THE EFFECTS OF STRESS AND F₀ ON THE VOICE SOURCE

Jacques Koreman

Abstract

In this paper, we present data on the qualitative difference in the voice source characteristics in stressed versus unstressed syllables; for the unstressed syllables, we will further point out the qualitative differences dependent on high versus low F₀ (fundamental frequency of the vocal fold vibrations). The voice source characteristics will be described in terms of a set of parameters extracted from the inverse filtered signal, consisting of the parameters OQ (open quotient), rₖ (pulse skewing), Eₑ (excitation strength), AC (peak-to-peak airflow) and DC offset (leak flow). We attempt to give a physiological explanation for our findings by comparing our data with those from other experiments, discussed in the literature. We also compare our findings with the predictions from modelling studies, and suggest some improvements of the phonation models to obtain a better correspondence between their predictions and our results. The physiological interpretation of our findings remains tentative, since no direct physiological measurements were made. Nevertheless we are able to propose a consistent explanation for the behaviour of the glottal parameter values that were observed.

1. Introduction

Phoneticians often describe the speech signal in terms of the source-filter model (Fant, 1960). In this model (see figure 1), a division is made between the vocal tract excitation (source) and its (anti-)resonances (filter), which enhance certain frequency components present in the source signal, while reducing others.

Especially the filter has been investigated extensively. Much less is known about the source signal, which, for voiced sounds, is produced at the glottis. One obvious reason for this gap in our knowledge is that the airflow at the vocal folds is hard to measure due to their inaccessibility. In speech synthesis this has led to a situation in which many of the properties of the voice source have been included in the filter characteristics, while a simple source is used which consists of a (number of) pulse(s), which may be filtered before they serve as the excitation of the filter which represents the
vocal tract. As a result, it is obviously impossible to provide a flexible control of the voice source for synthesizing different voice qualities, for instance for different speakers or depending on the suprasegmental properties of the utterance. Quite apart from that, it does not help us to understand the true physiological behaviour of the voice source. The need for a more flexible and physiologically interpretable voice source forms the background to this study.

This paper discusses voice source characteristics which have been obtained from the inverse filtered oral airflow signal in a larger project, titled "The influence of consonantal environment and F0 movement on the voice source characteristics", which was carried out at the university of Nijmegen, the Netherlands, and funded by the Dutch Organisation for Scientific Research (NWO).

1.1. The voice source signal

We have chosen a set of parameters derived from the true glottal airflow (cf. Anathapadmanabha & Fant, 1982) as our representation of the voice source. This signal was obtained by inverse filtering (I.F.) of the airflow measured at the lips (Koreman & Cranen, 1989). The true glottal airflow is a time-dynamic representation of the voice source which can replace static voice sources such as the ones described above. Using the true glottal airflow to represent the source guarantees linear separability of source and filter (i.e. the airflow at the lips can be computed from the true glottal airflow by multiplying that signal with the vocal tract properties), so that the true glottal airflow is especially useful in speech synthesis. A disadvantage of this signal is that, since our filter estimate is based on the (relatively stable) vocal tract characteristics during the closed glottis interval of the glottal period, time-variant interactions which occur when the glottis is open (caused by coupling of the sub- and supraglottal cavities) are not compensated for and occur as ripple in the true glottal airflow; the amount of ripple depends on the segment that is being articulated. This need not be a problem for a more physiological interpretation of the signal, which we shall aim at, as long as we take only the most robust properties of the glottal pulse shape into account when interpreting the signal.
1.2. Modelling the voice source

The first attempt to a better (physiologically oriented) understanding of the voice source was made in the article "On the air resistance and the Bernoulli effect of the human larynx" by Van den Berg et al. (1957), which marks the beginning of an era of research into voice source modelling. It describes the voice source in terms of the relationship between transglottal pressure, the dimensions of the glottis, and volume velocity of the glottal airflow in voiced sounds on the basis of experiments with a plaster cast model of an excised human larynx. Since then, computer models have been developed which incorporate many properties of normal vocal fold behaviour. Two models have been and still are particularly influential: the two-mass model (Ishizaka & Flanagan, 1972) and the body-cover model (Titze & Talkin, 1979). Both of them consist of a self-oscillating source: in the two-mass model each vocal fold is composed of two stiffness-coupled masses, whereas in the body-cover model each fold is represented by a layered structure (mucosal, ligamental, and muscular layer), in which properties of the layers affect the wave propagations through the tissue, which determines vocal fold vibration. Of course, other models of the voice source do exist (e.g. Fant et al., 1985; Klatt & Klatt, 1990), but these only model the acoustic behaviour of the voice source. This makes them particularly suitable for use in speech synthesizers, enabling us to control the behaviour of the voice source independently of the vocal tract filter characteristics. At the present time, the usefulness of any of the existing flexible voice source models for speech synthesis purposes (both physiologically oriented and acoustic ones) is limited, because only relatively little is known about the behaviour of the voice source in natural speech utterances. This study is intended as a step toward filling this gap by investigating the voice source characteristics in different linguistic conditions: we will report on qualitative distinctions in the glottal airflow parameters under the influence of stress and F₀.

1.3. Physiological interpretation of the glottal airflow

It is of course one thing to find out how the glottal airflow characteristics differ for several linguistic conditions (in itself theoretically sufficient to improve the naturalness of synthesized speech), and another to be able to explain these differences in terms of the underlying physiology. We believe that voice source research, also for synthesis purposes, can only be successful in the long run, if a possible physiological basis for a description of the voice source characteristics is provided. Models like the two-mass model and the body-cover model, both of which are based on physiological data, can help us to come to a better understanding of the relationship between characteristics of the glottal airflow on the one hand, and the responsible physiological mechanisms on the other. Unfortunately, there are many uncertainties in the physiological basis of the voice source models. The models mainly restrict themselves to a mathematical description of the relationship between glottal airflow, transglottal pressure, and glottal area. The glottal area in the two-mass model is controlled by spring constants, which are functionally related to muscle activity. Titze & Talkin (1979) state that "[their] ultimate aim is to propose a model for muscular control of the larynx which predicts acoustic consequences of contraction of a small set of intrinsic laryngeal muscles", while restricting themselves to the "systematic variation of a larger set of geometric and elastic parameters". It is clearly very difficult to provide a description of the role of muscle activity in parameter changes, especially in a quantitative model. Using converging evidence from the literature reporting on muscle activity and transglottal pressure in different experimental conditions, we shall attempt a tentative qualitative explanation of the observed voice source characteristics in our experiment in terms of the role of muscle activity and transglottal pressure. This is the aim of the work reported in this paper.
2. Stress, F0 and voice source characteristics

We shall attempt to give a physiological explanation of the behaviour of the voice source in stressed versus unstressed syllables, and in unstressed syllables with a high and low F0, respectively.

The voice source characteristics consist of a set of glottal airflow parameters: open quotient (OQ), pulse skewing (r_k), excitation strength (E_e), peak-to-peak airflow (AC) and DC offset of the airflow (DC). Together, these parameters determine the main *shape* characteristics of the glottal pulses and thus the spectrum of the voice source signal (see figure 2). Overall patterns which have been shown to be different were obtained through Hidden Markov Modelling (Koreman et al., 1991; Koreman et al., 1992). We shall take the results from this procedure as a starting point, and try to give a physiological interpretation to them.

\[
\begin{align*}
OQ &= \frac{T_{\text{open}}}{T_0} \\
E_e &= \text{amplitude } U_g' \text{ at } T_e \\
AC &= \text{amplitude } U_g \text{ at } T_{\text{max}} - \text{amplitude } U_g \text{ at } T_{\text{min}} \\
r_k &= \frac{T_-}{T_+} \\
DC &= \text{value } U_g \text{ at } T_{\text{min}}
\end{align*}
\]

![Figure 2. Glottal airflow parameters and their derivation from the glottal airflow (U_g) and its time derivative (U_g').](image-url)

2.1. An experiment for four intonation patterns

An experiment was carried out in which the voice source characteristics (glottal airflow parameters) were compared for the vowels of stimuli of the form /pEpECEpE/, where C was replaced by all Dutch consonants. The stimulus word was stressed on the second (2) or on the third (3) syllable, the stress being realised either by a falling (F) or a rising (R) F0 movement. We thus have four stimulus classes: F2, F3, R2 and R3. All stimulus words were placed in a carrier phrase and realised with the "flat hat" intonation pattern intended by the experimenter. The goal of the experiment was to find characteristic *glottal*
pulse shape patterns for the four different intonation classes. The pulse shape can be described on the basis of the five glottal airflow parameters given above.

On the basis of our experimental results and corroborating evidence from the literature, we shall try to come to a physiological interpretation of our data. For the description of stress, we shall compare our findings to the effects of pressed voice. For the description of the effect of F_0 class, we shall take the two-mass and the body-cover models as a basis, since both have explicitly incorporated the effects of this parameter on some aspects of the glottal pulse shape. We must, however, point out that, due to limitations in these models, our interpretation must be seen as very tentative. Nevertheless, we shall attempt to provide a coherent picture of the possible physiological mechanisms underlying the parameter changes we observe in our experiment.

2.2. Voice source characteristics and stress

The effects of stress on the voice source parameters were found by comparing the glottal airflow parameters for the vowels in the stressed with those in the unstressed syllables of the stimulus words, regardless of the direction of the F_0 movement. Our findings are summarized in table 1.

<table>
<thead>
<tr>
<th>Stress</th>
<th>OQ</th>
<th>r_k</th>
<th>E_e</th>
<th>AC</th>
<th>DC offset</th>
</tr>
</thead>
</table>

The glottal airflow pulses are a little less sinusoidal in stressed than in unstressed syllables, with smaller open quotients and less pulse skewing. Their DC offset is also smaller, indicating less leak flow in the closed glottis interval. Excitation strength and AC flow are greater in stressed than in unstressed syllables due to the stronger closure of the vocal folds.

2.2.1. A physiological interpretation of the effects of stress

The main articles which we shall use to attempt to provide a physiological basis for our observations are Gobl (1988) and Pierrehumbert (1989). Before we start to compare our findings to those reported in the literature, we want to point out that it is often difficult to separate the effects of stress on the voice source parameters from those of F_0 (movement): stress is often realised with a pitch accent, and therefore there is an F_0 movement on the stressed syllable. This makes the two effects (i.e. of stress and F_0 movement) inseparable. In reports in the literature, it is not always clear with what type of pitch accent the stress is realised. The relationship between stress and the parameter values may be partly due to changes in the articulatory settings necessary to produce the F_0 movement used to realise the pitch accent. Of course, the stress in our stimulus words is also realised with a pitch accent, but because we can compare falling and rising F_0 movements, we can compensate for this to find the effects of stress alone on the pulse shape parameters.

Fant (1980) notes that "the source correlates of lexical stress patterns are not very distinct" (p. 29). He does, however, find changes in some of his parameters for the stressed syllables, among others an increase in F_0. Since the words used in his study were spoken in isolation, the stressed syllables were probably realized with a rising-
falling pitch accent; this is at least one possible reason for the increase in $F_0$. Another reason for the higher $F_0$ values may be that there is an increase in subglottal pressure in the stressed syllables, caused by increased activity of the respiratory muscles (Ladefoged, 1967; Van Katwijk, 1974). An increase in subglottal pressure has been shown to correlate with an increase in $F_0$ (Ishizaka & Flanagan, 1972; Strik, 1994).

Of course, the effects on the pulse skewing and the excitation strength which Fant finds may depend on or be enhanced by the higher $F_0$ (cf. section 2.3), which was not controlled for in Fant's study; the same is true for Gobl's findings (see below).

Fant notes that his asymmetry factor $K$ is higher in stressed syllables. Since a higher value for $K$ implies a more strongly skewed pulse form, his findings corroborate ours: we found (a weak tendency towards) lower $r_k$ values in stressed syllables, and thus a more skewed glottal pulse shape.

Fant notes that the greater skewing correlates with an increase in the excitation strength, which we also find. The same relationship between $r_k$ and $E_e$ was also found by Gobl (1988, p. 152). He writes that the tendency for a stronger excitation in stressed syllables is not absolute in his data (the $E_e$ level is sometimes greater in the unstressed vowel immediately preceding the stressed one). Gobl proposes two different explanations for the higher excitation strength in stressed vowels: first, increased respiratory effort and consequently higher subglottal pressure; and second, increased medial compression of the vocal folds, causing a higher $E_e$ due to a more abrupt closing phase of the glottal cycle.

Since the latter mechanism should cause a decrease of OQ (and $r_a$ (dynamic leaking), a measure not used in this study), which Gobl does not find in his data, the second hypothesis is discarded. Our data do not completely support Gobl's conclusions, since we do find a weak tendency for OQ to be lower in stressed vowels. We do not therefore exclude medial compression of the vocal folds as a possible physiological explanation for the observed voice source characteristics in stressed syllables.

Unfortunately, the other parameters used in our study were not systematically investigated in either Fant's or Gobl's study. Fant merely notes that the vocal pulse amplitude is not very different in stressed versus unstressed /e/ in his two stimulus words, which is not confirmed by our data.

Summarising our own results (see table 1), we find that stress has the strongest effects on the three amplitude parameters in our study: $E_e$ and AC are high in the stressed vowel, while DC offset is low. The effects on the time parameters OQ and $r_k$ are much weaker, but both tend to decrease in the stressed vowel.

The parameter changes that we find in the stressed vowel (compared to the unstressed ones) are the same as those noted for raised voice. Pierrehumbert (1989) writes: "A raised voice is typically more pressed [italics not in original text, JK]: the vocal folds are more adducted, giving rise to a shorter open quotient and a sharper closing than for a normal voice. These adjustments increase the amplitude of the high frequencies relative to the fundamental." The sharper closure of the vocal folds would lead to lower $r_k$ and higher $E_e$ values (Pierrehumbert assumes that the relative shortening of the opening phase is not greater than that of the closing phase of the vocal folds, and that that the amplitude of the glottal airflow waveform does not decrease; these assumptions are confirmed by our data). Moreover, adduction of the vocal folds, probably enhanced by medial compression of the vocal folds in pressed voice (see Laver, 1980, p. 145), would lead to lower DC offset values.

Pierrehumbert notes that a raised voice is also characterized by heightened subglottal pressure (see also Laver, 1980, p. 146). This increase in subglottal pressure can either be passive or active, and will result in higher maximum airflow values. Strik (1994) writes that increased activity of the laryngeal muscles (cf. greater adduction and medial tension of the vocal folds for pressed voice) leads to an increase in the impedance
of the glottis. This results in higher subglottal pressures, which in turn raises AC. The increase in AC may also be caused by an active increase in subglottal pressure due to increased activity of the respiratory muscles in stressed syllables (see above).

According to the two-mass (Ishizaka & Flanagan, 1972) and the body-cover model (Titze & Talkin, 1979), the increase in subglottal pressure would also lead to the decrease in OQ observed in our data, as well as to a decrease in $r_k$ (not discussed in the two-mass model).

Although all the parameter changes observed in the stressed syllables in our data can be explained from muscular activity alone, it is very likely (given the evidence from the literature) that (some of) the parameter changes are enhanced by an active increase in subglottal pressure. Our data provide further support for the suggestion made by Ní Chasaide (1987) that "stress is potentially manifest by increased muscular activity at every level of production; the respiratory, the laryngeal, and frequently, at the level of supralaryngeal articulation."

### 2.3. Voice source characteristics and $F_0$

The relationship between the glottal airflow characteristics and $F_0$ were obtained from the vowels in the three unstressed syllables in the stimulus words, which were divided up into syllables with a high versus syllables with a low $F_0$. For the intonation conditions F2 and F3, the syllables before the stressed one (the second, respectively third syllable of the stimulus word) have a high $F_0$, the ones after the stressed syllable have a low $F_0$, and vice versa for the other two intonation conditions. The voice source parameters for the stressed syllables are not taken into consideration, because the effects of $F_0$ are obscured by the effects of stress and because $F_0$ changes during the stressed syllable, so that the voice source characteristics cannot be related to either a high or a low $F_0$. The relationships that we found for the glottal airflow parameters with $F_0$ are summarised in table 2.

<table>
<thead>
<tr>
<th>$F_0$</th>
<th>OQ</th>
<th>$r_k$</th>
<th>$E_e$</th>
<th>AC</th>
<th>DC offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

The vowels in syllables with a high $F_0$ have greater open quotients and more skewed pulses (smaller $r_k$ values) than in syllables with a low $F_0$. The excitation strength is greater, indicating a faster closing movement. AC (the peak-to-peak amplitude of the glottal pulse) is smaller and the leak flow (DC offset) is greater.

### 2.3.1. A physiological interpretation of the effects of $F_0$

Three studies are particularly relevant for the discussion of our experiment. The first is Pierrehumbert (1989), a study in which the relationship between pitch and intensity on the one hand and glottal airflow parameters on the other was systematically investigated. The other two studies, reported in Ishizaka & Flanagan (1972) and Titze (1992) are modelling studies which attempt to provide a physiologically plausible explanation for the relationship between $F_0$ and intensity, and glottal airflow parameters (see also Titze (1988) for a comparison of his body-cover model with the two-mass model, and Story & Titze (1994) for a hybrid model which combines the two-mass model with the body-
In her study, Pierrehumbert measured OQ, $r_k$, $E_e$, and $T_{a}$\(^1\) for H and L tones in sentences pronounced at five different voice levels. Pierrehumbert notes that "the F\(_0\) values for H's were higher than for L's, in all voice levels". We can therefore compare the parameter values found by Pierrehumbert for H tones to those in syllables with high F\(_0\) in our experiment, and the parameter values for L tones to those at low F\(_0\) in our experiment (it should be borne in mind, though, that Pierrehumbert's measurements were made in stressed syllables!).

Another remark about a difference between Pierrehumbert's and our experiment is in order. In her experiment, Pierrehumbert distinguished five voice levels (intensities). Since intensity is one of the correlates of stress (Fry, 1955; Fry, 1958; Lieberman, 1960), voice level is to some extent related to (linguistic) stress (our experiment). Nevertheless, the non-linguistic voice level variable is likely to cover a greater range of variation than our (linguistic) stress variable, and may therefore affect the voice source parameter values more strongly. We must therefore only compare the effect of Pierrehumbert's H and L tones on the voice source parameters with the effect of our F\(_0\)'s within the same voice level. For instance, although within each of the five voice levels H tones correspond to higher open quotients than L tones, low tones at voice level 1 still have higher open quotients than high tones at level 5 (Pierrehumbert's figure 7). In what follows, all comparisons to Pierrehumbert's data are made within each of the voice levels.

Following Pierrehumbert, we compare our findings to the predictions of the two-mass model (Ishizaka & Flanagan, 1972). It should be pointed out that although the model is able to predict many of the parameter changes that are found in observed natural data, its physiological basis, and therefore its explanatory power when we want to relate parameter changes to physiology, has some limitations. We therefore try to find further support for our interpretations in other studies. The second modelling study we shall refer to is Titze (1992). The limitations of the model used in that study were mentioned in Titze & Talkin (1979); the simplification of the dynamic stress-strain characteristics of the laryngeal tissues in that model (which are assumed to be linear) may affect its predictions concerning the relationship between physiological parameters on the one hand, and glottal airflow descriptors on the other. In general, the relationships between muscle activity, subglottal pressure, aerodynamic factors and F\(_0\) is far too complex to expect them to have all been incorporated in the models correctly. The meaningfulness of an attempt to find an explanation for our observations on the basis of existing models is therefore limited from the start, but may shed some light on the shortcomings of these models, and thus lead to improvements.

Nevertheless, both models are based on widely accepted assumptions and observations, and are able to produce changes in the glottal airflow parameters that have been found experimentally. They can therefore provide a useful basis for discussing the physiological origin of the data derived from the acoustic model, so as to arrive at a more comprehensive phonetic view of voice excitation. In order to do so, we attempt to build up a consistent picture of the physiological mechanisms which may be responsible for the observed changes in our glottal airflow parameters.

\(^1\)\(T_{a}\), (dynamic leakage index) was not used in our study. Pierrehumbert shows in Fig. 5 in her article that $ST_{-\{a\}}$ and $SE_{-\{e\}}$ are highly negatively correlated, so that $ST_{-\{a\}}$ is predictable from $SE_{-\{e\}}$. She further notes that "[a] more detailed breakdown of the data, not shown here, indicates that the difference between H and L made no further contribution to explaining the statistical variability in $St_{-\{a\}}$ values."
2.3.2. The role of transglottal pressure

Possible physiological mechanisms causing an increase in $F_0$ are muscle activity and subglottal pressure. Let us first see whether in the two voice source models alone a change in subglottal pressure can explain the parameter changes which we have found. In neither of the two models do we find an increase in subglottal pressure (for higher $F_0$'s) to cause an increase in OQ; on the contrary, subglottal pressure and OQ are negatively correlated. On the basis of subglottal pressure alone, the body-cover model predicts the decrease of $r_k$ observed in our data, but it also predicts an increase of AC, which is the opposite of what we found; an increase in AC with subglottal pressure is also predicted by the two-mass model. On the whole, using the theoretical framework of the models, subglottal pressure does not consistently predict the parameter changes which we have observed in our data.

It is of course well known that an increase in subglottal pressure has often been found to be positively correlated with an increase in $F_0$. This "knowledge" has been incorporated in the voice source models. If this is correct, the question naturally arises why we do not observe the parameter changes predicted by the models in our data. Either the predictions of the effects of an increase of subglottal pressure on the glottal airflow parameters are incorrect, or the assumption that the subglottal pressure increases with $F_0$ must be wrong. Let us see whether we can find a possible explanation for our problem.

An active increase in subglottal pressure caused by "increased activity of the respiratory muscles for stressed syllables was found by Ladefoged (1967) and Van Katwijk (1974)" (Strik, 1994). This is not likely to occur in our case, since the data presented in section 2.3 are taken from unstressed syllables. However, a passive increase in subglottal pressure is possible: this "might be due to a change in the impedance of the glottis, which, in turn, results in changes in the activity of the laryngeal muscles" (Strik, 1994). Note, however, that Strik's measurements were also made in stressed syllables, so that we do not know whether even a passive relationship between $F_0$ and subglottal pressure is to be expected in our data. It is possible that no changes in subglottal pressure (relative to the pressure in stressed vowels) occur in unstressed syllables. However, a passive increase in subglottal pressure is possible: this "might be due to a change in the impedance of the glottis, which, in turn, results in changes in the activity of the laryngeal muscles" (Strik, 1994). Note, however, that Strik's measurements were also made in stressed syllables, so that we do not know whether even a passive relationship between $F_0$ and subglottal pressure is to be expected in our data. It is possible that no changes in subglottal pressure (relative to the pressure in stressed vowels) occur in unstressed syllables at all, so that the assumption that subglottal pressure always increases with $F_0$ may be an overgeneralization. If an increase in subglottal pressure does occur for higher $F_0$, it is clear that its effects must be weaker than the effects of muscular tension (which we shall discuss below). This conclusion is supported by Titze & Talkin (1979), who note that "[s]ubglottal pressure is a less effective fundamental frequency controlling mechanism [than muscular tension, and especially longitudinal stress, p. 73, JK]". For that reason, we shall now confine ourselves to a discussion of the way $F_0$-related muscle activity can affect the glottal parameters in the models.

2.3.3. The role of muscle activity

We shall now discuss the role of muscle activity for the behaviour of our glottal airflow parameters. It should be stressed that we are only trying to find model-internal explanations for the effects of muscle tension on the glottal parameters. Limitations of the models therefore continue to influence our interpretations.

OQ

Pierrehumbert reports higher open quotients for H tones than for L tones. This correspond well with the increase in OQ values which we found for higher $F_0$'s.
Pierrehumbert (1989) explains the higher observed open quotients for H tones from the greater tension of the vocal folds, caused by contraction of the cricothyroid muscle which is used to produce high tones. This relation between cricothyroid activity and $F_0$ has been noted by many other researchers (e.g. Collier, 1975; Atkinson, 1978, and references therein; Ishizaka & Flanagan, 1972; Strik & Boves, 1987), although other muscles (e.g. sternohyoid, lateral cricoarytenoid, sternothyroid, and vocalis) can be involved in the regulation of $F_0$ as well (see Atkinson, 1978).

The relationship between $F_0$ and OQ can be modelled in both the two-mass model (Ishizaka & Flanagan, 1972) and the body-cover model (Titze, 1992). In the two-mass model (see figure 3), the authors suggest that "this feature [namely the increase of open quotient with fundamental frequency, JK] can also be given to the vocal-cord model by modifying the coupling-tension parameter $k_c$ to increase more than in linear proportion to Q." For very large values of $k_c$, "close to the bounds of the oscillation range, both the glottal flow and area waveforms become sinusoidal on a dc component, and the glottis does not close." Since high values of $k_c$ reduce the phase difference between the lower and upper parts of the vocal folds $A_{g1}$ and $A_{g2}$ (making the model behave more like a one-mass model), and therefore correspond to more symmetrical waveshapes, this mechanism would not explain some of our other observations (as we shall see in the description of the relationship between $F_0$ and $r_k$ below); also, it seems unlikely that, even for the higher $F_0$'s, the vocal fold vibrations are close to the bounds of the oscillation range.

![Fig. 3. The two-mass model of the vocal folds](image)

There is, however, another possible parameter in the two-mass model which can predict the observed relationship between $F_0$ and OQ, and which at the same time leads to the observed changes in $r_k$. The parameter we are referring to is $k_2$ (the linear stiffness of spring $s_2$ in figure 3). As Ishizaka and Flanagan note, contraction of the cricothyroid muscle (which is one of the main muscles involved in $F_0$ control) leads to an increase in $k_2$, which can in turn lead to "no closure of $A_{g2}$ while $A_{g1}$ can close completely during the cycle. Owing to the small amplitude of $A_{g2}$ and its dc component, the glottal flow increases in upward roundness and also increases in duty ratio" (p. 1253).

The increase of OQ with $F_0$ was also modelled by Titze (1992). An increase in $F_0$ corresponds to an increase of his model parameter $c$ (the mucosal wave velocity in the vocal fold cover) and a decrease of $T$ (vocal fold thickness); at least the latter can be
expected to be related to tension of the cricothyroid muscle, the folds becoming thinner when the tension increases. Both of these changes lead to an increase in $P_{th}$, the phonation threshold pressure, according to the formula $P_{th} = k_1 Bc \xi_0 / T$. Since the open quotient in the airflow waveform $Q_o$ is related to $P_{th}$ through the empirical formula $Q_o = k_2 + (1 - k_2) (P_{th} / P_L)$, an increase in the threshold pressure in turn leads to an increase in $Q_o$ (assuming that the change in lung pressure $P_L$ is relatively small). Although Titze's predictions are borne out by our data, there are serious limitations to this interpretation, which we will come back to below when we describe the model prediction for the relationship between $F_0$ and $r_k$.

The effects of stress and $F_0$ on the voice source

$r_k$

The data reported by Pierrehumbert indicate the same relationship between pitch and $r_k$ as we found in our experiment: L tones have a higher $r_k$ value, and therefore a more symmetric pulse shape, than H tones.

Before turning to the possible effects of muscle tension on $r_k$, we must refer to Titze (1992), who says that "[a]t high fundamental frequencies, $Q_s$ [the inverse of $r_k$, JK] may decrease with $F_0$ because vocal tract inertance is less effective in skewing the pulse". The $F_0$'s in our data are not likely to come close to $F_1$, so that vocal tract inertance is not likely to have a significant effect on pulse skewing. Moreover, $r_k$ shows the opposite behaviour in our data, so that we have discarded acoustic loading as a possible cause of the parameter change with $F_0$, and have assumed that the skewing effects with $F_0$ are completely attributable to skewing of the glottal area, related to muscle tension.

Pierrehumbert claims that "this result [low values for $r_k$ at high $F_0$, and vice versa, JK] is surprising, since in both the Ishizaka & Flanagan (1972) and the Titze & Talkin (1979) models, the higher vocal fold tension responsible for $F_0$ raising results in a more symmetric pulse shape" (and therefore high $r_k$ values). Her conclusion that the two-mass model predicts high $r_k$ values for H tones is correct if we assume, as Ishizaka and Flanagan do themselves (!), that an increase in $F_0$ should be modelled by an increase in $k_c$, which in turn would lead to more symmetrical glottal waveforms (p. 1252-1253). The $k_c$ parameter models the coupling stiffness between the lower and upper vocal fold masses, which is related to an increase of the tension in the vocalis muscle (Ishizaka & Flanagan, 1972, p. 1237).

It was already hinted at in the description of $OQ$ above, however, that the cricothyroid muscle (which primarily affects the longitudinal tension of the vocal folds) is probably more important for $F_0$ control than the vocalis (cf. Farley, 1994, p. 1027). Titze & Talkin (1979) support this conclusion: "For a given laryngeal size, fundamental frequency is controllable primarily through longitudinal stress in the muscular layers, and in a less obvious fashion, through additional transverse stresses resulting from coupling between the longitudinal and transverse stresses." The assumption that $F_0$ is mainly controlled by the cricothyroid muscle is also supported by quantitative experimental data presented in Atkinson (1978). Pierrehumbert is therefore probably correct in assuming (as we shall) that "[i]n the mid to high region of the pitch range, the dominant mechanism for $F_0$ control appears to be the cricothyroid contraction and relaxation", rather than the vocalis (thyroarytenoid) muscle. But if this assumption is correct, contrary to Pierrehumbert's argument based on changes in $k_c$, the predictions of the two-mass model can describe both Pierrehumbert's and our data very well (as was also the case for the relationship between $F_0$ and $OQ$), since relaxation of the cricothyroid muscle (for L tones) is related to smaller $k_2$ values in the two-mass model (p. 1237 of Ishizaka & Flanagan, 1972), which has the result that "the glottal waves tend to a symmetrical form" (p. 1253). This prediction corresponds to the high $r_k$ values which Pierrehumbert and we
found for low pitch in comparison to the \( r_k \) values for high pitch. Since there are so many indications, both from our own data, from the literature and from Ishizaka & Flanagan's article, that \( k_2 \) should play an important role in the regulation of \( F_0 \), we shall base our other model predictions for the effects of \( F_0 \) on the voice source parameters on the model parameter \( k_2 \) instead of \( k_c \).

Now let us see what predictions the body-cover model makes with regard to the relationship between \( F_0 \) and \( r_k \). We have already seen that in Titze (1992) the phonation threshold pressure \( P_{th} \) increases with \( F_0 \). Through a second empirical formula in the article, \( Q_s = 1.0 + k_3 (2P_m - P_L - P_{th}) (P_L - P_{th}) \) Titze describes a positive relationship between \( P_{th} \) and \( r_k \) (of which \( Q_s \) is the inverse). Assuming the other variables remain constant, as we did for \( Q_0 \), \( r_k \) should therefore increase with \( F_0 \), which is the opposite of the relationship between the two parameters observed in our data. The empirical formula is based on a subset of the data reported in an article by Sundberg et al. (1993), and concerns the airflow characteristics of five tenors. Their task was not to read out sentences, but to produce sequences of identical syllables at five different pitches, spanning 1.5 to 2 octaves. In the following section, we shall try to explain why the data obtained from such a task may not be comparable to our data.

**E_e and AC**

\( E_e \) is greater at higher \( F_0 \)'s in our experiment, while \( AC \) is lower. For normal and (two) lower voice levels, the same relationship between \( F_0 \) and \( E_e \) is also found by Pierrehumbert: H tones have somewhat greater excitation strengths than L tones (although, as was the case for \( OQ \), the excitation strength is mainly determined by voice level). At higher than normal voice levels, this relationship is absent (either inverse or unclear).

In the two-mass model, the behaviour of \( E_e \) and \( AC \) with changes in \( F_0 \) are a complex function of many model parameters. As a consequence, these parameters are difficult to predict from the behaviour of individual model parameters. Let us attempt to tentatively formulate how the parameter changes we find could be generated with the model. The two-mass model shows a decrease of \( AC \) with higher \( F_0 \) values (due to the decrease in the amplitude of \( A_{g2} \) when \( k_2 \) increases). This relationship between \( AC \) and \( F_0 \) was also found in our data, and would (without a change in the closing duration) lead to lower values of \( E_e \) for high \( F_0 \)'s. On the other hand, we have already mentioned that higher \( k_2 \) values (for higher \( F_0 \)'s) typically correspond to smaller \( r_k \) values, i.e. more skewed pulse shapes with a relatively short closing duration; this tendency for short closing durations is strengthened simply by the shorter period duration corresponding to a high \( F_0 \). The short closing duration can lead to higher \( E_e \) values for high \( F_0 \) values, despite a decrease in \( AC \).

It is unfortunately not possible to make any clear predictions on the basis of Titze (1992) about the relationship between \( AC \) and \( E_e \) on the one hand, and \( F_0 \) on the other.

**DC offset**

Finally, let us see how well the models are able to predict the increase we found in the DC offset with \( F_0 \). As was the case for \( AC \) and \( E_e \), it is not possible to make any statements about DC offset on the basis of the body-cover model; this parameter is left out of consideration in the model altogether.
The two-mass model predicts that, when glottal conditions are kept at their "typical values", an increase of $k_2$ (which is related to an increase in $F_0$) "will lead to no closure of $A_{g2}$ while $A_{g1}$ can close completely during the cycle." The model therefore does not predict the increase in the value of the DC offset with increasing $F_0$ which we found. It may be possible to improve the predictions of the model by increasing $k_1$ (longitudinal stiffness of the lower masses) as well as $k_2$ when longitudinal tension on the vocal folds increases. This would allow the lower masses in the model to also remain apart during the complete glottal cycle, creating the possibility of greater DC offset values. Another way to model this is by assuming that $k_2$ and $k_c$ are not completely independent, which is in fact very likely, so that $k_c$ also increases when $k_2$ does. Then the increase of DC offset can be predicted by the model.

**Summary**

In summary, we cannot hope to offer a complete physiological explanation for the correlation of our glottal flow parameters with $F_0$ on the basis of existing models. We therefore only try to come to a qualitative, global description of the relationship between voice source parameters at different $F_0$'s on the one hand, and the underlying physiological mechanisms on the other. As was explained above, differences in subglottal pressure are probably only of minor importance for the behaviour of the glottal parameters in the unstressed syllables in our investigation, since they cannot explain our observations; the same is true for acoustic loading.

We therefore want to point out how some of our voice source parameters may be related to tension of the vocal folds, which is likely to be the main physiological correlate of high $F_0$ in unstressed syllables. The greater (longitudinal) tension of the vocal folds for high $F_0$'s causes a greater resistance of the folds to displacement from their neutral position, thus decreasing AC. The DC offset increases with increasing tension, because the inward displacement of the vocal folds (i.e. the negative modulation of the rest area) also becomes smaller. For the negative correlation between $F_0$ and $r_k$, it is very difficult to come up with a clear physiological explanation (although, as we have explained above, it can be explained in the two-mass model on the basis of the parameter $k_2$). In fact, on the basis of the suggestion in Gobl (1988) that "the opening time in normal phonation may be more or less conditioned by the time constant of the mass and compliance of the vocal folds, whereas the closing time is affected by the Bernoulli forces" we would expect exactly the opposite correlation: the decrease in mass and compliance of the vocal folds for higher $F_0$'s should lead to shorter opening durations, and the reduction of the effects of the Bernoulli force due to the greater resistance to displacement should lead to longer closing durations. The combination of these effects should cause the pulse shape to become more symmetrical, i.e. it should cause an increase of $r_k$. This is exactly the opposite of the relationship we observed. The increase in $E_e$ that we find can be explained by a combination of the decrease of $r_k$ and the shorter duration of the glottal period ($T_0$); the effects of the decrease of $T_0$ and $r_k$ clearly outweigh those of the decrease of AC, since otherwise the changes in $E_e$ would have been in the opposite direction. As Ishizaka and Flanagan point out, smaller displacement of the vocal folds at high $F_0$'s lead to higher values of $OQ$, since the relative duration during which the vocal folds are in contact decreases. With the exception of the behaviour of $r_k$, we have formulated a very tentative physiological hypothesis for the observed effects of $F_0$ on the voice source parameters. Much further research, both concerning the relationship between measurements in the acoustic, aerodynamic and physiological domain and concerning voice source modelling, is obviously necessary before we can hope to acquire a complete and consistent picture of
the physiological mechanisms which are used to produce different voice source characteristics.

### 2.3.4. The possible division of labour between cricothyroid and vocalis muscles

Although not immediately relevant for the interpretation of our data, we want to attempt to resolve some of the clashes between our data on the one hand and reports in the literature and model predictions on the other. We shall point out how the tasks in F0 control may roughly be divided over the cricothyroid and vocalis muscles (which caused some difficulties in the interpretation above), in order to provide (a small step towards) a clearer picture of voice source control. It is obvious that, of the glottal parameters whose behaviour was investigated in the two-mass and the body-cover models, rk presents the most problems. In both models, rk is positively correlated with F0, reflecting the findings from many studies; in our experiment, we found that rk decreases with an increase of F0, the waveform thus becoming more skewed. This contradictory behaviour of rk in different studies may find a possible explanation in its dependence on the subjects' task, and/or on frequency range.

Pierrehumbert notes that "[her] own informal observations suggest that the extremely high F0 values found at the end of yes/no questions exhibit more symmetric pulse shapes than the H tones in this study, whose associated F0 values were far lower." A positive correlation between F0 and rk was also found by Holmberg et al. (1989) when they compared normal and high pitch. In the two-mass model, such findings could be modelled by an increase in kc, which is related to the activity of the vocalis muscles, for pitch increase at the more extreme end of the pitch range (but see Farley, 1994, p. 1027). The hypothesis that the cricothyroid muscle is responsible for linguistic pitch changes within the normal pitch range, while the vocalis muscle is mainly used to control non-linguistic pitch changes or, especially, pitch changes in the higher frequency range, although very tentative, is at least partly supported by Atkinson (1978), who writes that "there is a direct causal relationship between CT [cricothyroid, JK] and F0" (p. 214), while "several studies have shown a general correlation between V [vocalis, JK] activity and F0. The most consistent and convincing studies have involved singing or isolated words" (same page). It therefore seems that it is not implausible to assume that the task carried out by our (and Pierrehumbert's) subjects, is crucially different from the one that the two-mass and body-cover model are based on, and that it is precisely this difference which may explain the contrasting findings of rk in different studies.

### 3. Conclusion

This paper describes a physiological interpretation of the behaviour of the voice source for four different intonation classes. The voice source characteristics consist of a set of parameters computed from the glottal airflow, which together describe the shape of the glottal pulses. It consists of the parameters OQ (open quotient), rk (pulse skewing), Ee (excitation strength), AC (peak-to-peak airflow) and DC offset (leak flow). We looked at

1Holmberg et al. (1989) did not measure rk, but say that the speed quotient (which is the inverse of rk) shows a significant difference (p = 0.027) between normal and high pitch, the speed quotient being higher for lower pitch. The differences between their pitch conditions are non-linguistic, however.

2From his own data, which concern short statements and questions, there is no direct evidence for the presupposed increase in vocalis activity in the higher frequency range. This may be explained by the stimulus material he used: higher frequencies are only to be expected in his question sentences, where he does find a tendency towards a higher correlation between vocalis activity and F0, though it is not significant, possibly because the correlations were made over the whole utterance.
the effects of stress and F₀ on the parameter patterns found in each of the intonation classes, and presented related findings from the literature.

Given the complexity of the physiological mechanisms regulating stress and F₀, it was difficult to come up with a complete physiological explanation of the findings. We have nevertheless attempted to present such an explanation on the basis of subglottal pressure and vocal fold tension, supported by findings from studies involving physiological measurements and from modelling studies. There are many uncertainties in the interpretation of the data in terms of the underlying physiology, so that much further work, using physiological measurements to improve existing voice source models or to propose new ones, will be necessary to come to a better understanding of so complex a physiological system as the control of the voice source.

4. References


