to CT activity for the overall intonation pattern of the utterance. Here we have the counterpart to the effect that the analysis of the vowel segment incorporated a bias towards higher activity in the tense vowels, i.e against the hypothesis of more activity in the lax vowels. In the C2 segment the bias now works in the opposite direction: because the lax vowels tend to end before the overall CT peak, the postvocalic consonant tends to include the peak region, whereas the tense vowels normally extend right up to the vicinity of the peak so that the postvocalic consonant is then located to a greater extent on the falling flank of C2 activity. To make this point easier to see, we reproduce here at a larger scale one of the panels from the ensemble averages.

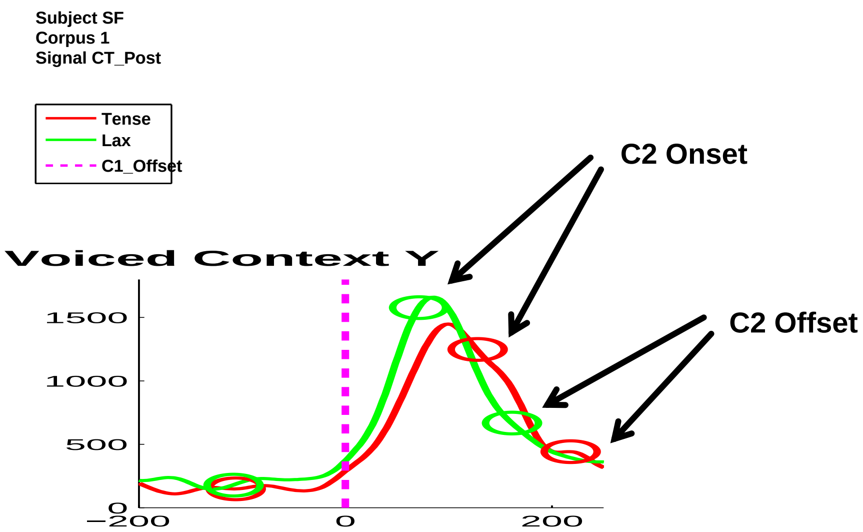


Fig. 5.15: Detailed view of segment boundaries with respect to typical ensemble-averaged CT activity; speaker SF /**y**ː/ (red) and /**ʏ**/ (green) in voiced context.

At this point one could argue that the result is thus largely an effect of the overall utterance prosody and not worth considering further. However, let us recall the key finding from the first results section that higher CT activity is associated with voiceless consonants. We now observe here higher CT activity in the consonants following lax vowels, and it is a well-known fact of German that voiced obstruents following lax vowels are dispreferred. This could provide a physiologically motivated reason why this has remained a salient feature of German sound structure whatever the historical roots of the pattern may have been (see Becker, 1998, p.50 for further background), and avoids the necessity for simply stipulating that there is a stronger affinity between lax vowels and strong consonants.

One might object that the argument depends here on the assumption of a very specific utterance prosody, but the intonation contour in our utterances is actually a very typical one precisely for the case that words are being uttered clearly (for example cited, or demonstrated).

The results also tie in quite neatly with findings of Ladd (Ladd et al., 2000; Atterer & Ladd, 2004; see also Pape & Mooshammer, 2004) regarding the timing of pre-tonic rises. His results suggest that the pre-tonic rise is typically aligned later in German than in English, and later in southern German than in northern German. Since it is the late timing of the peak that leads to this effect it is interesting that rareness of voiced consonants after short vowels is more obvious in German than English, and those that are found in German tend to have a northern "flavour", e.g "Ebbe" (often origin in Niederdeutsch, cf. Becker, p. 50).

5.4.2 Relations between EMG and F0

The EMG results thus provide some evidence that speakers actively modify their laryngeal behaviour as part of the tense-lax distinction. As discussed above, this also opens the way to resolving the apparent puzzle of intrinsic pitch in German. If muscle activity is increased on the lax vowels then the typical pattern in German of similar F0 for tense-lax pairs need not contradict the tongue-pull hypothesis. However, in order to make this contention more watertight it is necessary to show that the relationship between EMG and F0 differs between the two vowel classes. For example, if our speakers simply all had higher F0 on the lax vowels, then F0 might still simply follow CT activity leaving no room for a tongue-pull effect. If, however, it can be shown that at any given level of CT F0 is lower for the lax vowels then this would imply the presence of a non-muscular influence on F0.

To provide the background for examining this relationship we first show ensemble averages of F0 arranged (as in the EMG ensembles above) with tense-lax pairs juxtaposed (these are the next double page of figures; they are immediately followed by a further double page of figures that are explained below).

Ensemble averaged F0

These ensembles essentially confirm the well-established finding for German that there is no clear distinction in F0 between tense and lax vowel pairs. Often the F0-contour for lax vowels is slightly higher (e.g CK, voiced contexts), but there are also a fair number of cases where it is slightly lower (especially speaker CG). The crucial point is that, as in previous investigations, there is absolutely no indication that for example lax /ɪ/ is lower than tense /eː/ as would follow from tongue height. Accordingly statistical analysis of the main effect of TENSE is of only limited interest. If desired, reference can be made back to Fig. 5.9 (on p. 67 in the section on the results with respect to consonant voicing) for median F0 in each vowel category[22].

[22]though note that once again because of the rising intonation contour there is a bias towards higher F0 on the tense vowels.

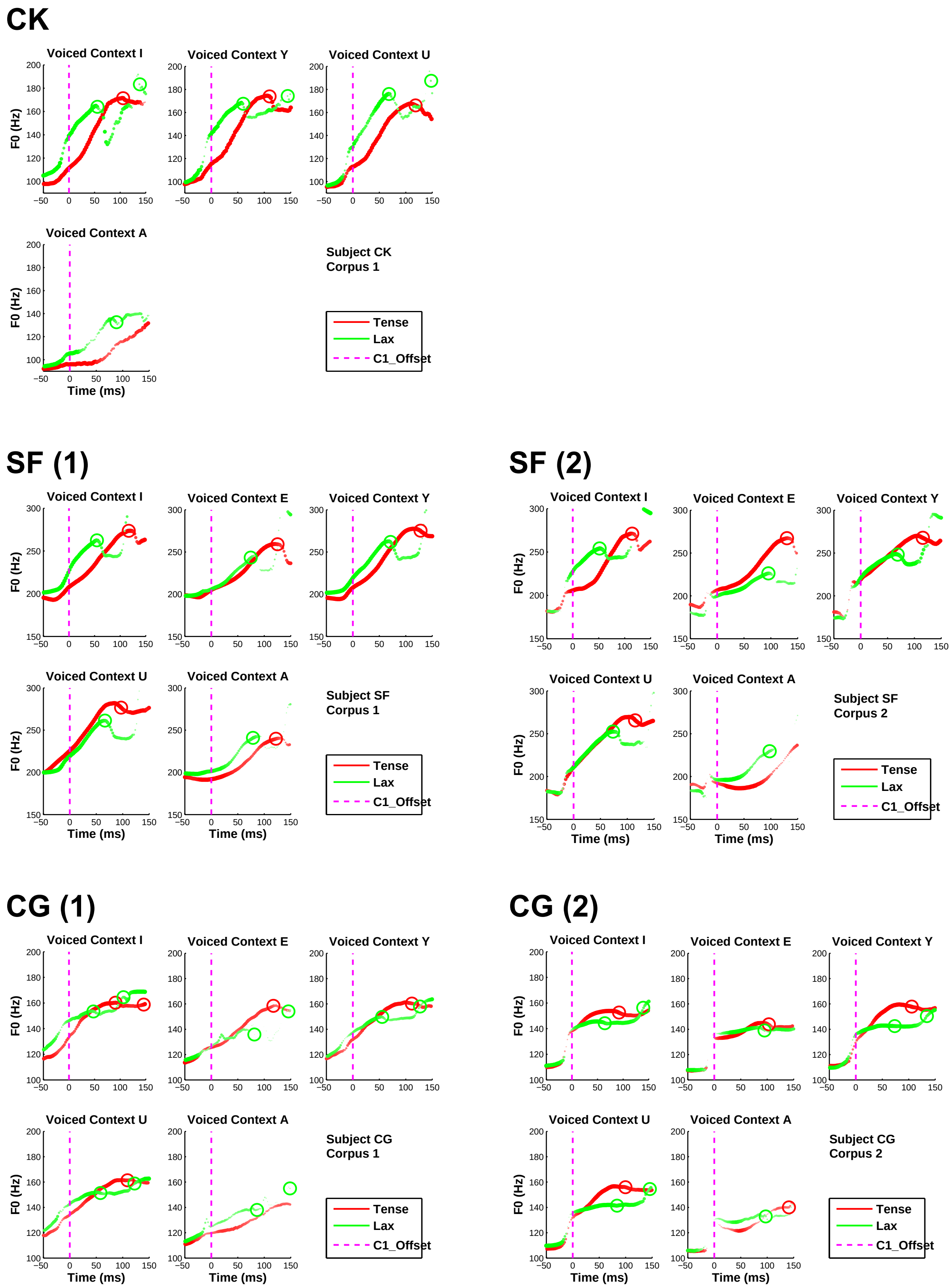


Fig. 5.16: Ensemble averages of F0 contours for tense (red) vs. lax (green) vowel pairs. Voiced consonantal context. Other details as for corresponding figures of EMG activity.

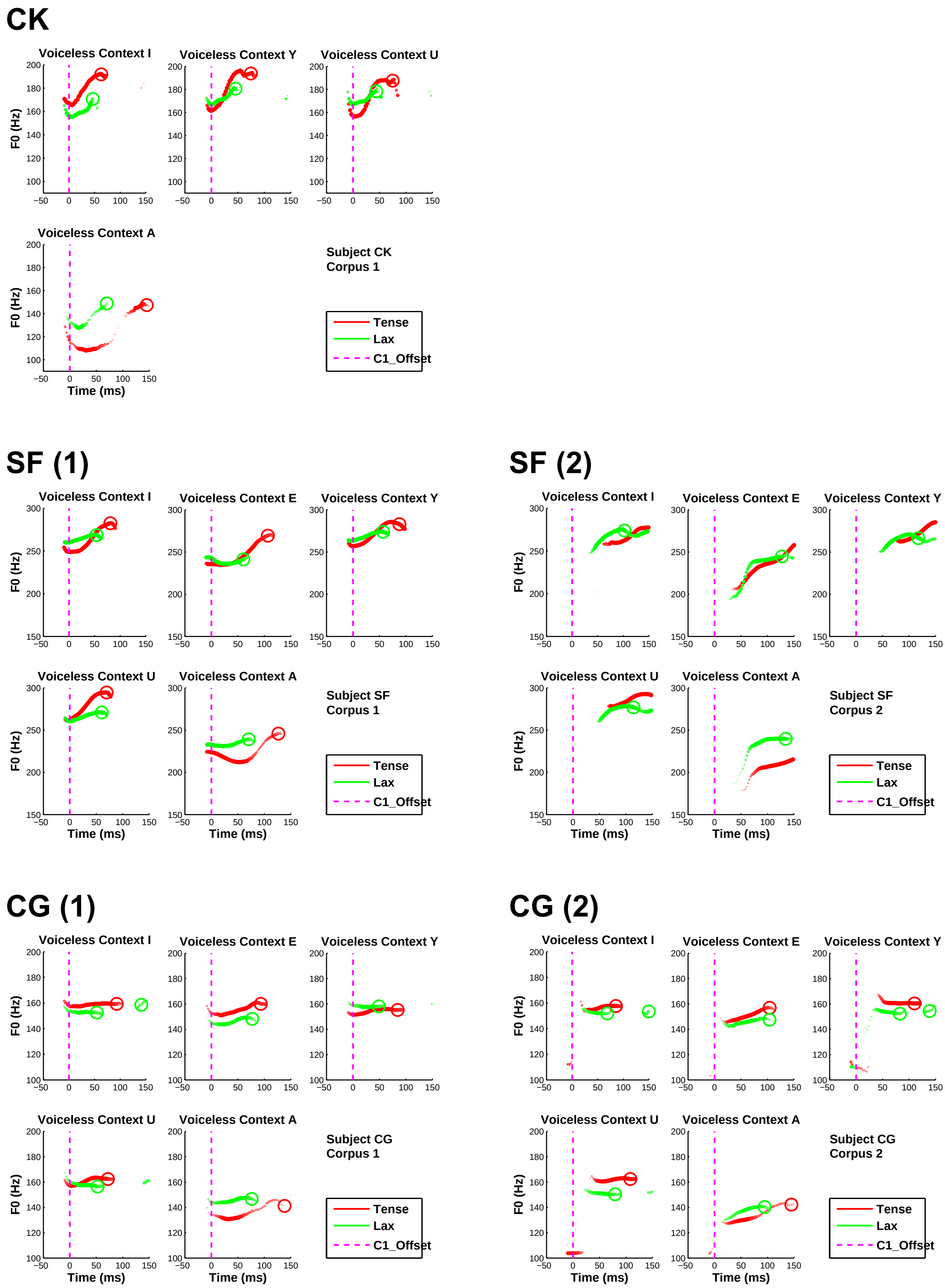


Fig. 5.17: Ensemble averages of F0 contours for tense (red) vs. lax (green) vowel pairs. Voiceless consonantal context. Other details as for corresponding figures of EMG activity.

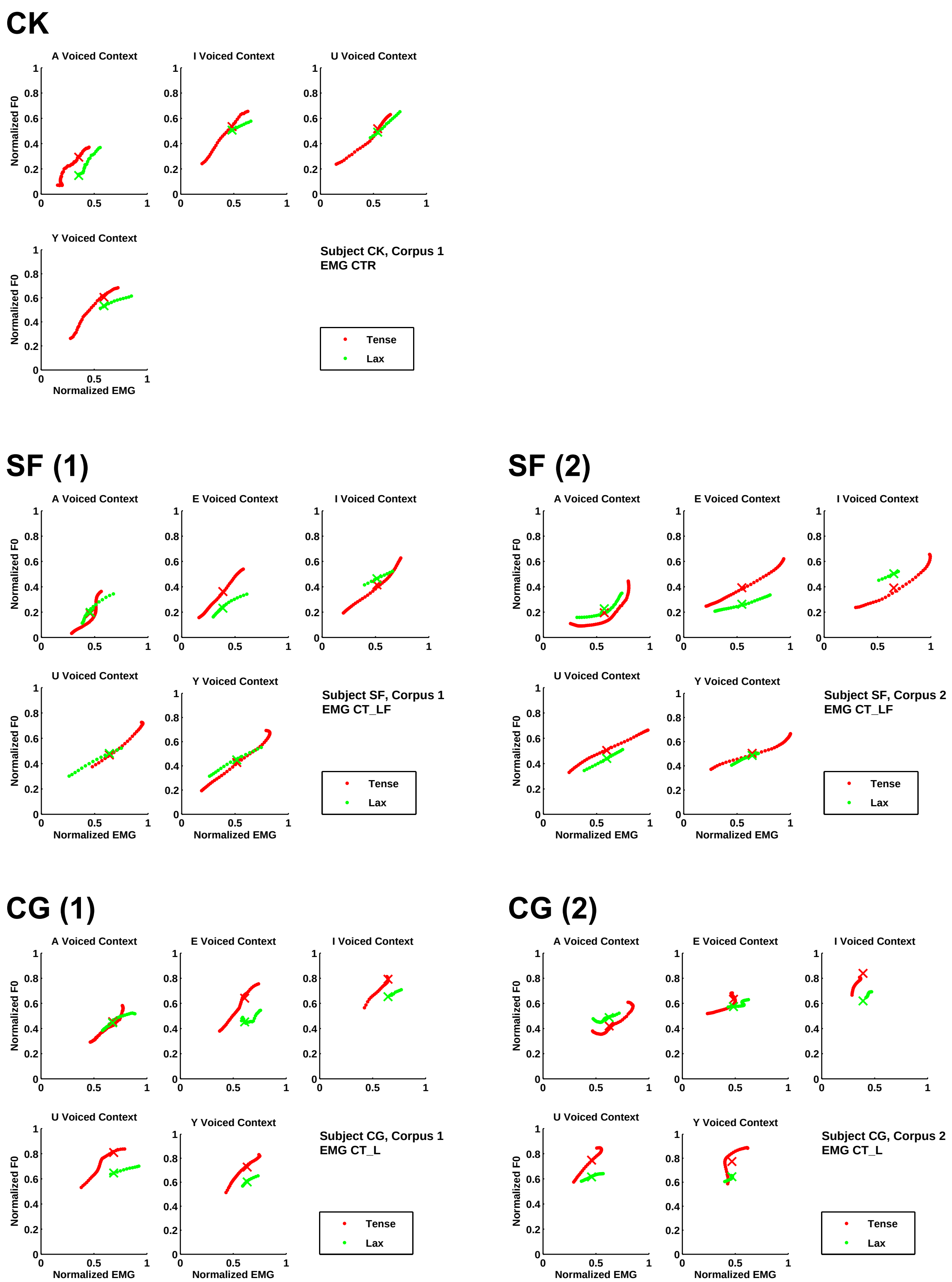


Fig. 5.18: Relationship between ensemble-averaged F0 and EMG over the vowel segment. Voiced consonantal context. Trajectories for tense vowels in red, for lax vowels in green. Both F0 and EMG data have been normalized to a range of 0 to 1. The cross symbols indicate the F0 values used to compare tense and lax vowels at a common EMG level in the regression analysis below (see text for details).

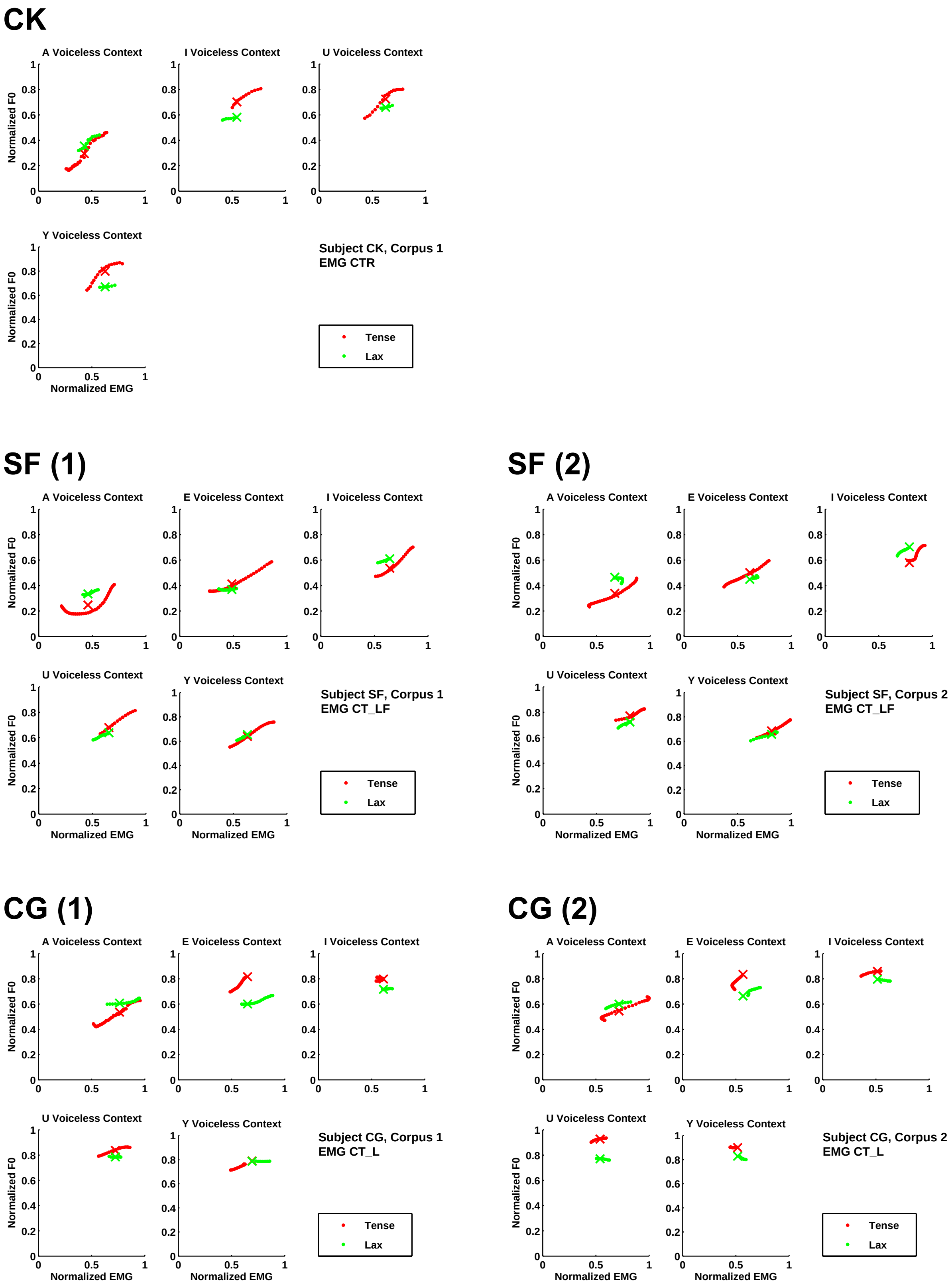


Fig. 5.19: Relationship between ensemble-averaged F0 and EMG over the vowel segment. Voiceless consonantal context. Trajectories for tense vowels in red, for lax vowels in green. Both F0 and EMG data have been normalized to a range of 0 to 1.

F0 vs. EMG: (1) Inspection of scatter plots of ensemble-averaged data

On this background, we are now in a position to look explicitly at the relationship between EMG and F0. To do so, we will examine only ensemble-averaged data from the target vowel (a briefer analysis below is based on individual vowel tokens). Since there was generally quite strong F0 movement through this target segment, it is basically well suited for carrying out a regression analysis. Nevertheless, we eliminated 15ms at the start and end of each vowel to avoid F0 values that might be perturbed by transitory aerodynamic effects (in Corpus 2 15ms from the onset of voicing after the aspirated consonant was eliminated). If F0 is then displayed in scatter plots as a function of muscle activity, then it is possible to gain a first impression of whether tense and lax data points lie on the same regression line (see previous double page of figures). If it is not possible to reject the null-hypothesis that tense and lax data lie on the same regression line, then we cannot assume that there is a biomechanical effect on F0. More specifically we would expect that the regression line for lax data is either displaced globally downwards and to the right, with respect to the tense data, or at least shows a shallower slope (i.e either a given level of F0 requires more muscular activity in the lax vowels, or a given change in F0 requires more change in activity). An additional specific hypothesis is that these effects should not be apparent for the vowel category /**a**/. Up to now we have treated /**ɑː**/ vs. /**a**/ as comparable to the other tense-lax pairs. But there are of course important differences. Unlike the other pairs, this pair is often considered to show negligible differences in vowel quality. While this may not be strictly true (see the articulatory representation of the vowel space in Fig. 2.6) we can be sure that long tense /**ɑː**/ will not show tongue-root advancement compared to lax /**a**/, whereas for all the other pairs shown here this will definitely be the case. Thus, under the tongue-pull hypothesis /**a**/ should either show negligible differences in the F0-EMG relationship, or possibly even a reversal compared to the other tense-lax pairs if indeed tense /a:/ does have stronger pharyngeal constriction. (Note that these figures use amplitude-normalized versions of the F0 and EMG ensemble averages, i.e the range found for each speaker over the complete material is converted to a range of [0 1])

Explicit analysis of the regression lines will be given further below in another set of figures. Looking first at the patterns in the scatter plots, the immediate impression is that subject CG, in particular, shows quite a clear trend in the hypothesized direction. This applies both to the basic hypothesis that lax vowels should show lower F0 at comparable EMG values, as well as to the detail hypothesis that this trend should be absent or reversed for /**a**/ (for the voiced contexts of CG (1) (i.e left side of Fig. 5.18) the lines are virtually indistinguishable (the /**a**/ panel is always the top left panel in each block); in his other three /**a**/ panels (right side of Fig. 5.18 and left and right side of Fig. 5.19) the lax case appears shifted to higher F0 at comparable EMG). For CK, the general trend is also in the hypothesized direction, the main exception being that /**a**/ in voiced context does not show the expected absence or reversal. The least clear results are for speaker SF. While the special case of /**a**/ conforms largely to expectations (actually quite similar to CG's results for this vowel), the other vowels (which should behave as a group) have rather inhomogeneous behaviour, from supportive (generally the case for /**e**/) via unclear (/**u**/, /**y**/) to contrary to the hypothesis (/**i**/).

F0 vs EMG: (2) Regression analysis of ensemble-averaged data

Following these qualitative remarks we next attempt to assess the regressions more quantitatively. The relationship between EMG and F0 was approximated by a simple straight-line fit. There are a couple of obvious cases where a fit with a higher-order polynomial would have been more accurate (e.g SF (1), tense /**ɑː**/ in voiceless context), but as many of the shorter vowels would have been difficult to fit robustly at a higher order, it was decided to keep a linear fit for all vowels. The first parameter based on this analysis is simply the gradient of the straight line. This is shown on the first of the two pages of figures below together with the 99% confidence intervals. In a similar way, one could have compared the intercept term in the regression equation, but it was considered more illuminating to compare predicted F0 for tense-lax vowel pairs at the same EMG value. Ideally, the EMG value on which to base the comparison should be close to the centre of the distribution of EMG values for both vowels. For each tense-lax pair the value was determined by computing the mean of the tense and lax EMG data individually, and then averaging these two values. At this EMG value the predicted F0 value was then computed from the regression equations for the tense and lax member of the pair. The results of this prediction (again with the 99% confidence intervals) are shown in Fig. 5.21. The location of the comparison values derived in this way is indicated by a cross symbol on the previous scatter-plot figures[23]. Referring back to these figures confirms that by and large the distribution of EMG data for the tense and lax vowels overlaps sufficiently for the comparison value to be located at an intuitively satisfying location. Inevitably there are a couple of cases where the EMG distributions do not overlap (e.g CG (1) /**y**/ voiceless context). Here we are forced to compare predicted F0 values for an EMG value that was not observed in the data (this is also why it seemed important to indicate confidence intervals in the figure)[24].

Looking first at the results for the gradient, the picture that emerges is remarkably consistent: lower values (i.e shallower gradient) for the lax vowels. This even applies quite well to speaker SF, whose scatter plots had appeared least supportive of the basic hypothesis. Across speakers, in the majority of cases the error bars do not even get close to overlapping, even though the use of 99% confidence intervals can be regarded as fairly conservative. Of the three speakers, CK shows the clearest difference in terms of gradient between the (non-/**a**/) tense and lax categories; this is also the speaker for whom the /**a**/-vowels most clearly differ from the other vowels. The tense-lax difference for CG is also consistent, mitigated only by the fact that /**a**/ does not clearly pattern differently from the other vowel categories.

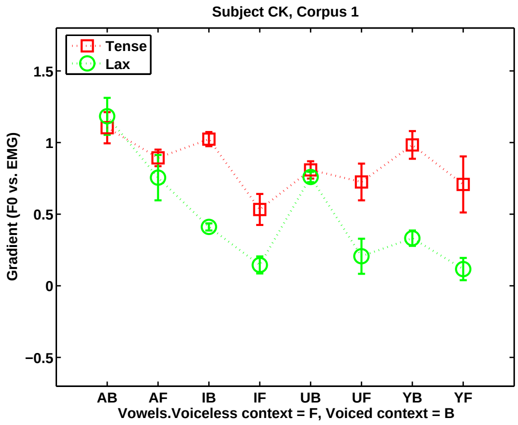
Turning now to the predicted F0 values at the reference EMG value, there is now in effect an exchange of roles between CG and CK. CG shows the clearest effect in the hypothesized direction (lower F0 values for lax vowels) with the /**a**/-vowels also differing most clearly from the other vowels. CK shows a consistent but weaker trend in the expected direction, mitigated by the fact that /**a**/ does not differ clearly from the other vowels. For SF no consistent trend can be identified.

---

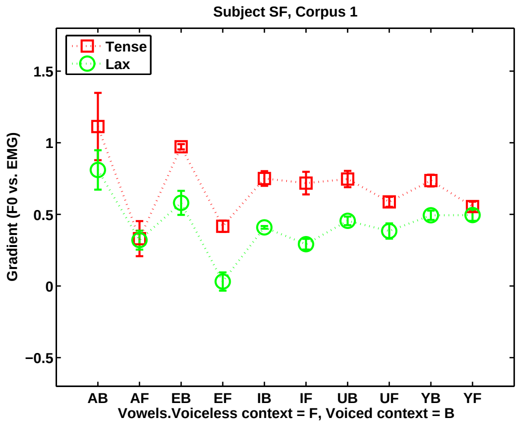
[23]Cases where the cross symbols are not located on the scatter-plot lines correspond to cases where the linear regression was not fully appropriate to the data.

[24]This procedure was preferred to simply averaging over all data in the pair, to avoid biasing the result towards the longer vowel in the pair , i.e the tense vowel.

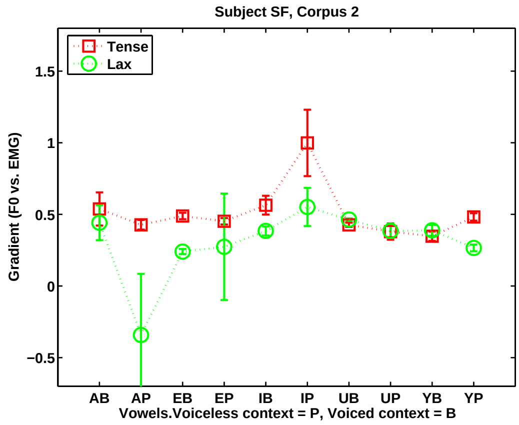
CK



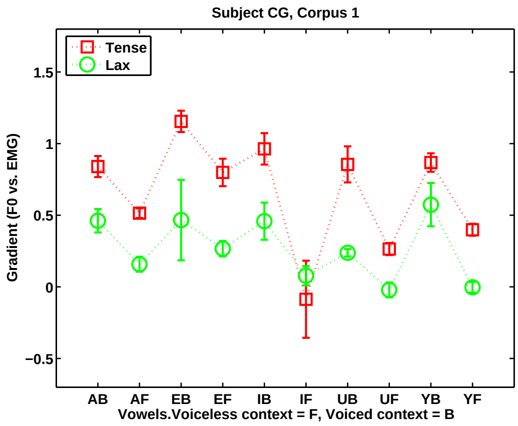
SF (1)



SF (2)



CG (1)



CG (2)

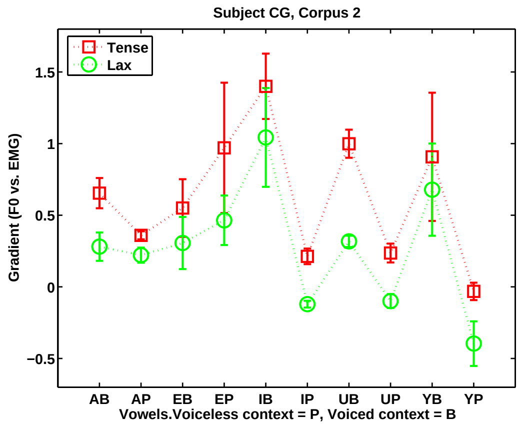


Fig. 5.20: Mean (over 10 repetitions) and standard error of gradient of relationship between F0 and EMG activity, comparing tense (red squares) and lax (green circles) vowel pairs (gradient based on normalized F0 and EMG). Separate panels for each speaker (and separate panels for each corpus for speakers CG and SF). The two-letter upper-case codes on the abscissa indicate vowel category (first letter) and consonant context (second letter). This letter corresponds to the identity of C1. Thus 'B' always indicates voiced context, while 'F' and 'P' indicated voiceless context for corpus 1 and corpus 2, respectively.

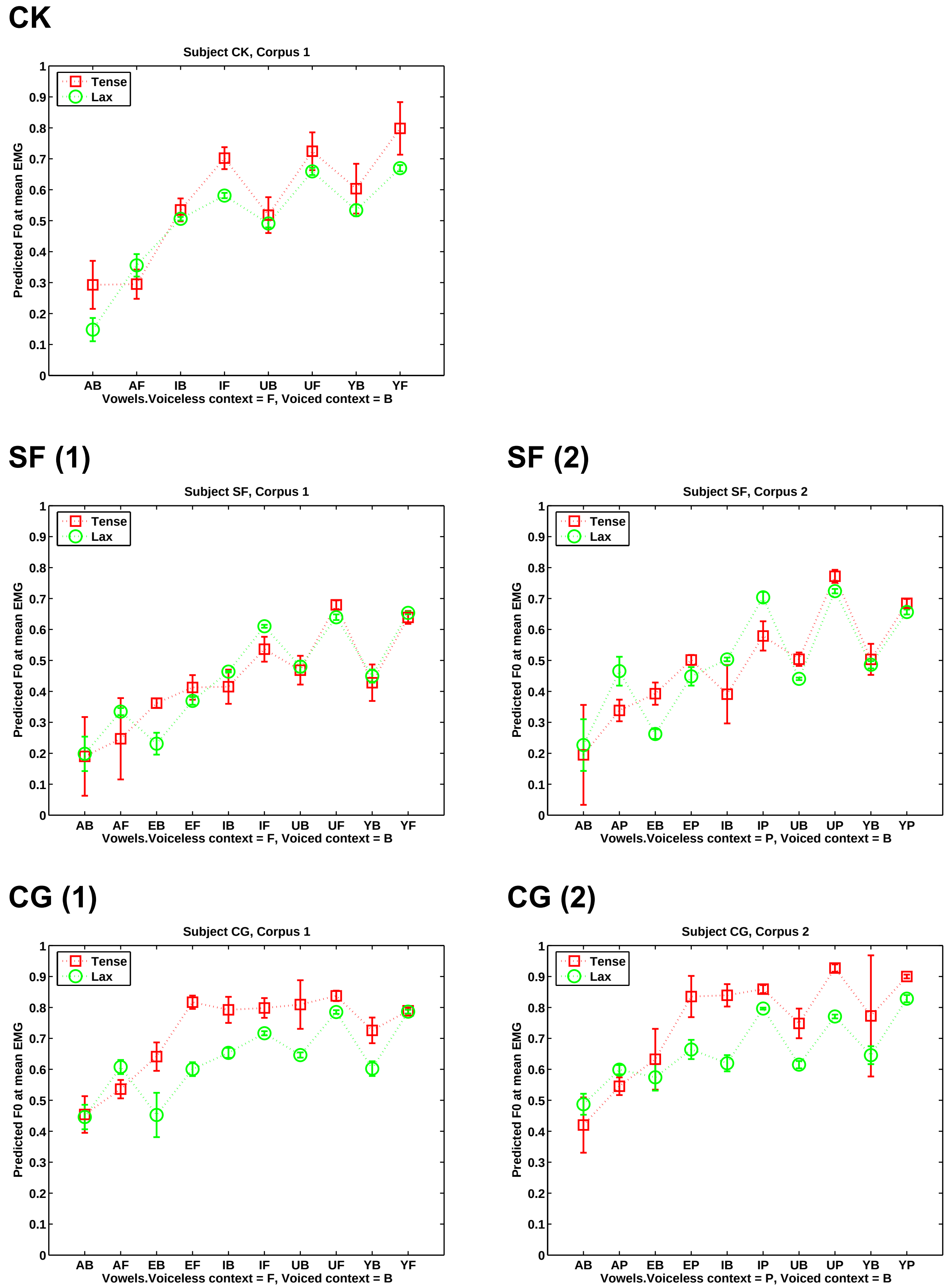


Fig. 5.21: Mean (over 10 repetitions) and standard error of predicted F0 at mean EMG activity, comparing tense (red squares) and lax (green circles) vowel pairs. Other details as for previous figure on F0 vs. EMG gradient.

The balance of the evidence presented above suggests that there could well be a difference in the relationship between muscle activity and F0 for tense vs. lax vowels. This was clear for CK and

CG (though expressed in a somewhat different way), rather marginal for SF. Exploring this relationship for single speech segments is not without its hazards, however. In particular, short vowels, especially in voiceless consonant contexts, provide only a small number of values for the regression analysis. This was a reason for using conservative confidence intervals. Given this, the robustness of the results is actually quite encouraging. Nevertheless, it appeared worthwhile attempting to cross-check with an alternative procedure. Before we turn to this we will make a brief diversion to expand on the relationship between EMG and F0 with respect to the consonant voicing distinction.

A diversion on consonant voicing

In the discussion of consonant voicing effects on p.67 it was mentioned that the regression analysis to be introduced for examination of the tense-lax distinction would also be useful for formalizing some of the observations made in that earlier section. The two figures just presented on the gradient of the F0 vs. EMG relationship and on the F0 predicted at a mean EMG level can indeed also be usefully inspected with respect to the voice-voiceless contrast. In both these figures, pairs contrasting in voicing of the consonant are located at adjacent positions on the abscissa, so relevant effects are quite easy to identify. Looking first at the predicted F0 at mean EMG it will be seen that there is a fairly consistent effect for higher F0 in the voiceless context. This restates the finding made in the earlier section that there may be salient F0 differences on the vowel related to consonant voicing even if there is not much difference in the level of EMG activity. The observation here from the regression analysis can be paraphrased as saying that F0 in the context of voiceless consonants is somewhat higher in the vowel than would be expected from the level of EMG activity measured in the vowel itself. Following the logic behind the regression analysis for intrinsic pitch effects this in turn indicates that at the current state of our analysis we are missing an effect that must also be contributing to the higher F0 in the vowel following voiceless consonants. In the concluding discussion we will offer some suggestions as to the source of this “missing” effect.

Regarding the gradient of the F0 vs. EMG relationship, there is fairly consistent effect for a shallower gradient for voiceless compared to voiced consonant contexts. In the concluding discussion, we will argue that it may not be a coincidence that those sounds that attract higher levels of CT activity, namely voiceless consonants and lax vowels, are those sounds that show a shallower slope for the F0 vs EMG relationship, and that this may provide an explanation for the repeatedly encountered VOICE x TENSE interaction (i.e clearer VOICE effect on the consonant with tense vowel context, and clearer TENSE effect on the vowel with voiced consonant context).

The patterns discussed here can also be found in more condensed and perhaps clearer form in the following main section of the results (p. 89ff on vowel-specific EMG activity, the relevant figures being Fig. 5.25 on p. 94 (predicted F0 at mean EMG) and Fig. 5.26 on p. 95 (F0 vs. EMG gradient).

F0 vs. EMG (3): Alternative approach to regression analysis

Following the diversion, we now return to the influence of the tense-lax distinction of the F0 vs. EMG relationship. As an alternative to the previous approach we re-calculated the regressions based on the mean EMG and median F0 from each individual token and pooled all vowels and consonant contexts.

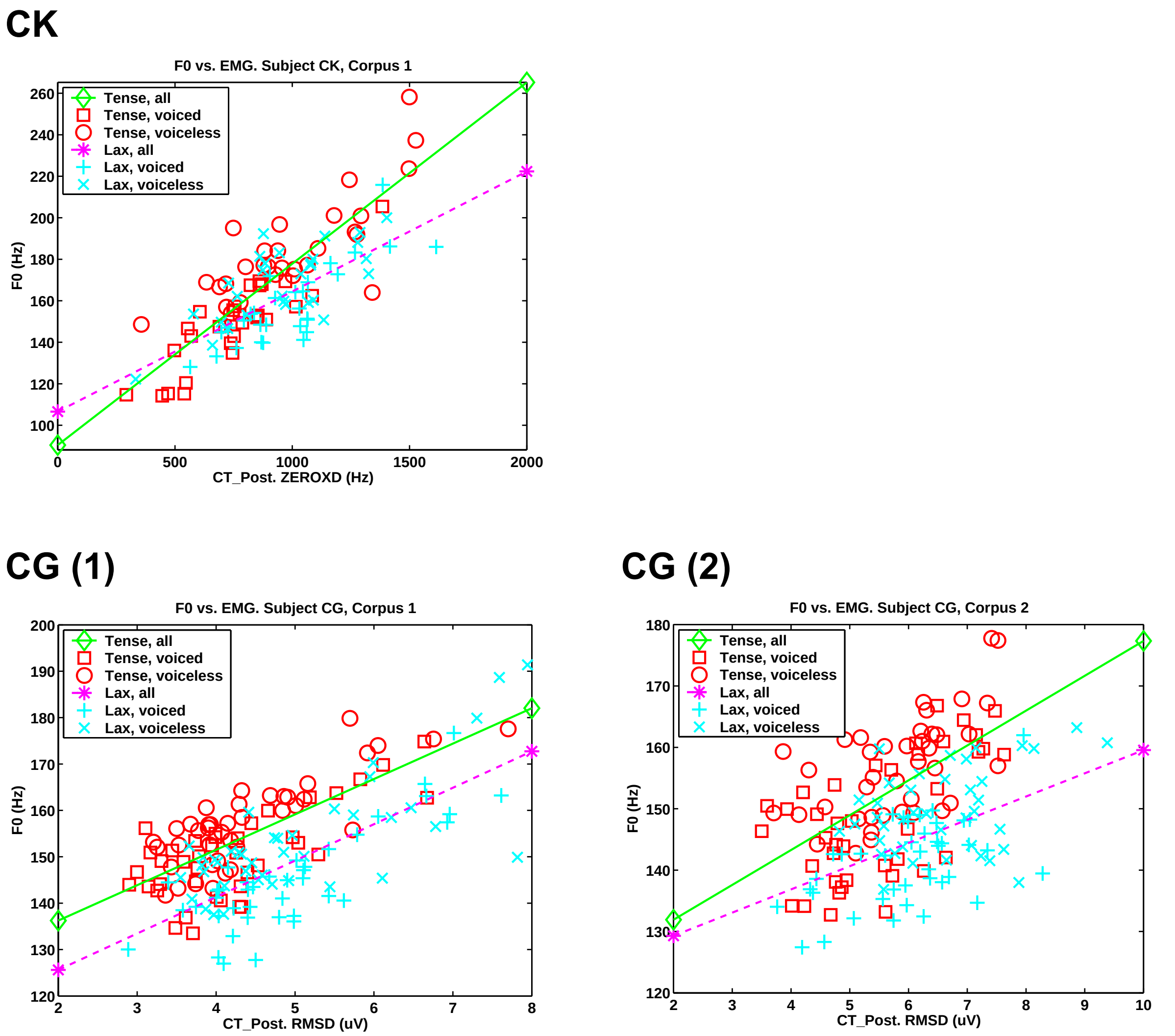


Fig. 5.22: Relationship between F0 and EMG based on average values over vowel segment of individual tokens. Regression lines shown for all tense data points (solid green line, diamonds at end-points) and all lax data points (dashed mauve line, asterix at end-points)

This gives quite a large number of data points, but also has definite disadvantages. /**a**/-vowels were excluded, in effect for the reasons given above, i.e the relationship should be systematically different from the other vowels. But it cannot necessarily be assumed that conditions are homogeneous for the other vowels, for example we saw in the literature review that tongue-pull effects may work somewhat differently for /**u**/ vs. /**i**/. Also, as just discussed in the diversion, consonant voicing appears to affect the gradient of the F0-EMG relationship. Thus the alternative approach smears over a number of effects that can hardly be disentangled at this level. For SF no consistent effects emerged, so these data are not shown here. For CK and CG the previous findings were basically confirmed: CK seemed to distinguish between tense and lax vowels in the gradient of the EMG-F0 relationship, whereas for CG the regression lines were shifted with respect to each other (in effect a difference in intercept). Given the smearing effects just alluded to, we did not necessarily expect to find statistically significant differences: in fact the difference for CK just reached significance at $p<0.05$, whereas that for CG did not.

5.4.3 Discussion

We will briefly recapitulate here the main findings with respect to the tense-lax distinction, and then come back to them in the concluding discussion of Chapter 6 to consider their implications in a wider perspective.

The key finding was the fairly consistent evidence that speakers may show increased laryngeal activity on lax compared to tense vowels. This opened the way to showing that the well-established intrinsic-pitch relation in German tense vs. lax vowels do not necessarily constitute an anomaly for mechanical models of intrinsic pitch. The regression analyses confirmed that a mechanical effect is most likely present: At comparable EMG levels, F0 tends to be lower for the lax vowels.

Given the delicate nature of EMG experimentation it would be hazardous to claim that we have conclusively demonstrated how a long-standing puzzle can be resolved. Nevertheless, the scenario based on higher laryngeal activity in the lax vowels had always been logically one of the main possibilities, and one can certainly say that it has now gained in plausibility.

The main complication in the results was the interaction between TENSE and VOICE effects, i.e the tense-lax difference appeared more salient in the voiced consonant contexts.

A further result that was not anticipated was the very clear tense-lax difference postvocalically, i.e on C2. This was interpreted as an interaction between segmental timing and timing of the intonation contour, rather than a tense-lax difference per se. It was, however, speculatively suggested that the effect is nonetheless interesting because it could give a physiological substrate for the disfavouring of voiced occlusives after lax vowels in German, since, as we saw in the first section of the results, higher CT activity forms part of the laryngeal manoevres for voiceless consonants.

5.5 Vowel-specific EMG activity

In this section we will look at EMG activity and F0 not from the rather specific point of view of the German tense-lax vowel distinction but from the perhaps more familiar point of view of "traditional" intrinsic pitch studies. As discussed in the literature review, the key question revolves around whether the very robust intrinsic pitch difference between low and high vowels involves an active component in terms of laryngeal muscle adjustments, or whether the automatic biomechanical consequences of vowel articulation provide sufficient explanation.

We will first consider briefly the statistical results for the main effect of VOWEL in the EMG measurements (i.e the remaining independent variable in the design; refer back to the beginning of Chapter 5 (p. 52ff for the complete overview of the statistical results). We will then look in more detail at the relationship between EMG and F0. As in the previous section on the tense-lax contrast, 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.

5.5.1 The main effect of VOWEL

As vowel-specific information on EMG activity is already available in figures used to present the results for the voicing and tense-lax contrasts reference to these figures can be made, if desired. In particular, referring back to the section on the tense-lax distinction, Fig. 5.12 on p. 72 showing average EMG for the vowel segment is fairly clear. New figures presenting much of the same information in a slightly different way will be shown below.

For CK, the high vowels as a group show quite consistently higher EMG 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**/. The pattern for CG was different. In the first corpus no significant effect of VOWEL was found, while in the second corpus /**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 ($p<0.001$). The higher activity in /**a**/ was unexpected; while there is some debate as to the interpretability of results reported in the literature for EMG differences between high and low vowels, there has been no suggestion that the difference could go in the opposite direction to intrinsic pitch.

5.5.2 Basic F0 patterns in the vowels

Previous figures also provide sufficient information to view vowel-specific F0-patterns and will not be repeated here. See in particular Fig. 5.9 (p. 67) in the section on the voicing contrast. Basically, all speakers are unremarkable, showing the expected intrinsic pitch 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**/.

5.5.3 Relationship between F0 and EMG: Overview

In this section we look at the relationship between F0 and EMG averaged over each combination of vowel category, tenseness, and consonant voicing. In the next section we look at the regression pattern within each of these categories.

The following figure shows for each subject and corpus the relationship between EMG and F0 when the data points in the scatter diagrams consist of the average values within each linguistic condition.

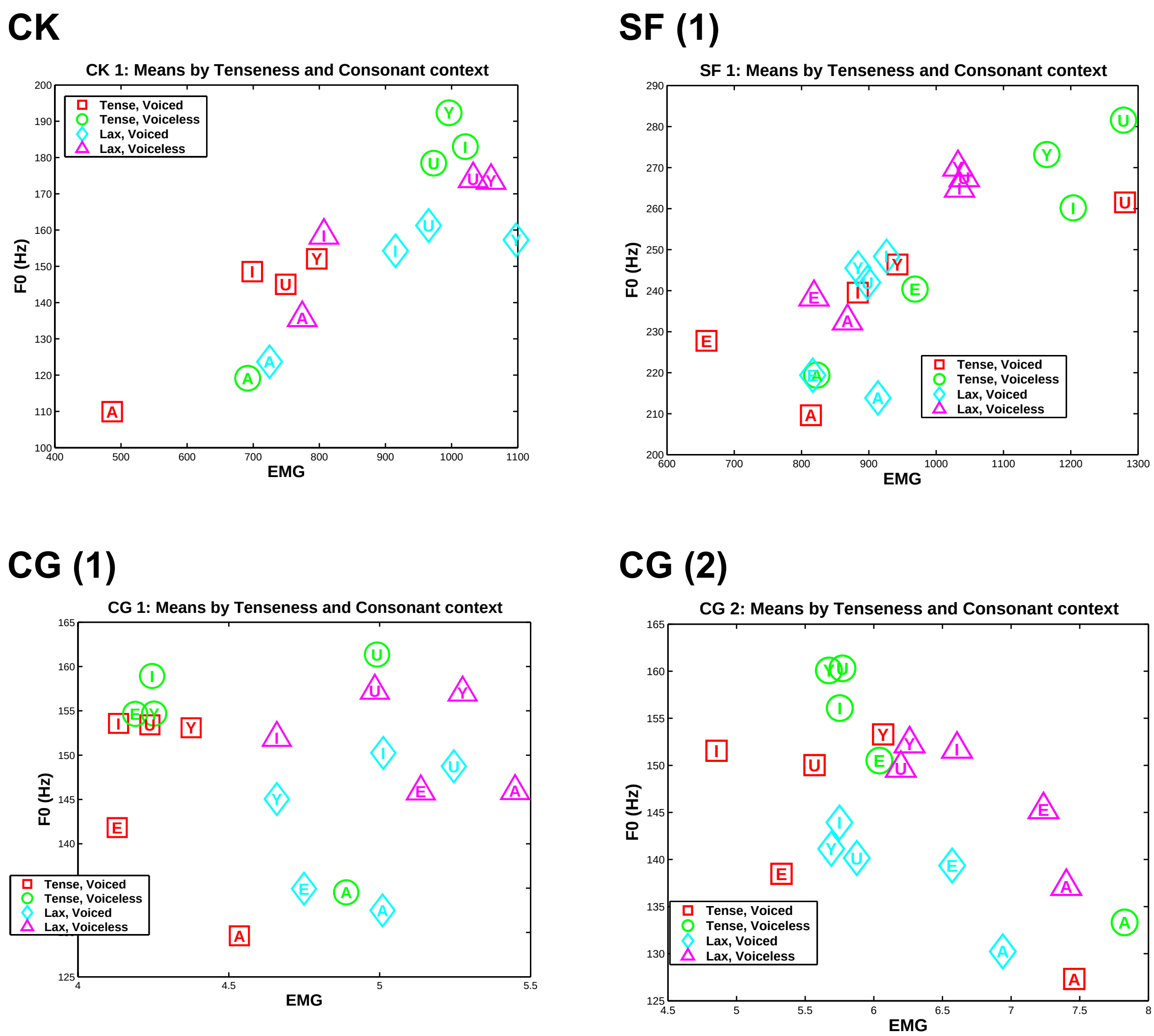


Fig. 5.23: Scatter plot of F0 vs. EMG based on mean value of each linguistic category in the corpus (vowels * tenseness * consonant voicing)

The patterns that emerge follow on from the ANOVA results just reported, 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 EMG 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 red squares and green circles) having a clear association between low EMG and low F0, and vice-versa. The lax vowels (cyan diamonds and magenta 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 EMG activity than for low vowels (i.e in the direction of intrinsic pitch) it is tempting to conclude for these speakers that intrinsic pitch is in fact the result of an active laryngeal adjustment. We will see below that this conclusion is probably 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 EMG and F0, while for the second corpus the coupling of the unexpectedly high EMG activity for /**a**/ (already mentioned in the ANOVA results) with completely typical intrinsic pitch values result in a tendency to a *negative* relationship between EMG and F0 (all /**a**/ data points are located at the bottom right of the plot).

In the next section we compare over vowels the relationship between EMG and F0 *within* each vowel category. The reasoning is similar to that employed for the analysis of the tense-lax distinction. The linking of high F0 and high EMG for CK and SF can only be taken at face value if a single regression line captures the relationship between EMG and F0 for both high and low vowels. If not, then the presence of further influences on F0 must be assumed.

5.5.4 Relationship between F0 and EMG: Regression analysis

In this section we employ the same kind of regression analysis as used above for the analysis of the tense-lax vowel contrast. In other words we will first present scatter diagrams of the relationship between F0 and CT activity based on ensemble average data (now juxtaposing the vowel categories /**a**, **e**, **i**, **y**, **u**/) , and then summarize the characteristics of these scatter plots in terms of the gradient of the relationship, and the F0 value associated with the same typical EMG value for each vowel.

For the scatter plots it is possible to give a more compact representation of the data than for the tense-lax or voicing distinction, since there are now 5 vowel categories (4 for CK) to compare in each panel (whereas for tenseness or voicing there are, of course, only two categories to compare). Four panels per subject are then sufficient to cover each combination of tenseness and voicing[25].

Fortunately - although a large amount of information is now compressed onto a single page - the key finding emerges very clearly.

[25]The scatter plots also give an additional compact summary of the tense-lax differences, particularly in terms of the gradient of the F0-EMG relationship, since contrasting tense-lax items are located in adjacent panels.

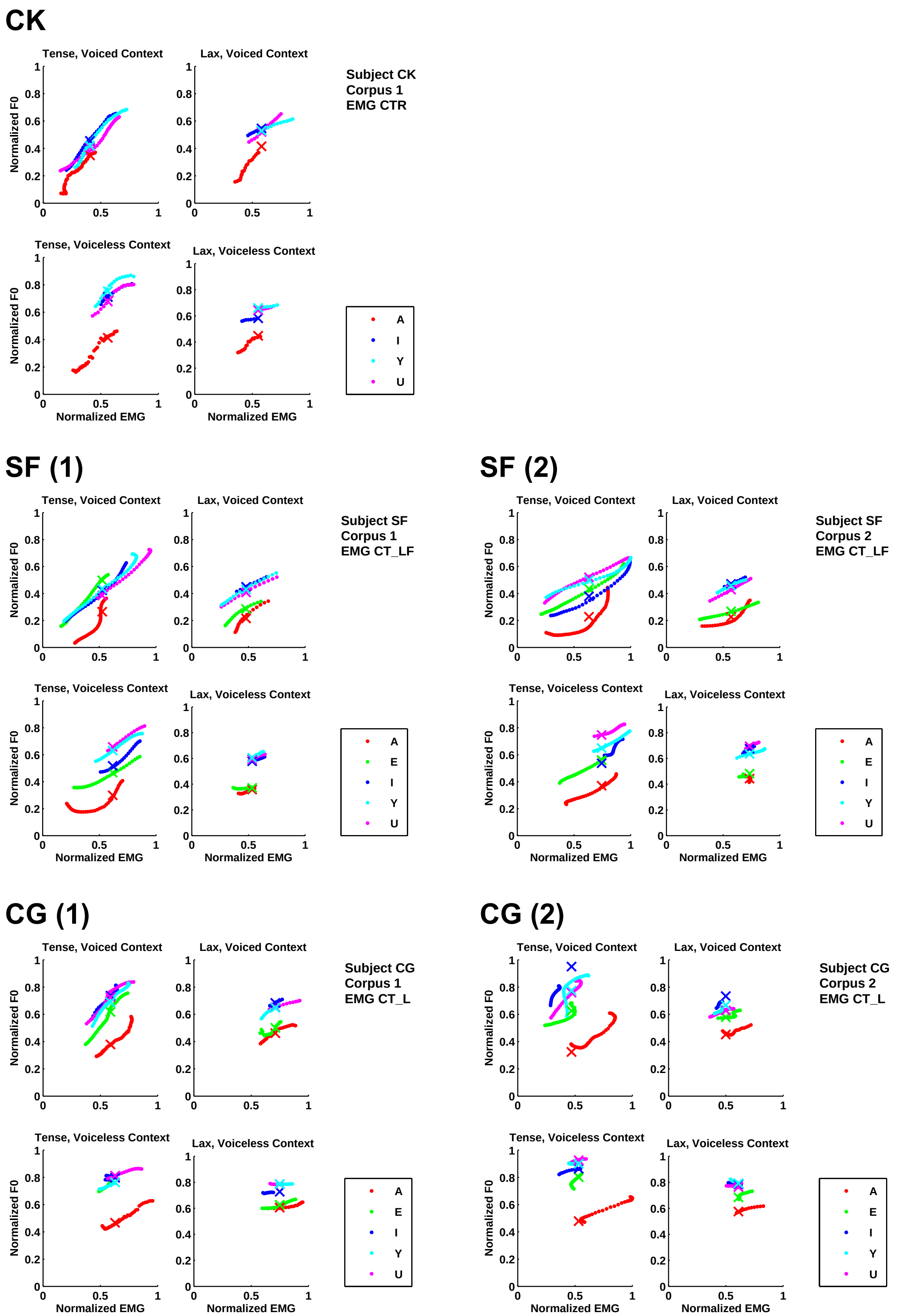


Fig. 5.24: Relationship between ensemble-averaged F0 and EMG over the vowel segment. Trajectories colour-coded for each vowel category.The four panels for each speaker represent the four combinations of vowel tenseness (tense, lax) with consonant context (voiced, voiceless). Both F0 and EMG data have been normalized to a range of 0 to 1.

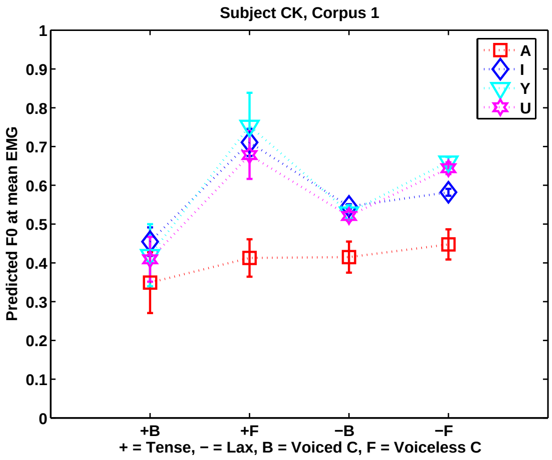
There is an absolutely consistent difference between the low and high vowels (i.e essentially /**a**/ (red) vs. /**i**/ (blue), /**y**/ (cyan), /**u**/ (magenta) (the possibly intermediate case of /**e**/ will be considered immediately below)). At comparable EMG levels F0 is lower for /**a**/ (as for the tense-lax analysis, the data points marked by crosses are displayed in an additional summary figure below). This, of course, is exactly what tongue-pull theories of intrinsic pitch would predict. Note, in particular, that the result now applies to all subjects in equal measure, i.e the distinction between CG and the other two subjects no longer applies. The robustness of the effect is underlined by the fact that it remains clear even for lax vowels in voiceless context, where the amount of data on which to base the analysis is sometimes rather restricted. As a point of detail, the position of /**e**/ also makes good sense. The basic expectation might be that it is intermediate between the high and low vowels. It turns out to be much closer to /**a**/ in the lax case. Tense /**iː**/ and /**eː**/ may actually have quite similar tongue position, whereas lax /**ɛ**/ is undoubtedly the notionally front vowel (apart from lax /**œ**/) with least tongue advancement (or conversely most pharyngeal constriction). Since, as already discussed, a possible tendency for lax /**a**/ is for less pharyngeal constriction than tense /**ɑː**/, it is thus quite plausible that lax /**ɛ**/ and /**a**/ could be located quite close together in these scatter plots.

Inspection of the scatter plots suggests that the differences between the vowels emerge more clearly from comparison of F0 at a common EMG value than from comparison of the regression gradients, so the summary figures for the predicted F0 values will be presented first. They essentially recapitulate (with the addition of 99% confidence intervals) the finding just discussed that the regression lines for /**a**/ are displaced relative to the other vowels. Each position on the x-axis corresponds to one panel in the previous figure, i.e one of the four possible combinations of vowel tenseness and consonant voicing. The confidence intervals serve to emphasize the robustness of the effect (the relatively large confidence intervals for /**a**/ of speaker SF are due to the departure from a purely linear relationship (curved pattern in the above figure)).

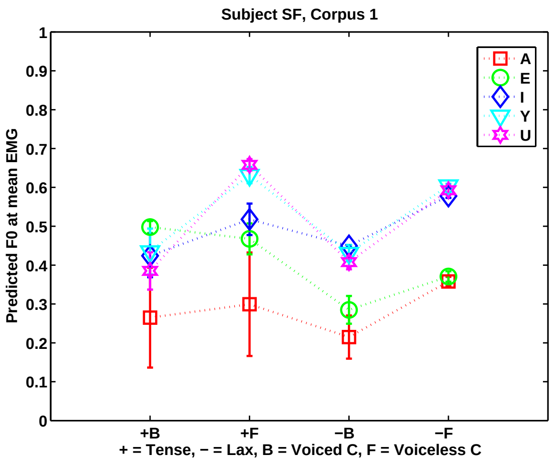
As an aside: This is one of the figures that have already been alluded to regarding the dependency of the F0 vs. EMG relationship on consonant voicing (see initial discussion on p. 67, followed by further discussion on p. 86 relating to Figs. 5.20 and 5.21). As in the earlier figures voiced-voiceless pairs are located at adjacent positions on the abscissa which makes it easy to observe a quite consistent effect of higher F0 for voiceless compared to voiced at comparable EMG (in similar terms the following figure also illustrates the shallower slope for voiceless compared to voiced).

Unlike the predicted F0 at a common EMG value, the gradients of the F0 vs. EMG relationship do not show any obvious pattern (and, in fact, are probably not crucial to the argument given that differences in the regression pattern are already clearly documented with the predicted F0). On the one hand, it is interesting that speaker CK, for whom F0-EMG gradient was relevant for the tense-lax distinction, is the only speaker where clear differences emerge, namely *steeper* gradient for /**a**/ vs. the other vowels. On the other hand, it is not clear what significance can be attached to this. The effect is essentially opposite to that found for the tense-lax distinction, since there the tendency (also visible in the present figures) was for shallower slopes with the lax vowels. Thus if shallower slope is associated with less tongue advancement, one might have expected shallower rather than steeper slope here for /**a**/. Most likely, for the present intervocalic comparisons the slopes are simply not directly comparable. CK is the speaker where the amount of overlap between /**a**/ and the other vowels is particularly small both in terms of EMG and F0. Since it is probably not reasonable to expect a linear relationship over the whole F0-EMG space anyway comparisons are risky when there is so little overlap between the vowel categories.

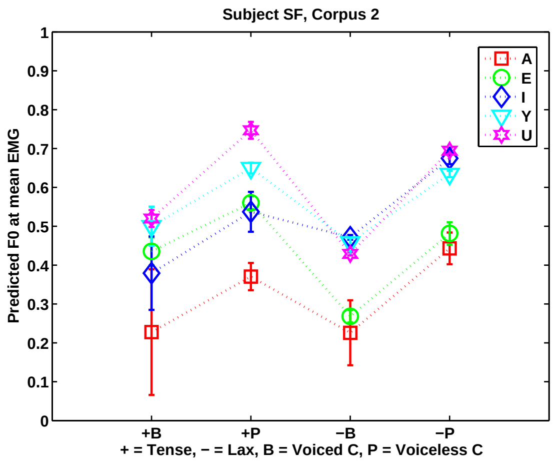
CK



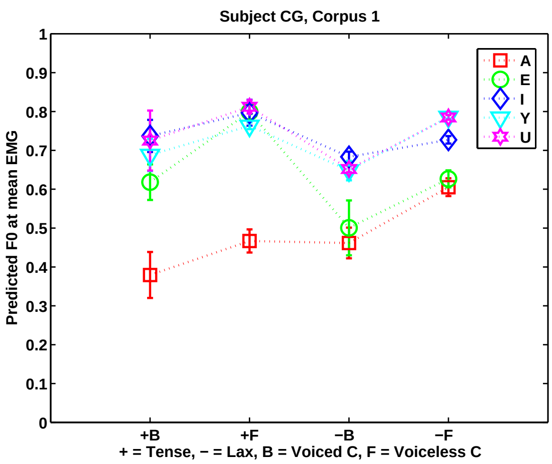
SF (1)



SF (2)



CG (1)



CG (2)

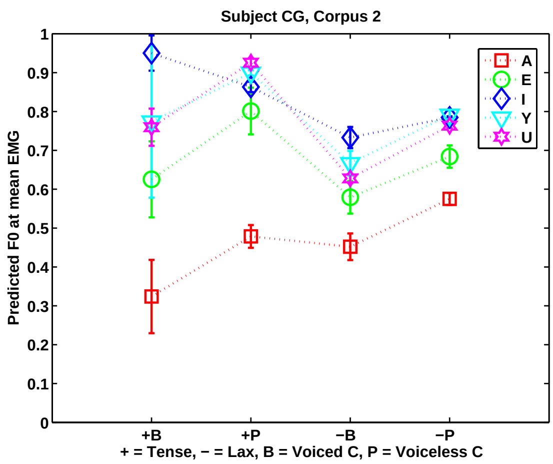
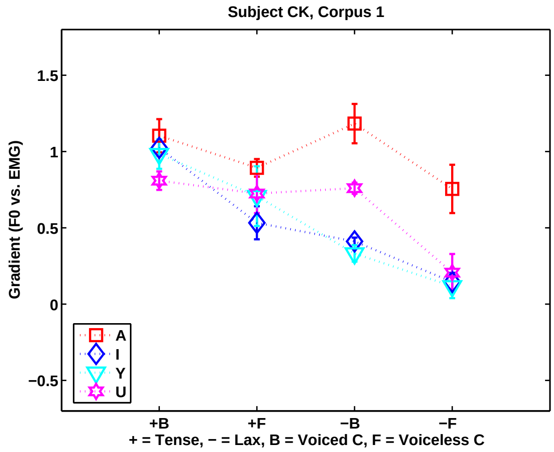
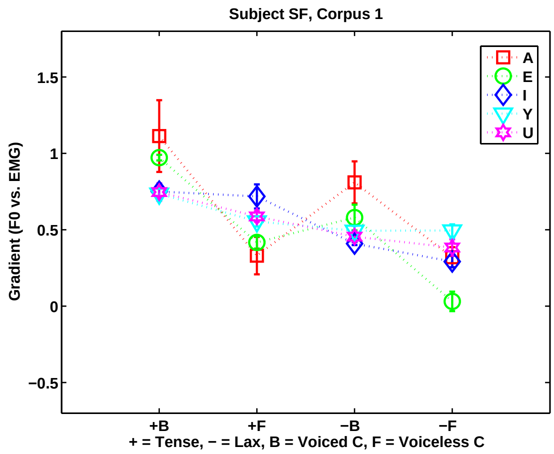


Fig. 5.25: Mean (over 10 repetitions) and standard error of predicted F0 at mean EMG activity level comparing vowel categories (coded by colour and symbol). Positions on the abscissa represent the possible combination of vowel tenseness with consonant voicing; from left to right: tense voiced, tense voiceless, lax voiced, lax voiceless. Based on normalized F0 and EMG values.

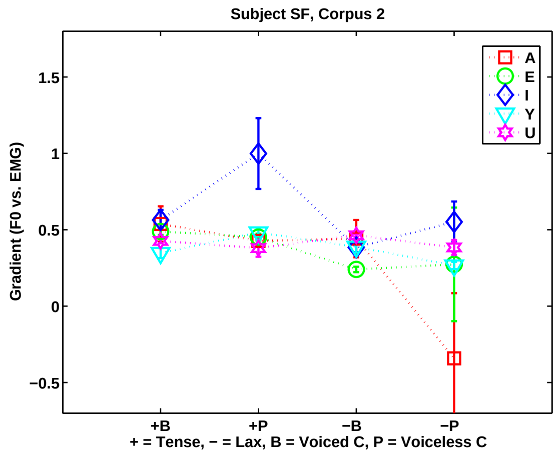
CK



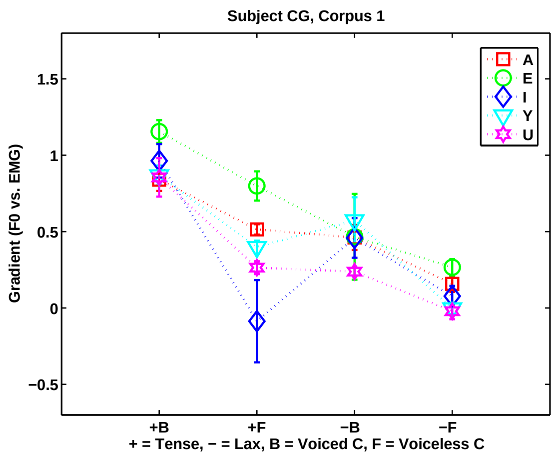
SF (1)



SF (2)



CG (1)



CG (2)

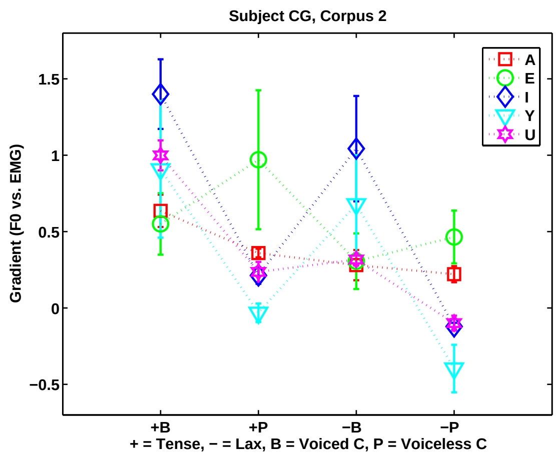


Fig. 5.26: Mean (over 10 repetitions) and standard error of gradient of relationship between F0 and EMG activity comparing vowel categories (coded by colour and symbol). Other details as for previous figure.

5.5.5 Discussion

What interpretation can be given of these results?

Since the most robust effect was that at comparable EMG levels F0 is lower for the low vowels, we assume that the tongue-pull mechanism is the basic mechanism underlying intrinsic pitch. This fits in with the tense-lax results which showed, perhaps less conclusively, that the specific intrinsic pitch patterns in tense vs. lax vowels need not be in conflict with the tongue-pull hypothesis. Thus we follow Whalen et al. in rejecting an active enhancement mechanism at the heart of intrinsic pitch. However, the fact remains that two speakers did also have higher EMG 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 (discussed in Chapter 2 above, p. 8): 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. Note for example that our speaker CK has very large intrinsic pitch effects. This would also fit in with the thrust of the tense-lax findings, since these indicate that a basic biomechanical effect can be overlaid by planned muscle activity in specific cases.

A hybrid explanation could also explain why in the perceptual experiments of Fowler & Brown (1997) the amount of the putatively mechanical contribution to F0 in high and low vowels that is parsed out by listeners comes out substantially less than the typical difference in F0 actually occurring in vowel production (cf. p. 14 above; further discussion of perceptual effects in the concluding discussion below).

The fact that active laryneal muscle contributions are not an essential component towards supporting spectral contrast in vowels via F1-F0 distance is indicated by the fact that they were not present in all three speakers. In fact, the finding for which we have no explanation is the tendency towards more EMG activity in the low vowels of CG (2), which would constitute a weakening of intrinsic pitch effects.

5.6 Functional differentiation of pars recta and obliqua of the cricothyroid?

In this section we look first at the background for the assumption that a functional differentiation between pars recta and obliqua could be an interesting concept in understanding laryngeal function in speech, and then consider the relevant experimental results for speaker CG. Since we have here only one subject on whom to base discussion, any conclusions will be of necessity particularly tentative.

5.6.1 Background

The cricothyroid joint is generally considered to have two degrees of freedom, namely rotation and translation, and it is tempting to assume that the two compartments of the cricothyroid, the pars recta and pars obliqua can be associated respectively with these two movement components. To supplement the anatomical background given in Chapter 4 above (p. 36ff) the following figure summarizes this idea from a recent review of F0 mechanisms by Honda (2004).

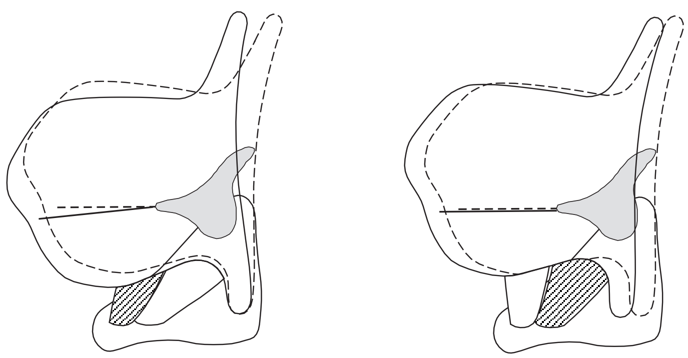


Figure 2: *Rotation and translation of the cricothyroid joint. The pars recta and oblique of the cricothyroid muscle have been assumed to cause rotation and translation of the joint, respectively.*

Fig. 5.27: Illustration of the possible two functional components of the cricothyroid muscle (from Honda, 2004, Fig. 2)

Honda notes that the rotational component is better documented. In a recent study of his own using a high-resolution MRI technique customized for the laryngeal region he found a rotational difference of 5deg. and a translational difference of 1.25mm between sustained phonation at frequencies of 120 and 180Hz. The translational difference may appear small, but Honda notes that it could be particularly effective in changing vocal fold tension and hence F0. Regarding a functional differentiation between pars recta and obliqua, he shows a small amount of data from an earlier investigation suggesting that pars obliqua showed a stronger relationship with F0.

The pairing of rotation and translation with pars recta and obliqua plays a prominent role in the influential Fujisaki model of intonation. The basic ideas have been presented in numerous publications (e.g recently Fujisaki, 2004). The key feature of the model is the superposition of phrase and accent components to generate F0 contours. The phrase component is modelled as the response of a second-order linear system to an impulse input, while the accent component is modelled as the response of a further second-order system to a step input. The time-constant of the phrase component is longer than that of the accent component. Accordingly, the phrase component is associated by Fujisaki with the translational/pars obliqua cricothyroid activity and the accent component with rotation/pars recta, since in his opinion the translational movement has a much longer time-constant (e.g Fujisaki, 2004, p.5). (Honda, 2004, also refers to the possibility that "pars obliqua determines slower changes in vocal fold tension" (p. 740)). In one of the earliest presentations of the model (Fujisaki, 1988) some calculations are given to show that the assumption of a longer time-constant for translation is plausible, but it is conceded that it is not yet clear that time-constant ratios for the phrase and accent component typically derived from modelled intonation contours actually match the time-constant ratio of translation and rotation, and also that evidence for differential activation of pars recta and obliqua is not yet available. In fact, in a comment on Fujisaki, 1988, Hirano remarks (p.355) that the hypothesis of differences between pars recta and obliqua is worth testing but that in his own earlier EMG work no obvious qualitative differences between the two parts was found[26].

It is not clear that in the intervening period stronger evidence for differentiation of pars recta and obliqua has emerged (the admittedly small amount of EMG data presented by Honda is not really supportive). It is also not clear to me whether further attempts have been made in the literature to estimate the time-constants of translation and rotation.

On this background, it seemed interesting within the framework of the present experiments - and to the extent that the "hit-rate" of the insertions would permit - to compare activity of pars recta and obliqua. It should be emphasized that this was not intended as a test of the Fujisaki model itself. In my opinion the central superposition feature of the model could well be valid even if a specific physiological substrate cannot be documented. In addition, the corpus was not designed to look specifically at phrase and accent components of intonation. Rather, on the basis of the preceding discussion, it seemed at least plausible that if the rotational component is better suited for short fast changes, then it might be particularly important for the presumbably short-term nature of the effects in which we are interested, i.e voicing contrast in consonants, and tense-lax contrast in vowels.

5.6.2 Data

Table 4.1 on p. 46 above gave correlations between EMG activity and F0 for the different insertion locations. This allowed no firm conclusions about a difference between the insertion locations. The posterior insertion and ANT1 appeared fairly similar, while the ANT2 insertion had generally rather weaker correlations (details depending slightly on the corpus and signal parameters used (RMS amplitude vs. zero-crossing rate).

The following figure serves simply to illustrate that the different insertions have broadly similar patterns, with no marked differences in timing.

The designation of the EMG signals is as follows:

- CT_R1: First Anterior Insertion
- CT_L: Posterior Insertion
- CT_R2: Second Anterior Insertion

---

[26]In a further comment Kahane considers on anatomical grounds the possible extent of translational movement to be very restricted.

Each EMG signal is shown as the raw signal overlaid by the RMS curve of the differentiated signal. Part of the utterance "habe Liebe besucht" is shown. Note that the EMG signals have not been time-shifted relative to audio and F0.

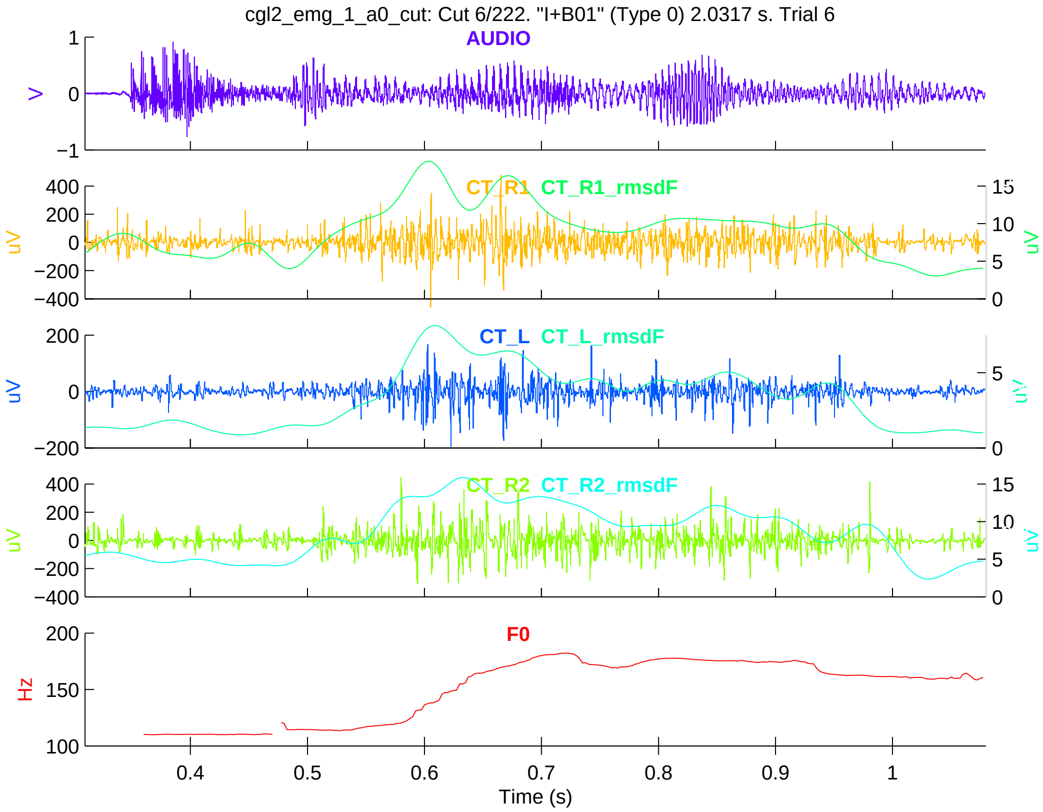


Fig. 5.28: EMG signals from two anterior insertions (CT_R1 and CT_R2; panels 2 and 4, respectively, counting from top) and one posterior insertion (CT_L; panel 3). Each EMG panel shows both the raw EMG data (axis on left) and RMS amplitude (axis on right). Utterance is "habe Liebe besucht" (slightly truncated).

Thus the initial impression might be that there are no systematic differences between the insertion locations. In fact, it was possible to observe differences from two points of view, firstly in a non-speech task, and secondly in the sensitivity to the effect of the linguistic variables in the experiments.

One of the control tasks used in the first part of the experiment was swallowing. This is shown in the next figure. Here, completely different patterns were found. The anterior insertions showed essentially no activity while the posterior one showed a pattern of activation and suppression.

The same EMG signals are shown as in the figure above (note that scaling for CT_L has been changed).

The activation-suppression pattern was considered by K. Honda (p.c) to be more typical of cricothyroid activity. But the main point is simply that this demonstrates that we were recording from parts of the muscle with potential for differential activation. By contrast, all insertions showed similar activity for glottal attack, which was used as a further control task later in the experiment. Strong activation related to each glottal closure would indicate lateral cricoarytenoid rather than cricothyroid insertion. Our insertions all showed some activity, but undifferentiated relative to the glottal closure events, thus probably representing a background level of overall laryngeal activation.

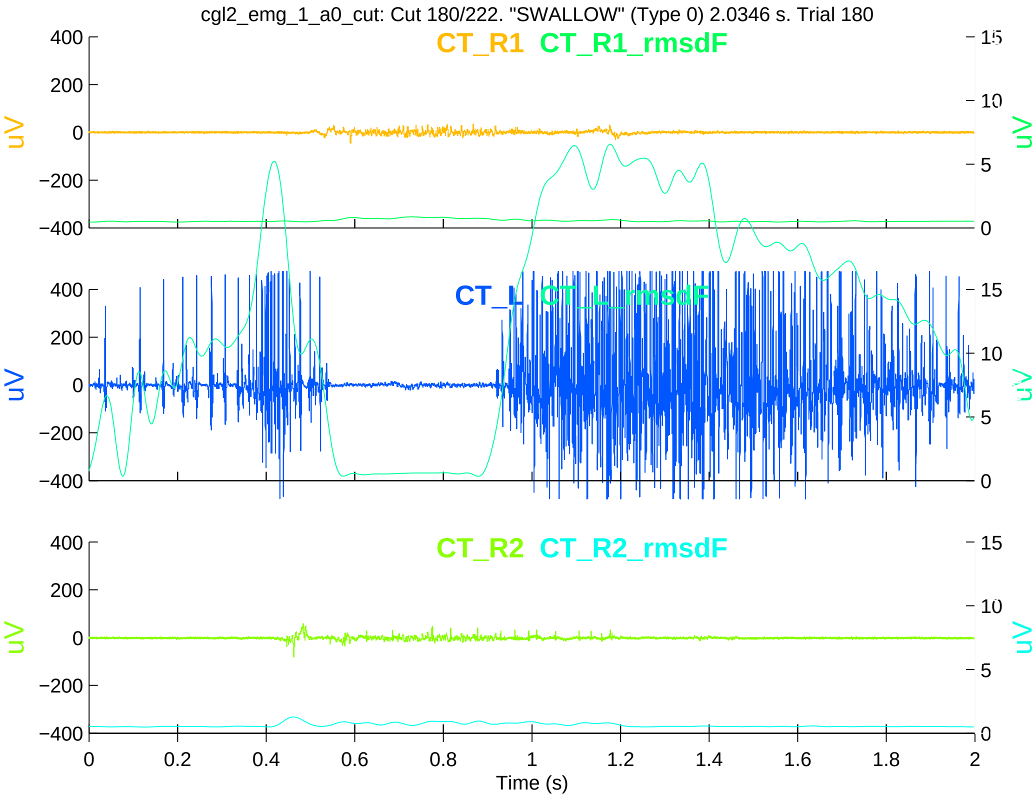


Fig. 5.29: Swallowing task for same CT insertions as shown in previous figure

With respect to speech tasks, the following figures compare results for the insertion locations for two selected cases that played a particularly important role in the main experimental analysis, namely the C1 segment with respect to the voicing distinction, and the Vowel segment with respect to the tense-lax distinction. In both figures the panel in the left column shows the results for the posterior insertion, i.e the results already seen in the corresponding sections above, while the right column shows the results from the two anterior insertions. The main point to make is that the two anterior insertions (while by no means identical, e.g ANT2 varies more with vowel category, as seen in the figure below for the tense-lax contrast[27]) show a much less clear separation with respect to the independent variable than does the posterior insertion. The overall direction of the effects is essentially the same for all three insertions (i.e higher activity for voiceless vs. voiced, and for lax vs. tense). The corresponding ANOVA results are given in the tables at the start of Chapter 5 (p. 52ff). There they are given for both Corpus 1 and Corpus 2; the figures below are restricted to Corpus 1.

[27]ANT2 here shows the unexpected vowel-dependent pattern that was found for the posterior insertion in Corpus 2, namely *higher* activity for the low vowels (cf. Fig. 5.12 on p. 72).

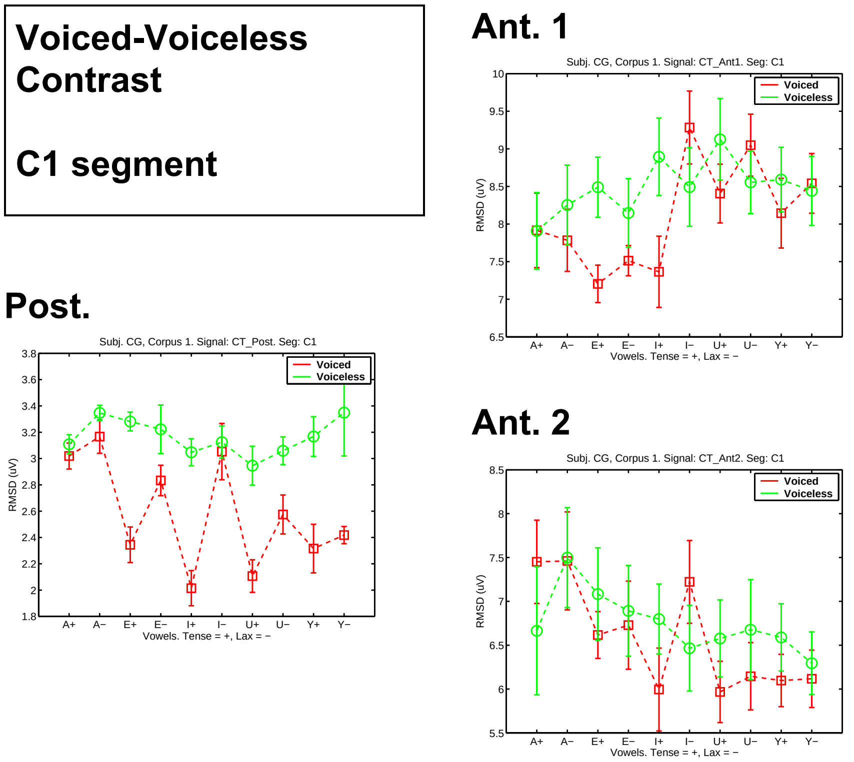


Fig. 5.30: Comparison of results for consonant voicing contrast with respect to the three different CT insertions.

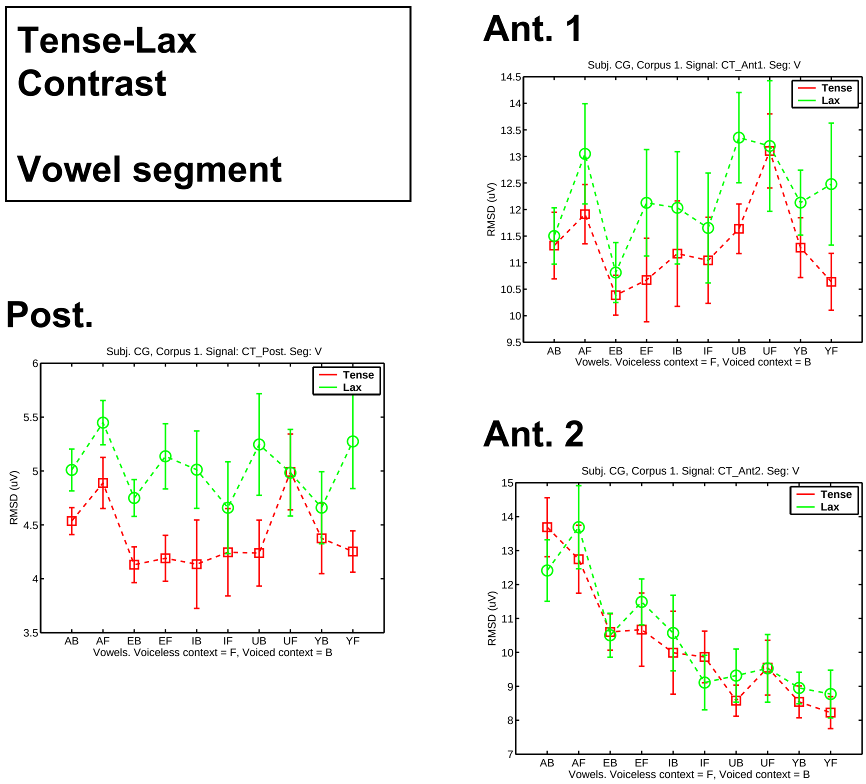


Fig. 5.31: Comparison of results for vowel tense-lax contrast with respect to the three different CT insertions.

From these observations we would draw the following conclusion: There is certainly no evidence that pars recta is particularly involved in fast segmental adjustments. If anything, it appears somewhat less sensitive to linguistic oppositions at this level. The result might in a sense confirm the earlier observation of Honda that the relationship with F0 was weaker for pars recta. Note, though, that in our data insertion ANT1 had a similarly strong overall correlation with F0 as the posterior insertion; it is only at the detailed segmental level that differences emerge. This could tie in with our speculation in the review of the intrinsic pitch literature that Dyhr found clearer evidence for vowel-specific differences than Whalen et al. because his recordings were based on pars obliqua insertions whereas those of Whalen et al. were pars recta. But this point, too, cannot go without qualification: The present speaker CG was the one who showed unexpectedly high EMG activity for /**a**/, and this turns out to be particularly clear for insertion ANT2.

Given this murky situation it is worth recalling that the swallowing task showed similarity for the two anterior insertions but a complete difference from the posterior insertion. Thus, it is certainly intriguing that potential differentiation is clearly present. Nevertheless, it would seem to be extremely difficult to show a clear functional differentiation for speech-related tasks. At the moment, the two most likely possibilities that remain seen to be either that there are genuinely no differences, or that posterior insertions are somewhat more sensitive to subtle segmental differences. It will remain a considerable challenge to gather a sufficient number of speakers with a sufficiently representative range of insertion locations to be able to distinguish between these possibilities given the differences that can occur even between insertions that are basically in the same part of the muscle.

6 General Discussion

Of the four areas covered in the previous chapter (consonant voicing, tense-lax distinction, vowel-specific effects, functional differentiation of the cricothyroid) we will not go further into the latter as it formed a short, self-contained section whose discussion can now be left as it stands. The remaining three topics have enough points of contact to warrant a more generally-framed discussion than could be given at the conclusion of each individual results section.

We will first consider the results for the tense-lax distinction in German, as this was the main impetus for conducting the investigation. This also provides a suitable platform for then moving on to consider the wider implications of the study for intrinsic pitch in general. The next main part of the discussion will examine consonant voicing. A theme common to the intrinsic pitch 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? The discussion will conclude by looking more briefly at the TENSE x VOICE interaction, and its implications both for experimental design and the status of the individual tense-lax and voicing related differences.

6.1 Tense vs. Lax

6.1.1 Introduction

Regarding the tense-lax distinction, the basic result was that the hypothesis of increased laryngeal activity for lax vowels gained in plausibility. The intrinsic pitch relationship of German tense-lax vowels has frequently been quoted as revealing the shortcomings of mechanical explanations of intrinsic pitch (e.g Diehl, 1991; also Kingston, 1992), but our analyses further showed that, on the contrary, the results could well be compatible with the presence of a mechanical effect of the kind often assumed. This is actually a pleasantly conciliatory result since, on the one hand, mechanical effects on vowel F0 are confirmed, but at the same time deliberate modulation of F0 as part of what might traditionally be regarded as simple segmental distinctions is clearly a strategy that is available to speakers. Whether this deliberate modulation of F0 also plays a role in the more traditional area of intrinsic pitch, i.e high vs. low vowels, will be discussed further below.

For German we now have the intriguing situation that the underlying laryngeal muscular activity is assumed to differ for tense vs. lax vowels, whereas the physical expression in the signal, i.e F0, is essentially the same. This raises the interesting question as to how F0 is perceived in such a situation. Following the logic of the Fowler & Brown experiment on the perception of /**i**/ and /**a**/ where a tendency was found that /**a**/ is perceived as higher in pitch if it has the same F0 as /**i**/, then one could ask whether lax vowels are perceived as higher in pitch than tense vowels at the same F0, i.e listeners perceive the fact that for such a situation to occur (identical F0) more laryngeal activity is required for /**a**/ on the one hand, and lax vowels on the other hand. It turns out that a complete treatment of the perception of intrinsic pitch would go far beyond the scope of this work (see references below for ongoing work by D. Pape), but some early results will be presented briefly, perhaps more as a cautionary tale than anything else.

Before doing so, it is convenient to discuss the question of whether laryngeal differences between tense and lax vowels have not already been observed elsewhere.

There is indeed one specific area where this could be the case: it has been suggested that tense and lax vowels can differ somewhat in terms of voice quality, with the tense cognates showing more of a tendency towards a breathy voice quality. This question has become somewhat entangled with the question of whether the tense-lax distinction in English (where the term is probably even more of a misnomer than in German) is similar to the ATR feature of many vowel harmony languages (in the latter case there can be no doubt that accompanying voice quality differences are worthy of discussion). We will not try and unravel this here. Restricting consideration to German, Jessen (2002) has shown in a very carefully argued contribution that there is some evidence for a manifestation of laryngeal differences. To some extent this is seen in the spectral balance of the radiated spectrum of the sounds (though the question of the appropriate normalization of vowels differing substantially in F1 is a thorny one), but more interestingly also in some features of the electroglottographic signal: the open quotient showed no difference, but there was a difference in the steepness of the slope of the closing phase of vocal fold vibration with steeper slope being found for the lax vowels (here the misnomer effect again!). The important point that Jessen makes in his discussion, however, is that there is overall very little justification for regarding this as an *active* laryngeal adjustment: firstly, it has been suggested that advancement of the tongue root can cause some weakening of the adductive force at the arytenoids; secondly, a strong constriction in the vocal tract as found for high tense vowels can also have some influence on the vibratory pattern of the source in the direction of weakening the closing phase (see Jessen for further discussion and references)[28].

6.1.2 Touching on perception

Returning to the question of perception, a first pilot experiment was carried out in which a pair of naturally spoken pseudowords with a tense-lax vowel contrast (**iː** vs **ɪ**) were manipulated so that the relative F0 of the pair varied over a range of +/- 10Hz in steps of 2.5Hz.

---

[28]If an active laryngeal component were present where would it be found? Since we have found increased activity in the cricothyroid for the lax vowels, and since the cricothyroid does not contribute to adductive tension (if anything, it could weaken it somewhat), then most likely thyroarytenoid activity would also have to be higher in lax than tense vowels. Testing this electromyographically has not yet been possible.

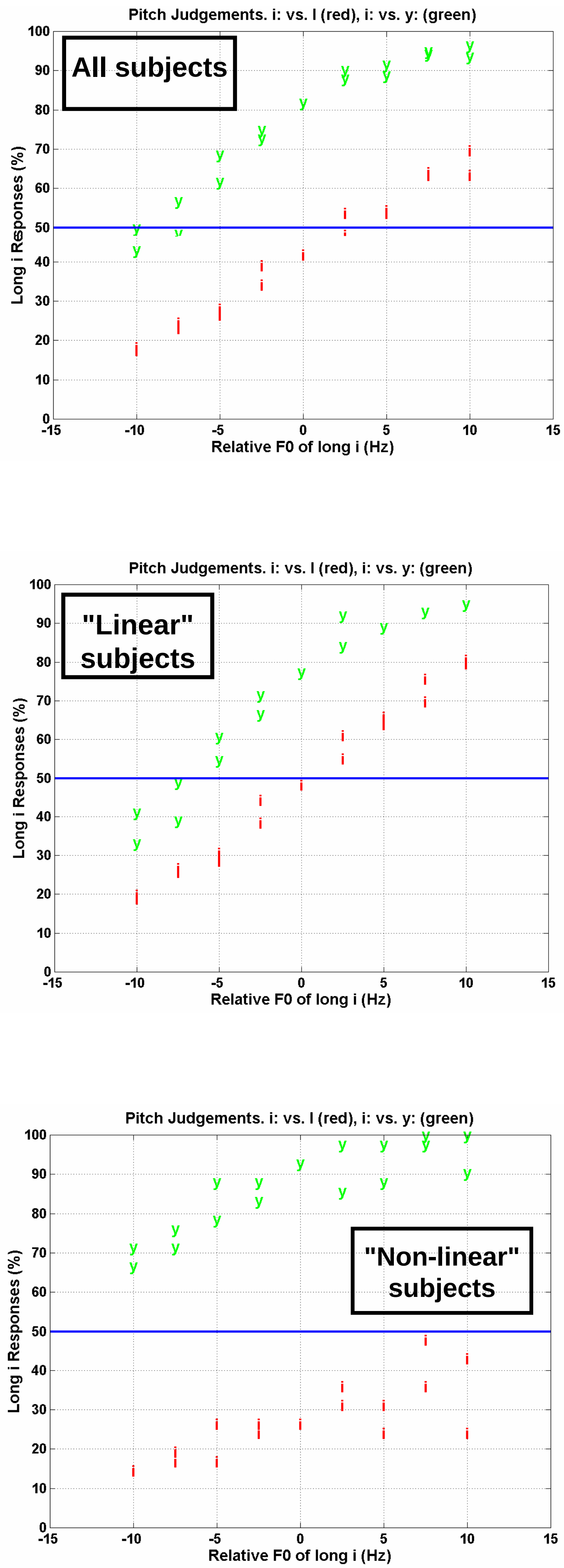


Fig. 6.1: Results of perception experiments into relative pitch of two types of vowel pairs: /**i**ː/-/ɪ/ (responses labelled with 'i') and /**i**ː/-/**y**ː/ (responses labelled with 'y'). Complete listener group in top panel; subdivided into "linear" and "nonlinear" group in middle and bottom panels. See text for details

Listeners were asked to judge which member of the pair sounded higher. In the following figure (top panel) the results correspond to the data points plotted with the symbol “i”.

If one looks only at these data points the result is superficially quite exciting: at equal F0 (0 on the x-axis) the number of stimuli in which the tense member of the pair sounded higher amounts to less than 50%. Looking along the x-axis to the intersection with the responses, one finds that /**iː**/ has to be a few Hertz higher than /**ɪ**/ to be perceived as having the same pitch. The effect is clearly not enormous, but it is in the hypothesized direction.

Following the pilot experiment which involved only /**iː**/ - /**ɪ**/ pairs it was pointed out by C. Mooshammer that the design should include a control condition as well. For this reason the test was repeated with both /**iː**/-/**ɪ**/ pairs and /**iː**/ - /**yː**/ pairs. Since the second pair involves just high tense vowels for which we assume that the laryngeal activation is the same at the same F0, then if listeners are indeed basing their responses on their interpretation of the underlying laryngeal activity, then the 50% point should be at 0 Hz. The “y” symbols in the figure show the result of including this control condition. Clearly there is a much larger shift than for the tense-lax pair (in effect, there is a strong tendency for /**yː**/ to be heard as lower in pitch). The source of this shift for /**iː**/-/**yː**/ is an interesting question in its own right, but will not be followed up here as it is not germane to the question at hand (see Pape et al., 2005, for further discussion). The point is that large shifts can occur; these cannot however be attributed to hearers’ perception of the underlying laryngeal behaviour, so clearly it would be hazardous to stick to the claim that the slight shift for /**iː**/-/**ɪ**/ reveals such an effect. The situation for the original hypothesis became even less promising after D. Pape looked closer at the individual response patterns of a larger group of hearers. This suggested that the listeners’ behaviour fell into two basic types. One group, labelled “linear subjects” (middle panel) clearly reacted systematically to the magnitude of the pitch differences between members of a pair. Unfortunately, for this group the intersection of the /**iː**/-/**ɪ**/ response curve with the 50% line does not differ significantly from 0Hz. The other group, labelled “nonlinear” (bottom panel) react only very weakly to the pitch manipulations in the stimuli, in effect making categorical judgements that /**ɪ**/ sounds higher than /**iː**/ (and /**yː**/ lower). However, since it must be questioned whether the experiment was successful in getting them to actually judge relative pitch at all, then it is also highly questionable if such behaviour is supportive of the hypothesis. The cautionary tale is that the original impression from the pilot data that the 50% intersection point shifted in the hypothesized direction was basically an artefact from averaging over two disparate groups.

There is clearly much more to be learnt about factors influencing the perception of pitch over different vowel categories (e.g see Pape et al., 2005, Pape & Mooshammer, 2006, for influence of musical education, and even native language), but for present purposes it suffices to note there is simply no evidence that a robust effect in the hypothesized direction is waiting to be teased out.

So the question remains as to what the perceptual effects of the articulatory adjustments could be: One possibility considered in a further pilot experiment was that the high F0 of lax vowels may help to distinguish not immediate tense-lax pairs like /**iː**/-/**ɪ**/ but rather such pairs as /**eː**/-/**ɪ**/, the formant frequencies being much more similar for the second pair than the first one. In stimuli where the duration difference between /**eː**/ and /**ɪ**/ was neutralized it appeared that further manipulation of F0 could make /**ɪ**/ sound more like /**eː**/. However, the ecological validity of such stimuli is dubious: one may question whether this kind of enhancement is really useful to listeners, given the clearly documented differences in the dynamic properties of tense vs. lax vowels[29].

6.1.3 Final perspectives

What, then, could be the motivation for speakers to raise F0 on lax vowels. Here the most down-to-earth explanation is perhaps the best one. This is that the tendency towards equalization of F0 on corresponding tense and lax vowels serves to indicate that the members of a pair do indeed belong together, that they are perhaps different prosodic expressions of a common category. Put another way, the F0 relationships underline the fact that, for example, the members of a pair like /**iː** - **ɪ**/ though differing radically in tongue height nevertheless form the highest representatives of their respective vowel series.

A final perspective could also fit the results elegantly into our previous work on the articulatory manifestation of the tense-lax distinction. In Hoole & Mooshammer (2002) we developed the idea of lax vowels as being characterized by pulsatile force input and tense vowels by distributed force input. This was based on analysis of the patterns of acceleration of the articulators over CVC movements (see there for more details). This was intended to emphasize in turn that traditional tense-lax terminology is indeed a misnomer: In this view, the reason for the centralization of lax vowels is not that the short duration does not allow sufficient time to reach a more extreme position; on the contrary the centralized location makes for a particularly strong modulation of the signal in moving from C to V and back to C (for many tense vowels the distance between vocalic and consonantal constriction can be extremely small). This leads to a single strong acceleration peak located squarely in the centre of the vowel. Assuming that speakers and hearers are very sensitive to the force patterns underlying speech utterances it would not be unnatural for this supraglottal articulation pattern to be enhanced by propagation to other speech subsystems, e.g the laryngeal one. Thus we closed the 2002 paper with the supposition - as it had to be at that time - that resolution of the German intrinsic pitch puzzle via higher laryngeal activity on lax vowels would represent precisely this coherent pattern of activity.

Within this perspective - and as an outlook for future work - there is no particular reason to expect the mechanisms observed here to also be present in English, for example, because the phonetics of English vowels are radically different from German. This is where the use of labels like “tense” vs. “lax” for both these languages can be genuinely misleading, coupled with the fact that familiarity with English tends to dull one’s perception for the fact that its vowel articulation is quite exotic: In many dialects there are practically no vowels anywhere near a cardinal vowel quality, and diphthongization is a very salient feature. Thus the very clean division in German between a peripheral, long vowel series and a centralized, short one is considerably muddied in English. Accordingly, there is probably no particular advantage in English to articulatory strategies that emphasize the pulsatile nature of the “lax” series. A more interesting comparison would be with a pure quantity language. If the perspective put forward here is correct that increasing laryngeal activity for lax vowels makes sense within the context of the dynamic organization of the German vowel system in turn within the context of German syllable structure, then this should *not* be observed on the short vowels of a quantity language, i.e in a language

[29]For further perception experiments related to the syllable-cut concept in German see Greisbach, 2001, Chapter 3.3.

where short duration is not accompanied by marked differences in tongue position and in syllable structure.

6.2 Vowel-specific effects

We have already indicated above in the brief discussion closing the section on vowel-dependent EMG and F0 (p. 96) 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. (1999) and look closely at the properties of the function linking EMG to F0 over the different vowel categories.

This showed - actually even more clearly than in the tense-lax case - that a mechanical effect must be at work. Nevertheless, the results also showed that an active difference in cricothyroid activity between high and low vowels is certainly also to be reckoned with, thus supporting findings made in the literature reviewed in Chapter 2 (though note that there was one clear case in our results of the differences not going in the expected direction of more activity for higher vowels).

Kingston has argued that many cases of covariation - in his view certainly the one here - 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 chose 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 (‘go with the flow’, cf. p. 8).

In this perspective we do not necessarily assume that speakers are initially aiming to enhance a specific property like F1-F0 distance. Possibly this could provide reinforcement when such a covarying pattern starts to emerge. Recall, however, Whalen’s point from the discussion of Chapter 2 that simple reliance on a property like F1-F0 can also have its disadvantages (p. 14).

As repeatedly emphasized, it would be beyond the scope of this work to aim at any definitive statements on perceptual organisation. For further discussion see, for example, Kingston’s (1991) extensive attempt to show that covarying articulations can merge into a single perceptual object.

In short, we suggest that the driving force behind this pattern of covariation is the articulatory contingency, and not the auditory perceptual goal, but that as suggested by Honda & Fujimura (1991; see p. 8 above) speakers may latch onto these effects and push them further as part of their linguistic behaviour.

6.3 Consonant voicing

6.3.1 Introduction: The open questions

We have seen that voiceless consonantal contexts reliably attract more cricothyroid 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 cricothyroid 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: Can we simply say that it helps to suppress voicing? And are the differences in F0 on the vowel sufficiently accounted for by the observed EMG activity?

6.3.2 Two conflicting points of view and a possible third

A convenient point to start discussion is with Kingston & Diehl (2004) since they raise some important points regarding the interpretation of the crucial paper of Löfqvist et al. (1989; refer back to p. 28ff). The thrust of Kingston & Diehl's argument is essentially the same as that for vowel intrinsic pitch: they see F0 as something controlled by speakers as part of their implementation of the phonological feature [voice] and not as an automatic consequence of the consonant articulation (the auditorily based impetus is in this case the enhancement of low frequency emphasis for [+voiced]). Löfqvist et al. take the opposite view; they see the cricothyroid as related to devoicing, but at the same time as accounting for F0 on the following vowel. Since we believe that there are some inconsistencies in the way both points of view are presented we need to go into some detail in order to arrive at what is hopefully a more water-tight scenario (ultimately again with similarities to the intrinsic pitch discussion: i.e a hybrid perspective).

We will use a lengthy quote from Kingston & Diehl to get the relevant issues up front:

> *Regarding differences in CT activity, Löfqvist et al.'s 1989 data, as well as those in Löfqvist at al., 1984, show that the amount that F0 is elevated in the following vowel does not depend on how long the temporal interval is between the end of the preceding vowel when the elevation of CT activity occurs and the beginning of the following vowel. F0 is elevated as much when the elevated CT activity is further away, as a result of longer intrinsic consonant duration or of a larger number of intervening consonants, as when the CT elevation is closer. Even given a latency of as much as 80ms between an increase in CT activity and an increase in horizontal fold tension (Collier 1974, Atkinson 1978, Baer 1981; Baer [personal communication, 1989] suggests a much shorter latency, of 20-50ms), the intervals between when CT becomes active at the end of the preceding vowel and when the following vowel begins appear to be too long for the contraction of this muscle to be responsible for F0 elevation in the following vowel in many of the utterances they consider.(pp.439f)*

This passage is immediately supplemented by the following footnote:

> *Löfqvist at al. (1989) suggest that the latency between CT contraction and F0 elevation would explain why the preceding vowel's F0 is not raised as much before a [−voice] stop as a following vowel's F0 is raised after one. But this would also mean that the increase in horizontal tension brought about by*

> *contracting CT would be too late to contribute much to extinguishing voicing at the beginning of the [−voice] stop closure when A_g [area of glottal opening] is still small. (p. 440)*

Let us look first at the argument that the CT activity is too far away from the vowel. This is a two-edged sword. Löfqvist at al. themselves point out that voicing-related differences are minimal in the vowel segment itself (in the figure from Löfqvist at al reproduced above as Fig. 3.4 on p. 29 this can be observed particularly neatly for subject NSM: assuming a delay of roughly 50ms between EMG and audio output the differences converge to about zero around the audio amplitude maximum for the vowel before diverging again for the following voiced or voiceless consonants). However, in order to round off their own argument Kingston & Diehl need the speaker to be actively doing something different on the vowel and it remains unclear where this then is to be observed. I think there are also several reasons why the contention that the putatively consonantal EMG activity is too far from the vowel needs moderating. For the Löfqvist at al. (1989) data I do not see any particular problem. Assuming an EMG delay of 50ms (which as we see again is admittedly a contentious matter to be returned to below) then it would appear that differences in the CT traces extend up to about the time that voicing would restart after the voiceless consonants. In other words differences in the CT traces roughly parallel the time-course of the laryngeal abduction-adduction gesture, a point that will also become important below. This again can be observed quite neatly in the Löfqvist at al. data: For the English speakers TB and NSM the CT traces rejoin later 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 LB this difference between fricatives and stops is not apparent (the voiceless stops being essentially unaspirated). Kingston & Diehl's reference to the earlier Löfqvist at al. (1984) paper is extremely cogent, but unfortunately I find the data very difficult to interpret. This paper briefly reported data for a single speaker of Swedish, the interesting point being that not only single consonants were recorded, but also sequences of voiceless consonants. Thus it is extremely relevant to ask where differences in CT activity are concentrated over these longer sequences, i.e whether they persist throughout or are more weighted to the beginning or end. It does appear that CT starts to rise at the start of the longer voiceless sequences but it is not easy to judge based on the plots given how far differences relative to a voiced control condition extend. Also I find it very difficult to see any difference between single /p/ and /b/ at all, making it in turn difficult to judge how far one can push this CT data. It certainly cannot be denied, however, that more information on CT behaviour in long voiceless sequences would be very useful.[30]

Assuming for the sake of argument that the relevant CT activity is at least not radically disjoint from the start of the vowel, then it is true that we still need a mechanism that will allow it to propagate its effects further into the vowel. It is not quite clear whether Löfqvist at al. see this as taken care of by the delay between EMG and its effect on F0. As Kingston & Diehl observe (footnote quote) this raises the logical problem that EMG would then be too late to suppress voicing at the consonantal onset (again a point to which we will have to return). In effect we have incorporated an estimate of the delay in our discussion above. However, there is a further aspect

[30]We have carried out an EMG recording of a Berber speaker which included CT as one of the muscles where one of the aims was to investigate the very long voiceless sequences that can occur in Berber, but it has not yet been possible to perform the analysis.

to the question of delay which is not picked up by Kingston & Diehl but which is mentioned briefly by Löfqvist at al. and which in our view could be crucial: this is that the delay for the effect of relaxation could be longer than the delay for the effect of activation. We showed earlier a figure illustrating the effect that following a rise in CT (and F0) the decline in F0 appeared to lag that in CT quite considerably. We will not get here into the issue of how F0 can be actively lowered (but see e.g Ohala, 1978; Erickson et al., 1983; Halle, 1994; Honda et al., 1999,). The point is simply that it appears plausible that once CT has caused a rise in F0 then comparatively little activity may be required to maintain it at the same level. In support of this we could quote a paper by Honda et al. (1995) which looked at procedures for predicting lip configurations from EMG data. They found that the predictions worked best when the position data was transformed to an 'incomplete first derivative', essentially a weighted sum of the measured signal and its first derivative. If one thinks in terms of an equation of motion with elastic, viscous and inertial terms (proportional to displacement, velocity and acceleration, respectively) then EMG can only be expected to be simply proportional to displacement when velocity is low. Honda et al. apparently found for their purposes that it was possible to neglect the acceleration-related term (for an early application of these concepts to speech motor control see Abbs & Eilenberg, 1976). In the present case we are confronted with short segmental changes usually superimposed on rapidly changing CT activity related to the intonation contour, so the velocity term is probably not negligible. We have no means at our disposal of readily incorporating this viscous-related term into a specific model prediction. The intention was simply to suggest where one could look for a possible scenario to generate the required propagation of CT effects in the consonant onto the following vowel. There is also a further specific indication from our data analysis that an effect of the required type is actually observable. When the regression analysis was carried out above to link EMG and F0 for voiced and voiceless consonants it was observed that F0 for voiceless consonants is actually predicted to be *lower* than for voiced at a comparable EMG level (see Figs. 5.21 and 5.25 on pp. 85 and 94). Put another way, the EMG level we can observe in the vowel following the voiceless consonant would lead us to underestimate the F0 actually occurring. Thus there must be an additional effect present contributing to a raising of F0. This could be an effect of the kind just discussed, i.e the effect of CT activation persists somewhat beyond the time when it has started to decline again. There is possibly another important candidate for such an effect which it is convenient to discuss here. 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 the Löfqvist at al (1984) study just quoted, and also for the aspirated plosives in the Hutters (1985) study (see Fig. 3.3 on p. 27). 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. It is impossible to say which of the two explanations just offered is more probable (CT hysteresis, or vocalis 'kick-in'), quite possibly some combination. The key point is that it seems to be possible to find effects that are undoubtedly consonantal in origin, but whose manifestation can emerge in the vowel. Note that these are also the kinds of explanations (rather than just EMG-F0 delay per se) that accord with the apparently robust observation (alluded to in Kingston & Diehl's footnote) that it is F0 of the post-consonantal rather than pre-consonantal vowel that is most affected. Kingston & Diehl are, however, quite justified in pointing out in their footnote that there is still a tension between the assumption that the purpose of the increased vocal fold tension is to assist in stopping vibration at the *onset* of the consonant and the location of the F0 effects *after* the consonant. We will now try and develop two further points of view that may help to resolve this tension. Under what now seems to be the traditional interpretation of the original Halle & Stevens suggestion that the increased vocal fold tension helps to suppress vibration then a logical expectation is that the mechanical effect of the EMG activity should peak around the onset of the consonantal occlusion. This is where the activity would contribute most to the suppression, since at this point in time the glottis has not opened very much and the intraoral air-pressure has also not risen very much. Thus these latter two factors would not yet be contributing strongly to devoicing. If, however, the maximum effect is reached later, say around the time of peak glottal opening, then it would be essentially useless as a direct contributor to voicing suppression, since by this time the effects of both glottal abduction and intraoral airpressure will be very strong. In fact, if our estimate of the delay between CT activity and its mechanical effect is realistic (a big caveat, admittedly) then we believe that the evidence both in Löfqvist at al. (1989) as well as our own data is that these mechanical effects quite closely follow the time-course of glottal abduction and adduction. This was indicated in the Löfqvist at al (1989) data by the fact that CT timing seemed to follow what is known about about differences in glottal devoicing gesture timing for fricatives vs. aspirated plosives with respect to the completion of the gesture. In our data a similar effect was pointed out for speaker SF (p. 58). There would be no particular reason for such a match if the crucial timepoint is the onset of the consonant. Referring back to our EMG ensemble averages (Figs. 5.1 and 5.2 on p. 56f), which take the estimated EMG delay into account and also mark the start of the oral occlusion for the consonant, it will be seen that for both speakers CK and SF there is absolutely no indication that the activity peaks around consonant onset. This is not the case for CG: for him the difference between voiced and voiceless consonants does seem to be close to its maximum here. It turns out, though, that this actually confirms in a very interesting way the contention put forward here that CT activity follows the course of glottal abduction and adduction: Both CG and SF were subjects in the experiments forming Part II of this monograph. It emerged as a very stable feature of the glottal timing of CG that glottal abduction started earlier relative to the formation of the oral occlusion than was the case for the other speakers, thus parallelling the timing of the EMG activity.

6.3.3 Consonantal articulation with vocalic effects: A shift of emphasis

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 not primarily to suppress vocal fold vibration (though it will still contribute to this indirectly) but rather to increase the mechanical efficiency of the abductory motion of the arytenoids with respect to the resulting glottal aperture.

This shift in emphasis as to the purpose of the CT adjustments could well make communicative sense, too. Given that onsets are perceptually more salient than offsets, it is probably not particularly important exactly how vocal fold vibration dies out at the VC transition (at least as long as voicelessness is in fact achieved when linguistically required). On the other hand, a clear glottal opening at release is clearly crucial for an aspirated plosive (and glottal opening is also a crucial requirement for a voiceless fricative). Support for glottal opening is probably also important in voiceless unaspirated plosives (when they form the voiceless category in a voicing contrast); even though the glottis will be more or less closed again by the time of the release the opening of the glottis during the occlusion phase will allow intraoral airpressure to rise quicker (than for a fully voiced plosive) and thus stabilize at a higher level by the time the release occurs, and thus in turn contribute to (perceptually salient) acoustic events occurring in the vicinity of the burst.

An example of laryngeal kinematics taken from the data collected for Part II below may help to support the contention that the offset of voicing is not the crucial locus of cricothyroid activity. Following that, we will move to the last piece of evidence in this discussion and look at the whole topic from a different point of view, namely in terms of modulation of vocal fold tension as a useful characteristic not at the *offset* of voicing in the VC transition but rather at the resumption of voicing in the CV transition.

The following figure shows an example of an utterance spoken by subject CG containing the target word ‘Type’ (/**tyːpə**/).

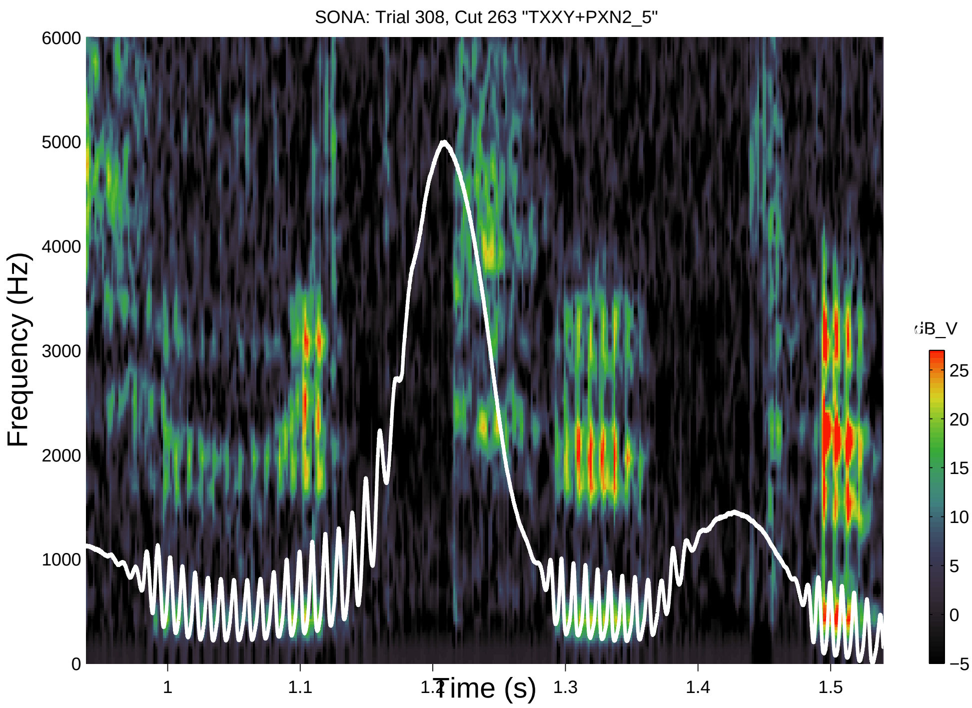


Fig. 6.2: Example of transillumination signal for subject CG. Target word ‘Type’ (together with part of carrier phrase). The large glottal abductory movement corresponds to initial prestressed /t/, the second smaller one to post-stress /p/

The first consonant, in pre-stress position, is voiceless and clearly aspirated, the second consonant, in post-stress position, is also voiceless but only weakly unaspirated. The transillumination signal (PGG) gives a measure of glottal aperture over time. Unlike the examples shown in Part II the signal has not been low-passed filtered so modulations of the signal caused by vocal fold vibration can be observed. Looking first at the pre-stress consonant it will be seen that residual vibrations of the vocal folds continue for some distance into the overall abductory movement. This is precisely the region where one might expect CT activity to be useful in swiftly suppressing vibration as the arytenoids start to abduct. So the impression here is that either CT activity is not very effective for this purpose, or this is simply not what the speaker is trying to accomplish. As already indicated, these residual vocal fold vibrations are probably extremely unobtrusive auditorily, given that the mouth is shut, and the glottis is no longer closing firmly. Turning to the post-stress consonant, there is a massive reduction in the overall size of the peak glottal opening (the synchronized video-films nevertheless clearly reveal that abduction at the arytenoids is still taking place). Moreover, for this speaker the EMG data indicated only a very weak cricothyoid difference between voiced and voiceless consonants in post-stress consonants, which we thus assume could well be the case here, too. In fact, there is not much difference between the pre- and post-stress case in terms of the residual vibration following onset of the occlusion (at least in relation to the huge laryngeal differences that are clearly present). And in the post-stress case voicelessness is nonetheless clearly achieved. Thus CT activity does not seem to be a crucial element in achieving voicelessness[31]. From this we derive the suggestion that it comes more into play when not just voicelessness is required, for which apparently only a small amount of glottal abduction is required, but rather when a sizeable glottal aperture is important[32].

It should be emphasized that we do not intend to suggest that the link between CT activity and modulation of glottal width is hardwired in any sense. First, while subject CG may show quite a close parallelism between amount of glottal abduction and amount of CT activation over the pre- and post-stress consonants, the picture for subject SF may be somewhat different. She probably shows at least as much reduction of glottal width in post-stress position (see Fuchs, 2005, for relevant transillumination data), but reduction in CT differences between voiced and voiceless from C1 to C2 is only apparent in lax vowel contexts. Secondly, we discussed above the case of the voiced aspirated consonants in Hindi (p. 24), where level of CT activity is clearly particularly low despite the presence of glottal abduction[33]. Thirdly, it will emerge immediately below from the final section of this discussion that speakers may be able to exercise deliberate control over the precise timing relationship between vocal fold tension adjustments and glottal width adjustments. The basic idea being put forward here is that these two components develop as a useful synergy for voiceless consonant production requiring a clear glottal aperture. Whether

---

[31]We have not tried to weigh up here all the factors impinging on how and when voicing occurs, in particular the very realistic possibility that voiced and voiceless plosives may also involve differences in the compliance of supraglottal tissue (see e.g Westbury, 1983; Svirsky et al., 1997).

[32]A supplementary motivation for the present EMG investigation of voicing in fact came from the frequent observation of very restricted glottal abduction but nonetheless consistent devoicing for post-stress stops in German speakers (the laryngeal kinematics of pre- and post-stress stops is not actually a topic in Part II of this work, but has been treated in detail by Fuchs, 2005). It was speculated that in such a situation CT activity might actually be particularly important for ensuring voicelessness. This now seems unlikely. Although CG is the only speaker who shows a consistent reduction of the voiced-voiceless difference for C2 compared to C1 (see Fig. 5.6), there is, on the other hand, no indication from the other speakers that CT activity can actually be enhanced on C2 as a kind of compensation for a weakening of glottal abduction.

[33]This actually illustrates very neatly that cricothyroid most likely does influence the ease with which the vocal folds can vibrate: The nominally voiced aspiration phase in voiced aspirated plosives is aerodynamically unstable and could easily become voiceless in unfavourable conditions. The point here has been that tension regulation is probably not particularly useful at the *onset* of normally voiceless consonants.

and how they generalize to other situations where sound production is voiceless may then vary over speakers.

In a paper devoted to the use of quasi-articulatory parameters to control the source generation in the speech synthesis system HLSyn Hanson & Stevens (2002) discuss the control of voicing in voiceless aspirated and (partially) voiced consonants. For the voiceless aspirated case they concentrate in particular on parameter **dc** (delta compliance), which is used to capture relative change in the compliance of the vocal folds[34]. They find it appropriate to modulate dc roughly in paralllel with glottal opening (i.e reducing the compliance during the occlusion), 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 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 (i.e a mode of phonation consistent with essentially complete closure of the glottis)[35]. In support of their interpretation of this pattern of events Hanson & Stevens quote an inverse filtering study by Ní Chasaide & Gobl (1993). The latter authors examined aspirated plosives in several languages, including German, and came to the conclusion that voicing is typically not initiated while the glottis is still appreciably open. Curiously, German seemed to pattern differently from the other languages, making less use of the possibility for delaying voice onset, and thus producing rather breathy voicing at the voiceless-voice transition. Whether this really is a consistent property of German is difficult to say since - in view of the time-consuming nature of inverse filter analysis - only a single utterance with aspirated plosive was analyzed for each of the four speakers of each language. The important take-home message from this study is, however, that speakers can be assumed to be in a position to deliberately control the precise acoustic properties of the transition from voiceless to voiced via the control of vocal fold tension.

Note that Hanson & Stevens do not refer at all to the modulation of the compliance as an aid to suppression of voicing at the start of the consonantal occlusion. This is consistent with other predictions of their model: In an additional example illustrating synthesis of a partially voiced American English /g/ it is found that with no glottal abduction and no change of compliance from the default value voicing still stops quite rapidly at formation of the occlusion. This is consistent with the empirical observations made above. In order to produce appreciable voicing into the closure it is necessary to increase the compliance and also modulate a further parameter in their model that is responsible for capturing active expansion of the vocal tract.

---

[34]also of the vocal tract walls, though the discussion of this example focusses on effects directly related to the vocal folds.

[35]This might also be seen as a good illustration of Stevens' quantal concept (e.g Stevens, 1989): Continuously varying articulations can result in a very abrupt transition in the properties of the acoustic output; this can be assumed to generate particularly salient information for the perceiver.

6.3.4 Approaching a conclusion

In the preceding paragraphs we have aimed to develop a scenario for the emergence of F0 differences on the vowel as a pattern of activity that is closely related to the abductory-adductory movement for voiceless consonants, emphasizing in particular the mechanisms that could account for propagation of the effect onto the ***post***-consonantal vowel.

This contrasts with the view of Kingston & Diehl who use data from a large number of languages to support the contention that the patterning of F0 differences can only be captured by a more abstract [voice] specification:

> *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 (p. 432).*

Particularly crucial here are those sounds that emerge auditorily as voiceless unaspirated:

> *.... voiceless unaspirated stops which realize a member of a [voice] contrast may both elevate and depress F0 ..... This result is paradoxical only if one ignores the phonological specification of this phone: when it represents the [−voice] category, as in Hindi, Thai, Spanish, French, Portuguese, Italian, and Japanese, the F0 is elevated; but when it represents the [+voice] category, as in English, German, Swedish, Danish and Korean, then F0 is depressed (p. 435).*

In fact, I do not think an appeal to the phonological specification is necessary to resolve the paradox: the two groups of languages probably quite simply differ in the management of the glottal width, even if the resulting acoustic output has considerable similarities: languages of the first type probably have an abductory movement at the arytenoids, even if in some of these languages it is not always very large, whereas languages of the second type do not[36].

A potentially more interesting test case involves languages of China such as Mandarin and Cantonese which contrast a robustly voiceless unaspirated plosive with voiceless aspirated. Here it seems possible that laryngeal abduction for the voiceless unaspirated is very restricted indeed, but the amount of parallel information on glottal adjustments and F0 is extremely meagre, so firm conclusions can hardly be drawn.

Kingston & Diehl discuss two further cases where in their view the presence of an abstract [voice] contrast is crucial for the presence of F0 differences in adjacent vowels. However, we believe - taking the same tack in both cases - that the situation may depend quite mundanely on the timing of the glottal abduction gesture.

---

[36]Danish is probably a slightly different case from e.g English and German. Hutters (1985) has documented some posterior cricoarytenoid activity for the Danish voiceless unaspirated consonant, and refers to it as an active devoicing gesture (even though the resulting glottal opening appears to be very small). Interestingly, although F0 is lower in the unaspirated case the effect appears to be weaker than in languages like English and German: The Danish unaspirated consonants generally have higher F0 than the fully voiced consonants (see Reinhold Petersen, 1983). Thus this pattern appears consistent with our model. Hutters found slightly higher CT activity for aspirated than unaspirated stops: What we admittedly do not know is whether the unaspirated in turn have higher CT activity than the fully voiced consonants.

The first case involves F0 following English [s+stop] clusters, where F0 is apparently very variable and often intermediate between the standard simple voiced and voiceless stop cases. Kingston & Diehl consider this as a prediction from the fact that here the [voice] contrast is neutralized and can no longer exert any control over F0 (p. 436). A more down-to-earth explanation is based on findings regarding glottal timing made in Part II of this work. We found that the three German speakers completed glottal adduction very early (relative to release of the plosive) in such clusters. For aerodynamic reasons, in this kind of sound sequence speakers probably have a great deal of freedom in precisely when the glottal adduction is completed (as long as it is completed roughly by the time of release) since in all cases voicing will simply start 10ms or so after the release when the intraoral pressure has declined sufficiently. The only acoustic trace of this freedom of timing will in fact be the variability of F0, if our assumption is correct that modulation of CT activity and glottal width proceed more or less in parallel.

The second case involves a contrast in Tamil between voiced simple stops and voiceless geminates. F0 differences following the two categories appear to be very slight. Kingston & Diehl argue for an analysis in term of a contrast of [length] rather than [voice]; the absence of the latter then predicts the F0 equalization. However, similarly to the cluster case above, we would suggest that in voiceless geminates, too, speakers have considerable freedom when the glottal abduction-adduction cycle is completed; even if adduction is completed some time before release it is very unlikely that voicelessness will be compromized.

Accordingly, in both cases it is difficult to assess what argumentative use can be made of the F0 patterns if precise information on glottal width is not available.

6.3.5 Final conclusion: Sitting on the fence

The bulk of the discussion in this section on the consonantal voicing contrast has been devoted to making a case for F0 differences on the vowel as emerging from typical articulatory patterns for the preceding consonant. However, the ultimate conclusion we would like to reach is exactly equivalent to that for intrinsic pitch effects: The driving force comes from the articulatory contigency, but once established speakers can deliberately emphasize its effects[37]. There was evidence of this in the ensemble averages of speakers CK and SF (Corpus 1) where at least in the case of the tense vowels the traces stay separated right through the duration of the vowel. We cannot completely exclude the possibility that this is simply an overlapping of effects from the preceding and following consonant, however in Löfqvist et al.'s data (where pre- and post-vocalic consonants were also identical in voicing status) a convergence of the voiced and voiceless traces in the centre of the vowel generally occurred. So it seems possible that our speakers on these occasions are actively enhancing the difference during the vowel segment. Nevertheless it would clearly be useful in future work to also include target items with alternating voicing for C1 and C2 (i.e C1 = voiceless, C2 = voiced, and vice-versa). On the lax vowels this effect was not observed. Differences between voiced and voiceless contexts were negligible. A possible reason for this often-mentioned TENSE*VOICE interaction will now be the subject of the final section.

[37]In defence of more sitting on the fence: It can be unconfortable and is also a highly skilled motor activity.

6.4 The TENSE*VOICE interaction: A cautionary tale in experimental design

We saw throughout this study that voicing effects on the consonant tended to be more easily observable in the context of tense vowels, and tenseness effects on the vowel more easily observable in the context of voiced consonants.

It was essentially fortuitous that we combined these two independent variables in this experiment, basically because of the wish to extract maximum usable material given the technical difficulty of conducting EMG experiments. It is sobering to consider what might have happened if we had decided to concentrate on just one of these independent variables. A balanced picture would not have emerged; the strength of the effect would either have been overestimated or underestimated depending on whatever level of the other factor we had happened to use to construct the linguistic material.

Is it possible to explain what lies behind this interaction? The basic situation is that we have two factors that seem to be potentially linked to higher CT activity: voicelessness of the consonant and laxness of the vowel. The interaction indicates that when they occur together they do not combine additively to give even higher CT activation. One very speculative suggestion as to why this happens can perhaps be derived from our regression analyses of the relationship between F0 and CT activity. The slope of the relationship was shallower for both lax vowels and voiceless consonants. This means that when either of these properties is present a further rise in F0 requires a relatively large amount of additional CT activity. If we assume that speakers have a very precise model of the distal effects of their muscular activity then one could speculate that they dispense with a further increment in CT activity when its effectiveness for further modulation of F0 is limited.

Whether this speculation could ever be tested is unclear. The final conclusion is quite simply that paring an experimental design down to the barest minimum may not be an advisable research strategy.

7 References

Abbs, J. H. & Eilenberg, G. R. (1976). "Peripheral Mechanisms of Speech Motor Control". in Contemporary Issues in Experimental Phonetics, edited by N. J. Lass (Academic Press, "Perspectives in Neurolinguistics and Psycholinguistics", N. -Y.), pp. 139-170.

Abramson, A. & Lisker, L. (1985). "Relative power of cues: F0 shift versus voice timing". in: Linguistic Phonetics, V. Fromkin (ed.), Academic, New York

Alfonso, P. J., Honda, K., Baer, T. & Harris, K. S. (1982). "Multichannel study of tongue EMG during vowel production". Paper presented at the 103rd meeting of the Acoustical Society of America, Chicago, Il..

Ansre, G. (1961). "The tonal structure of Ewe". Hartford Seminary Foundation.

Atkinson, J. & Erickson, D. (1977). "The function of the strap muscles in speech: Pitch lowering or jaw opening". Haskins Labs Status Report on Speech Research, 49, 97-102.

Atkinson, J. E. (1978). "Correlation analysis of the physiological factors controlling fundamental voice frequency". J. Acoust. Soc. Am., 63, 211- 222.

Atterer, M. & Ladd, D.R. (2004). "On the phonetics and phonology of "segmental anchoring" of F0: evidence from German". J. Phonetics, 32(2), 177-197.

Autesserre, D., Roubeau, B., Di Cristo, A., Chevrie-Muller, C., Hirst, D., Lacau, J. & Maton, B. (1987). "Contribution du cricothyroidien et des muscles sous-hyoidiens aux variations de la frequence fondamentale en francais: Approche electromyographique". Proc. 11th International Congress of Phonetic Sciences, Vol. 3, pp. 35-38. Tallinn: Academy of Sciences of the Estonian SSR..

Baer, T. (1981). "Investigation of the phonatory mechanism". Proceedings of the conference on the assessment of vocal pathology, C.L. Ludlow & M. O'Connell Hart (eds.) = ASHA Reports 11, pp. 38-46.

Basmajian, J.V. (1967). "Control of individual motor units". American J. Physical Medicine, 46, 480-486.

Basmajian, J.V (1962). "Muscles Alive. Their functions revealed by electromyography (5 ed. = Basmajian, J.V & DeLuca, C.J., 1985)". Williams and Wilkins, Balitmore.

Basmajian, J.V; Stecko, G. (1962). "A new bipolar electrode for electromyography". J. Appl. Phys., 17, 849.

Becker, T. (1998). "Das Vokalsystem der deutschen Standardsprache". Peter Lang, Frankfurt am Main.

Brooks, V. B. (1986). "The Neural Basis of Motor Control". Oxford University Press: New York; Oxford.

Chiba, T. & Kajiyama, M. (1958). "The vowel: its nature and structure". Tokyo: Tokyo-Kaiseikan Publ. Co. -Second edition published by Phonetic Society.

Clancy, E.A., Morin, E.L.,; Merletti, R. (2002). "Sampling, noise-reduction and amplitude estimation issues in surface electromyography". J. electromyography and kinesiology, 12, 1-16.

Collier, R., Lisker, L., Hirose, H., & Ushijima, T. (1979). "Voicing in intervocalic stops and fricatives in Dutch". J. Phonetics 7, 357-373.

Connell, B. (2002). "Tone languages and the universality of intrinsic FO: Evidence from Africa". J. Phonetics, 30(1), 101-129.

De Cheveigne, A., & Kawahara, H. (2002). "YIN, a fundamental frequency estimator for speech and music". J. Acoust. Soc. Am., 111(4), 1917-1930.

Diehl, R. L. (1991). "The role of phonetics within the study of language". Phonetica, 48, 120-134.

Dixit, P. R., and MacNeilage, P. F. (1980). "Cricothyroid activity and control of voicing in Hindi stops and affricates". Phonetica, 37, 397-406.

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

Erickson, D., Baer, T. & Harris, K.S. (1983). "The role of the strap muscles in pitch lowering". In D. Bless & J. Abbs (eds.), Vocal Fold Physiology: Contemporary Research and Clinical Issues. College Hill, San Diego, pp. 279-285.

Erickson, D., Honda, K., Hirai, H. & Beckman, M. (1995). "The production of low tones in English intonation". J. Phonetics, 23, 179-188.

Ewan, W. G. (1979). "Can intrinsic vowel F0 be explained by source/tract coupling?". J. Acoust. Soc. Am., 66, 358- 362.

Ferner, H. & Staubesand, J. (eds.) (1982). "Sobotta, Atlas der Anatomie des Menschen, 18th ed.". Urban & Schwarzenberg. Munich, Vienna, Baltimore.

Fischer-Jørgensen, E. (1990). "Intrinsic F0 in tense and lax vowels with special reference to German". Phonetica, 47, 99-140.

Fowler, C. A. & Brown, J. M. (1997). "Intrinsic F0 differences in spoken and sung vowels and their perception by listeners". Perception & Psychophysics, 59(5), 729-738.

Fowler, C. A. (1981). "Production and perception of coarticulation among stressed and unstressed vowels". J. Speech and Hear. Res., 46, 127-139.

Fromkin, V. (ed.) (1978). "Tone: A linguistic survey". Academic Press.

Fromkin, V. A. & Ladefoged, P. (1966). "Electromyography in Speech Research,". Phonetica, 15, 219-242.

Fuchs, S. (2005). "Articulatory correlates of the voicing contrast in alveolar obstruent production in German". ZASPiL, 41, 1-238.

Fujisaki, H. (1988). "A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour". Vocal physiology: voice production, mechanisms and functions, Fujimura, O. (ed.). pp. 347-356. Raven Press, New York.

Fujisaki, H. (2004). "Information, prosody, and modeling - with emphasis on tonal features of speech". Proc. Speech Prosody 2004 Nara, 1-10.

Gandour, J. & Weinberg, B. (1981). "On the relationship between vowel height and fundamental frequency: evidence from esophageal speech". Phonetica, 37(5-6), 344-354.

Geiger, S. R. et al. (1981). "Handbook of Physiology; Section 1: The Nervous System; Volume I: Motor Control". American Physiological Society: USA.

Greisbach, R. (2001). "Testmethodik in Phonetik und Phonologie". Peter Lang, Frankfurt am Main.

Guérin, B. & Boë, L. J. (1980). "Influences of acoustic coupling between source and vocal tract of the Fo of oral vowels. Consequence for the study of intrinsic characteristics". Phonetica, 37(3), 169-192.

Halle, M. & K. Stevens (1971). "A note on laryngeal features". QPR, Res. Lab. Electr., MIT, 101, 198-213.

Halle, P. A. (1994). "Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese". Language and speech, 37, 103 - 125.

Hanson, H.M. & Stevens, K.N. (2002). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn". J. the Acoustical Society of America, 112(I), 1158-1182.

Herman, R., Beckman, M. & Honda, K. (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, M. & Ohala, J. (1969). "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles". J. Speech Language and Hearing Researsh, 12, 362-373.

Hirose, H. & Gay, T. (1972). "The activity of the intrinsic laryngeal muscles in voicing control. An electromyographic study". Phonetica, 25(3), 140-164.

Hirose, H. (1971). "Electromyography of the Articulatory Muscles : Current Instrumentation and technique". Haskins Laboratories Status Report on Speech Research, 25/26, 73-86.

Hirose, H. (1979). "Laryngeal EMG". ABRILP, 13, 13-22.

Hombert, J.-M. (1978). "Consonant types, vowel quality, and tone". In V. Fromkin (ed.), Tone: A lingistic survey, pp. 77-111.

Honda, K. & Fujimura, O. (1991). "Intrinsic vowel F0 and Phrase-final F0 lowering: Phonological vs. Biological Explanations". In: Gauffin, J. & Hammerberg, B. (eds) Vocal Fold Physiology, pp. 149-157.

Honda, K., Hirai, H., Masaki, S. & Shimada, Y. (1999). "Role of vertical larynx movement and cervical lordosis in F0 control". Language and Speech, 42 (4), 401-411.

Honda, K., Kurita, T., Kakita, Y. & Maeda, S. (1995). "Physiology of the lips and modeling of lip gestures". J. Phonetics, 23, 243-254.

Honda, K. (1983a). "Relationship between pitch control and vowel articulation". In D. Bless and J. Abbs, Vocal Fold Physiology: Contemporary Research and Clinical Issues, pp. 286-297, College-Hill, San Diego.

Honda, K. (1983b). "Relationship between pitch control and vowel articulation". In I. Titze and R. Scherer, Vocal Fold Physiology: Biomechanics, acoustics and phonatory control, pp. 127-137, The Denver Center for the Performing Arts, Denver.

Honda, K. (1998). "Interactions between vowel articulations and F0 control". In O. Fujimura, B. Joseph and B. Palek (eds.). Proceedings of LP'98. Charles University, Prague.

Honda, K. (2004). "Physiological factors causing tonal characteristics of speech: from global to local prosody". Proc. Speech Prosody 2004 Nara, 739-744.

Hoole, P. & Kroos, C. (1998). "Control of larynx height in vowel production". Proc. 5th Int. Conf. Spoken Lang. Processing, 2, 531-534.

Hoole, P. & Mooshammer, C. (2002). "Articulatory analysis of the German vowel system". In: P. Auer, P. Gilles & H. Spiekermann (eds.), Silbenschnitt und Tonakzente. Niemeyer, Tübingen, pp. 129-152.

Hoole, P., Mooshammer, C. & Tillmann, H.G. (1994). "Kinematic analysis of vowel production in German". Proc. ICSLP 94, Yokohama, 1, 53-56.

Hoole, P., Munhall, K., & Mooshammer, C. (1998). "Do airstream mechanisms influence tongue movement paths?". Phonetica, 55(3), 131-146.

Hoole, P. (1999). "On the lingual organization of the German vowel system". J. Acoust. Soc. Am., 106(2), 1020-1032.

Hutters, B. (1984). "Vocal fold adjustments in Danish voiceless obstruent production". ARIPUC, 18, 293-385.

Hutters, B. (1985). "Vocal fold adjustments in aspirated and unaspirated stops in Danish". Phonetica, 42, 1- 24.

Jessen. M. (2002). "Spectral balance in German and its relevance for syllable cut theory". In: P. Auer, P. Gilles & H. Spiekermann (eds.), Silbenschnitt und Tonakzente. Niemeyer, Tübingen, pp. 153-179.

Kingston, J. & Diehl, R. (1994). "Phonetic knowledge". Language, 70, 419-454.

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

Kingston, J. (1992). "The phonetics and phonology of perceptually motivated articulatory covariation". Language and Speech, 35, 99-113.

Kohler, K. J. (1985). "F0 in the perception of lenis and fortis plosives". J. Acoust. Soc. Am. 78, 21- 32.

Kroos, C., Hoole, P., Kühnert, B. & Tillmann, H.-G. (1997). "Phonetic evidence for the phonological status of the tense-lax distinction in German". FIPKM, 35, 17-25.

Ladd, D.R., Mennen, I. & Schepman, A. (2000). "Phonological conditioning of peak alignment in rising pitch accents in Dutch". J. Acoust. Soc. Am., 107(I), 2685-2696.

Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. (1972). "An Auditory-Motor Theory of Speech Production". UCLA Working Papers in Phonetics, 22, 48-75.

Ladefoged, P. (1964). "A Phonetic Study of West African Languages". Cambridge University Press. Reprinted 1968.

Larson, C., Kempster, G.; Kistler, M. (1987). "Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles". JSHR, 30, 552-558.

Lisker, L. & Abramson, A. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements". Word, 20, 384-422.

Lowell, S. & Story, B. (2006). "Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration". J. Acoust. Soc. Am., 120(1), 386-397.

Löfqvist, A., Baer, T., McGarr, N. S. & Story, R. S. (1989). "The cricothyroid muscle in voicing control". J. Acoust. Soc. Am., 85, 1314-1321.

Löfqvist, A., McGarr, N. S. & Honda, K. (1984). "Laryngeal muscles and articulatory control". J. Acoust. Soc. Am., 85, 951-954.

MacNeilage, P. F., Sussman, H. M., Westbury, J. R., and Powers, R. K. (1979). "Mechanical properties of single motor units in speech musculature". J. Acoust. Soc. Am., 65, 1047-1052.

MacNeilage, P. F., Sussman, H. M., and Powers, R. K. (1977). "Discharge patterns accompanying sustained activation of motor units in speech musculature". J. Phonetics, 5, 135-147.

MacNeilage, P. F. (1973). "Preliminaries to the study of single motor unit activity in speech musculature". J. Phonetics, 1, 55-71.

Matisoff, J. (1973). "Tonogenesis in Southeast Asia". In L. Hyman (ed.), Consonant types and tone. Southern California Occasional Papers in Linguistics, pp. 71-96.

McClean, M. D. & Clay, J. L. (1995). "Activation of Lip Motor Units With Variations in Speech Rate and Phonetic Structure". J. speech and hearing research, 38, 772 - 783.

McClean, M. D. & Cooker, H. S. (1980). "Preliminary observations on the response of perioral single motor units to mechanical stretch". J. Acoust. Soc. Am., 67, 271- 275.

McClean, M. D. (1984). "Recruitment thresholds of lower-lip motor units with changes in movement direction". J. Speech and Hearing Research, 27, 6-12.

Meyer, E.A. (1896/7). "Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort". Phonetische Studien (Beiblatt zu der Zeitschrift Die neueren Sprachen), 10, 1-21.

Meyer, E.A. (1910). "Untersuchungen über Lautbildung". Die neueren Sprachen 18 (Ergänzungsband Festschrift Vietor), 166-248.

Mooshammer, C. (1998). "Experimentalphonetische Untersuchungen zur artikulatorischen Modellierungder Gespanntheitsopposition im Deutschen". Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation, München (FIPKM) 36, 3-191.

Ní Chasaide, A. & Gobl, C. (1993). "Contextual variation of the vowel voice source as a function of adjacent consonants". Language and Speech, 36 (2-3), 303-330.

Ohala, J. (1978). "Production of tone". In V. Fromkin (ed.), Tone: A linguistic survey, pp. 5-39.

Ohala, J.; Eukel, B. (1987). "Explaining the intrinsic pitch of vowels". In R. Channon and L. Shockey (eds.), In honor of Ilse Lehiste, pp. 207-215. Dordrecht, Foris.

Ohde, R. N. (1984). "Fundamental frequency as an acoustic correlate of stop consonant voicing". J. Acoust. Soc. Am., 75, 224- 230.

Pape, D. & Mooshammer, C. (2004). "Intrinsic pitch in German - Examining the whole fundamental frequency contour of the vowel". Proceedings CFA/DAGA 2004, 897-898.

Pape, D. & Mooshammer, C. (2006). "Is intrinsic pitch language-dependent? Evidences from a cross-linguistic vowel pitch experiment". Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Multilingual Speech and Language Processing, Stellenbosch, South Africa.

Pape, D., Mooshammer, C., Fuchs, S. & Hoole, P. (2005). "Intrinsic pitch differences between the German vowels /i:/, /I/ and /y:/ in a cross-linguistic perception experiment". Proceedings of the ISCA workshop on plasticity in speech perception (PSP2005), London, UK.

Reinholt Petersen, N. (1983). "The effect of consonant type on fundamental frequency and larynx height in Danish". ARIPUC, 17, 55-86.

Rischel, J. & Hutters, B. (1980). "Filtering of EMG signals". ARIPUC (Copenhagen) 14, 285-316.

Rosenbaum, D. A. (1991). "Human Motor Control". Academic Press: USA.

Rossi, M. & Autesserre, D. (1981). "Movements of the Hyoid and the Larynx and the Intrinsic Frequency of Vowels". J. Phonetics, 9, 233-249.

Sapir, S., McClean, M. D. & Luschei, E. S. (1984). "Time relations between cricothyroid muscle activity and the voice fundamental frequency (F0) during sinusoidal modulations of F0". J. Acoust. Soc. Am., 75, 1639-1641.

Sapir, S. (l989). "The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations". J. Voice, 3, 44-51.

Sawashima, M., Hirose, H., Yoshioka, H. & Kiritani, S. (1982). "Interaction between articulatory movements and vocal pitch control in Japanese word accent". Phonetica, 39(4-5), 188-198.

Schiefer, L. (1986). "F0 in the production and perception of breathy stops: Evidence from Hindi". Phonetica 43, 43- 69.

Schuh, R. (1978). "Tone Rules". In V. Fromkin (ed.), Tone: A linguistic survey, pp. 221-256.

Sievers, E. (1901). "Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen". Leipzig: Breitkopf und Härtel.

Silverman, K. (1986). "F0 segmemtal cues depend on intonation: The case of the rise after voiced stops". Phonetica, 43, 76-91.

Stevens, K. N. (1977). "Physics of laryngeal behavior and larynx modes". Phonetica, 34, 264-279.

Stevens, K. N. (1989). "On the quantal nature of speech". J. Phonetics 17. Theme issue: On the quantal nature of speech, 3- 45.

Strik, H., Boves, L. (1991). "A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech". J. Phonetics, 19, 367-378.

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S. & Wilhelms-Tricarico, R. (1997). "Tongue surface displacement during bilabial stops". J. Acoust. Soc. Am., 102(1), 562-571.

Titze, I. (1994). "Principles of Voice Production". Prentice Hall: Englewood Cliffs.

Traunmüller, H. (1981). "Perceptual dimension of openness in vowels". J. Acoust. Soc. Am., 69, 1465-1475.

Trubetzkoy, N. (1938). "Die phonologischen Grundzüge der sogenannten 'Quantität' in den verschiedenen Sprachen". Scritti in onore di Alfredo Trombetti, 155-174. Milano: Ulrico Hoepli Editore.

Trubetzkoy, N. (1939). "Grundzüge der Phonologie". Göttingen: Vandenhoeck & Ruprecht (7th edition 1989).

Vennemann, T. (1991). "Syllable structure and syllable cut prosodies in Modern Standard German". In: P. Bertinetto et al. (eds.): Certamen Phonologicum II: Papers from the 1990 Cortona Phonology Meeting, 211-243. Torino: Rosenberg & Sellier.

Vennemann, T. (2000). "From quantity to syllable cuts: On so-called lengthening in the Germanic languages". Rivista di Linguisitca, 12(1), 251-282.

Vilkman, E., Aaltonen, O., Laine, U. & Raimo, I. (1991). "Intrinsic pitch of vowels - a complicated problem with an obvious solution?". In Vocal fold physiology: acoustic, perceptual and physiological aspects of voice mechanisms (J. Gauffin & B. Hammarberg, editors), pp. 159-166. San Diego, CA: Singular Publishing Group.

Vilkman, E., Aaltonen, O., Raimo, I., Arajärvi, P., & Oksanen, H. (1989). "Articulatory hyoid-laryngeal changes vs. cricothyroid muscle activity in the control of intrinsic F0 of vowels". J. Phonetics, 17, 193-203.

Westbury, J. R. (1983). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing". J. Acoust. Soc. Am., 73, 1322-1336.

Whalen, D. H., Abramson, A. S., Lisker, L. & Mody, M. (1990). "Gradient effects of fundamental frequency on stop consonant voicing judgments". Phonetica, 47(1-2), 36-49.

Whalen, D. H., Abramson, A. S., Lisker, L. & Mody, M. (1993). "F0 gives voicing information even with unambiguous voice onset times". J. Acoust. Soc. Am., 93, 2152 - 2160.

Whalen, D. H., Gick, B., Kumada, M. & Honda, K. (1999). "Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0". J. Phonetics, 27(2), 125-142.

Whalen, D.H. & Levitt, A. G. (1995). "The universality of intrinsic F0 of vowels". J. Phonetics, 23(3), 349-366.

Wood, S. (1982). "X-ray and model studies of vowel articulation". Working Papers 23, Lund University.

Zawadzki, P. A. & Gilbert, H. R. (1989). "Vowel fundamental frequency and articulatory position". J. Phonetics, 17, 159-166.

Zhou, P.; Rymer, W.Z. (2004). "Can standard surface EMG processing parameters be used to estimate motor unit global firing rate?" J. Neural Engineering, 1, 99-110.