In the present paper we examine the simultaneous downtrend in fundamental frequency and subglottal pressure that is often observed for running speech. In particular, we will test the hypothesis that the downtrend in fundamental frequency is caused by a gradual decrease in subglottal pressure during the course of an utterance. In the literature, various ways to model the downtrend in fundamental frequency have been proposed. Our conclusion is that whether the hypothesis stated above is true depends on the model of downtrend adopted.

1. Introduction

A simultaneous downtrend in fundamental frequency ($F_0$) and subglottal pressure ($P_{sb}$) has often been observed for running speech (Lieberman, 1967; Ohala, 1970; Collier, 1974, 1975; Atkinson, 1978; Gelfer, 1987; Strik and Boves, 1993). As it is known that changes in $P_{sb}$ will affect $F_0$, everything else being equal (Titze, 1989), it seems plausible to assume that both downtrends are related. However, a considerable deal of controversy surrounds the relation between the two downtrends (see e.g., Ohala, 1978, 1990; Cohen, Collier and 't Hart, 1982; Ladd, 1984).

Research on the relation between the downtrend in $F_0$ and $P_{sb}$ is impeded by the fact that there is still no consensus on the correct way to model the downtrend in $F_0$. In the literature, various models have been proposed. Many of these models consist of two components: a short-term or local component and a long-term or global component. In these models, the global component is used to model the downtrend in $F_0$. Only some of these models provide a physiological explanation of both components. Öhman (1968), Collier (1975), and Fujisaki (1991) agree that the local component is controlled by the laryngeal muscles, but they do not agree about the control of the global component. According to Öhman (1968) and Fujisaki (1991), downtrend is also controlled by the laryngeal muscles, while according to Collier (1975) it is controlled by $P_{sb}$.

In Strik and Boves (1993), the relation between $F_0$ and some of the physiological mechanisms that are known to be important in the control of $F_0$ are studied by means of a qualitative analysis. Based on our own data and data from the literature, it was concluded that from a psychological viewpoint the following hypothesis is plausible: the downtrend in $F_0$ is due to the downtrend in $P_{sb}$. However, this
hypothesis is not unchallenged. In this article we will discuss the two main counter-arguments:

1. The lowering in $P_{sb}$ cannot explain all of the decrease in $F_0$ (Section 4.2.); and
2. Downtrend is part of the linguistic code, and thus it must be controlled by laryngeal muscles and not by $P_{sb}$ (Section 4.3.).

The fact that this issue is still controversial is expressed in the conclusion of a recent article by Ohala (1990): "It must be concluded that the question of whether $F_0$ declination is caused by laryngeal or by respiratory activity has still not been answered definitively." The purpose of this article is to clarify the relation between the downtrend in $F_0$ and $P_{sb}$.

In the literature, different models of intonation are available which are motivated both by phonetic and phonological considerations. The primary goal of the present article is to study the relation between the downtrend in $F_0$ and $P_{sb}$. For this reason, we look primarily at intonation from a physiological point of view. As a consequence, we try to avoid theory-laden terms like, e.g., "downdrift", "declination", and “baseline” as much as possible. Instead, we predominantly use the more neutral term “downtrend”. In some sections we refer to previous studies in which the term “declination” is generally used. In these cases we will also use the term “declination”. In this article, “downtrend” and “declination” are seen as synonyms, and are used to denote the gradual lowering of a signal during a whole utterance.

The outline of the article is as follows. In Section 2., material and method are described. Each experiment consisted of two parts. In part one the subjects were instructed to sustain vowels, and in part two they produced meaningful sentences. The results for “sustained phonation” are described in Section 3. These results are then used in the argumentation of Section 4., in which the results for “running speech” are presented. In Section 4.1., our physiological model of intonation is described. Subsequently, the two counter-arguments mentioned above are discussed in Section 4.2. and 4.3., respectively. Section 5. contains a general discussion. Finally, some conclusions are drawn in Section 6.

2. Materials and methods

Recordings were made of the audio signal, electroglottogram, lung volume ($V_t$), $P_{sb}$, and the activity of the sternohyoid (SH) and vocalis (VOC) muscles for two Dutch male subjects. Both subjects had normal phonation and hearing, but had not received special voice training. In addition to these signals, the activity of the cricothyroid (CT) muscle was also measured for subject LB (the second author), and oral pressure for subject HB. The electromyographic (EMG) signals of the laryngeal muscles were high-pass filtered, full-wave-rectified, and integrated over successive periods of 5 ms. All EMG signals were shifted forward over their mean response times, using the procedure described in Atkinson (1978).

The measurements were made while the subjects produced sustained vowels and meaningful Dutch sentences with different intonation patterns. The sentences spoken by subject LB were “Piet slikte zijn pillen met bier” (SU: Short Utterance); and “Piet slikte gisteren zijn vierentwintig gele pillen liever in stilte met bier” (LU: Long Utterance). The sentences produced by subject HB were “Heleen wil die kleren meenemen” (SU: Short Utterance); “Heleen en Emiel willen die kleren
Downtrend in $F_0$ and $P_{sh}$

liever wel weer meenemen” (LU: Long Utterance); and “Indien Emiel die kleren
wil meenemen, willen wij ze eerst wel even zien” (SWC: Sentence With Comma). These sentences contain mainly high vowels, in order to minimize the involvement of the SH in articulatory gestures.

The intonation contours produced were one “pointed hat” (HB-SU1, early stress); two “pointed hats” (HB-SU2, LB-SU2 and LB-LU2, early and late stress, $F_0$ is lowered in between); a “flat hat” (HB-SU3, LB-SU1 and LB-LU1, early and late stress, $F_0$ is kept high in between); and question intonation (HB-SU4, HB-LU4, LB-SU3 and LB-LU3). The intonation pattern of HB-SWC is more complex. For an explanation of the notions “pointed hat” and “flat hat” the reader is referred to ’t Hart, Collier and Cohen (1990).

Some sentences were also produced in réitérant form, using either the syllable /fi/ or /vi/. The subjects repeated each sentence 5 to 8 times. The raw signals of these repetitions were used to calculate median signals for each intonation contour. The method of non-linear time-alignment and averaging was used to average all signals, including $F_0$ (Strik and Boves, 1991). The procedures used for recording and processing the data are described in more detail in Strik and Boves (1992).

3. Sustained vowels

Before the actual measurements of the physiological signals were made, our subjects were trained to produce prolonged vowels for different combinations of $F_0$ and intensity level (IL). When the subjects were asked to sustain a given vowel, a gradual lowering of $F_0$ and IL was generally observed. Subsequently, when they were explicitly instructed to keep $F_0$ and IL constant, the downtrend in $F_0$ and IL diminished, but it was usually still present. Finally, the subjects were given on-line visual feedback of $F_0$ and IL. In this condition, they often managed to keep both $F_0$ and IL fairly constant during the production of a vowel.

After the training sessions actual measurements of the physiological signals were obtained. The subjects were given on-line visual feedback and were again instructed to keep $F_0$ and IL constant for a sustained vowel. This task was repeated for different combinations of $F_0$ and IL. The measurements show that the subjects usually managed to keep $F_0$ and IL at the target values. At the beginning of the utterances, some variation in $P_{sh}$ and the activity of the laryngeal muscles was observed, probably to reach the target levels for $F_0$ and IL. Apart from the initial variation, the physiological signals usually remained constant for the rest of the utterance. Different combinations of $F_0$ and IL were achieved by different levels of $P_{sh}$, SH, CT, and VOC. The results of this part of the experiment are described in more detail in Strik and Boves (1987).

This experiment shows that subjects who had no special voice training can keep $F_0$, IL, and $P_{sh}$ constant during a simple utterance (a sustained vowel), but only if they are supported by visual feedback. Subjects report that keeping $F_0$ and IL constant requires more effort than allowing a gradual decline, and feels less natural. Without visual feedback, $F_0$ and IL (and probably also $P_{sh}$) tend to fall gradually during the course of an utterance, even if subjects are instructed to keep $F_0$ and IL constant. The results obtained for sustained phonation will be used as support for the argumentation in the next section on running speech.
4. Running speech

4.1. A physiological model of intonation

In Strik and Boves (1993), we proposed a qualitative model of $F_0$ control in running speech. Our model describes consistent behaviour of $P_{sb}$, CT, VOC, and SH that was observed in the data of subjects LB and HB, and in other data presented in the literature. Figures with the average signals for the recorded utterances of subjects LB and HB can be found in Strik and Boves (1993). Here we will only display the average signals of a typical utterance (see Fig. 1), in order to illustrate our model.

The four physiological signals mentioned above were chosen because it is known that they are important in the control of $F_0$. In our model intonation and its physiological control take place at two levels, viz. a global and a local level. This is in accordance with other physiological models of intonation proposed in the literature (like Öhman, 1968; Collier, 1975; and Fujisaki, 1991).

Short-term variations in $F_0$, $P_{sb}$, SH, VOC, and CT have often been observed (see, e.g., Fig. 1), i.e., all five signals clearly have a local component. But it is not immediately clear whether all of these five physiological signals also have a global component.

4.1.1. Global level

A gradual lowering of $P_{sb}$ and $F_0$ during the course of a major syntactic constituent is often observed (see, e.g., Lieberman, 1967; Ohala, 1970; Collier, 1974, 1975; Atkinson, 1978; Gelfer, 1987; Strik and Boves, 1993). The domain in which the downtrends in $F_0$ and $P_{sb}$ occur has previously been given many different names,

\[\text{Figure 1. Average physiological signals for the Dutch utterance. "Piet slikte gisteren zijn vierentwintig gele pillen liever in stilte met bier" (LU1) spoken by subject LB. Also shown in the first and second panel are the global trend lines $F_{0,g}$ and $P_{sb,g}$, respectively (dashed-dotted lines).}\]
among other things “breath group” (Lieberman, 1967), “intonation group” (Breckenridge, 1977), “utterance” (Pierrehumbert and Beckman, 1988), “clause or clause complexes” (Clark and Yallop, 1990), or “major phrase” (Honda and Fujimura, 1991). In this article, we will use the term “utterance”. Within the recorded sentences there were no inspirations (resets of \(V_1\)), nor any resets of \(F_0\) or \(P_{sh}\).

Our definition of a global component is a gradual change spanning the total duration of an utterance. Therefore, in our model \(P_{sh}\) and \(F_0\) have a global component. The global component of \(F_0\) and \(P_{sh}\) in our model will be called \(F_{0,g}\) and \(P_{sh,g}\) respectively. In this article, the terms \(F_{0,g}\) and \(P_{sh,g}\) will be used for the global components of our model alone. Global components of other models will be denoted otherwise.

The model presented in Strik and Boves (1993) is a qualitative model. To illustrate our model, a possible quantitative decomposition of \(F_0\) and \(P_{sh}\) in a global and a local component is shown in Fig. 1. \(P_{sh,g}\) was obtained by manually fitting an exponential function through most of the valleys of \(P_{sh}\) (Fig. 1). Because it is assumed that \(F_0\) varies linearly with \(P_{sh}\) (Titze, 1989), \(F_{0,g}\) was defined in the following way: \(F_{0,g} = B_0 + B_1 \cdot P_{sh,g}\). The values of \(B_0\) and \(B_1\) that gave a satisfactory result for this utterance were 70 Hz and 5 Hz/cm H2O (Fig. 1), respectively. We would like to note that the manually fitted trend lines are only presented here to illustrate our qualitative model, and to give an example of a procedure that can be used to obtain the global and local components of \(P_{sh}\) and \(F_0\). These manually fitted trend lines are not used for further analysis in the present article. Instead, we will use a more objective statistical method in the following section.

A gradual change in the activity of \(SH\), \(VOC\), or \(CT\) during a whole utterance was not observed in any of our recordings nor in published data of other researchers (as far as we know). Sometimes the activity of these three laryngeal muscles varied slowly during part of the utterances, but no instance of a slow increase or decrease during the whole utterance (just like \(P_{sh}\) and \(F_0\)) was found. It must therefore be concluded, both from our own data and the data presented in various other papers, that in general \(SH\), \(VOC\), and \(CT\) do not seem to have a global component.

4.1.2. Local level

At the beginning of utterances \(CT\), \(VOC\), and \(P_{sh}\) may have extra high values, and the result will be a so-called “initial rise” of \(F_0\) (Fig. 1). At the end of utterances \(SH\) activity often increases while \(P_{sh}\) drops sharply. If these effects occur during voiced sounds at the end of the utterance, final lowering of \(F_0\) is observed (Fig. 1). Alternatively, increased \(SH\) activity and \(P_{sh}\) release may be delayed until after the last voiced sound, in which cases final lowering is absent (e.g., in most interrogative utterances). The initial rise and final lowering of \(F_0\) will add to the \(F_0\) fall that results from the downtrend in \(F_{0,g}\) alone (Fig. 1).

The local component of \(P_{sh}\) (\(P_{sh,l} = P_{sh} - P_{sh,g}\)) is generally positive. \(SH\), \(VOC\), and \(CT\) only have a local component, which is always positive because these signals can never become negative (see Section 2.). Finally, the local component of \(F_0\) (\(F_{0,l} = F_0 - F_{0,g}\)) is positive when the effect of \(F_0\)-raising mechanisms (\(VOC\), \(CT\), and \(P_{sh,l}\)) is larger than the effect of the \(F_0\)-lowering mechanisms (\(SH\)), and \(F_{0,l}\) becomes negative when the net effect of \(F_0\)-raising and \(F_0\)-lowering mechanisms is negative.
4.1.3. Hypothesis

To conclude this section, in our physiological model of intonation, SH, VOC, and CT do not have a global component, while $F_0$ and $P_{sb}$ do have a global component. A two-component model was chosen, because from a physiological point of view this seems to be the model that best describes the data. Because a downtrend in $F_{0,g}$ and $P_{sb,g}$ is often observed, the following hypothesis seems likely: The downtrend in $F_{0,g}$ is due to the downtrend in $P_{sb,g}$. This hypothesis has been challenged for different reasons. Two frequently adduced counter-arguments are discussed in the next two sections.

4.2. The $F_0$-$P_{sb}$ ratio

4.2.1. Counter-argument 1

An argument used against the above-mentioned hypothesis is that the variation in $P_{sb,g}$ cannot explain the total variation in $F_{0,g}$, because the $F_0$-$P_{sb}$ ratio (FPR) observed in running speech is often larger than 7 Hz/cm H$_2$O (e.g., Maeda, 1976; Ohala, 1978). Studies of the rate of $F_0$ change resulting from a change in $P_{sb}$ alone (generally by externally induced pressure variations) have revealed that the FPR should be in the range 2–7 Hz/cm H$_2$O (e.g., Ladefoged, 1967; Baer, 1979). In the present article, this range will be called the FPR-range. Because the FPR obtained for utterances often seems to exceed the FPR-range, the hypothesis is either rejected totally (Ohala, 1978), or an additional mechanism is invoked to explain (part of) the decrease in $F_0$ (the tracheal pull mechanism of Maeda, 1976).

Indeed, there seem to be no reasons to assume that the FPR obtained in experiments with externally induced pressure variations differs from the FPR in running speech. But the problem is that the FPR obtained for running speech depends on the way in which the downtrend in $F_0$ and $P_{sb}$ is defined and modelled.

4.2.2 Modelling the relation between $F_0$ and $P_{sb}$

In the literature, several methods have been proposed to model the downtrend in $F_0$, such as the difference between $F_0$ at the beginning and at the end of an utterance (see method 1 below), the baseline of Maeda (1976), and the bottomline and topline of Cooper & Sorenson (1981). Baseline, bottomline, and topline are trend lines which are generally fitted manually, just like $P_{sb,g}$ and $F_{0,g}$ in Fig. 1. Most probably the fitting is done manually because it is difficult to define a mathematical error function that could be used to derive the trend lines with an optimization algorithm.

We have done a number of experiments to determine the parameters of the downtrend components. The results of two experiments, in which different definitions of downtrend were used, are presented below. For this aim, six utterances of subject LB and six utterances of subject HB were used. For each subject, these are four declarative and two interrogative utterances (see Table I). All signals, including the $F_0$ signals, are average signals (Section 2.). Figures with the average signals for these twelve utterances can be found in Strik & Boves (1993). The average signals for one utterance of subject LB are shown in Fig. 1.

Method 1. In this method, the $F_0$ and $P_{sb}$ values are taken at two instances, one near the beginning ($T_1$) and one near the end ($T_2$). The following values are then
Downtrend in $F_0$ and $P_{sh}$

Table I. Listed from top to bottom are: utterance type, number of voiced samples (N), length of the utterance ($T = T_2 - T_1$) in s, $F_0$ values of first ($F_0(T_1)$) and last ($F_0(T_2)$) voiced sample in Hz, total fall of $F_0$ ($dF_0 = F_0(T_1) - F_0(T_2)$) in Hz, average rate of change of $F_0$ ($dF_0/T$) in Hz/s, $P_{sh}$ values for first ($P_{sh}(T_1)$) and last ($P_{sh}(T_2)$) voiced sample in cm H$_2$O, total fall of $P_{sh}$ ($dP_{sh} = P_{sh}(T_1) - P_{sh}(T_2)$) in cm H$_2$O, average rate of change of $P_{sh}$ ($dP_{sh}/T$) in cm H$_2$O/s, FPR$_1 = dF_0/dP_{sh}$ in Hz/cm H$_2$O, and the regression coefficient between $F_0$ and $P_{sh}$ (FPR$_2$) in a multiple regression equation, also in Hz/cm H$_2$O (for explanations, see also the text).

<table>
<thead>
<tr>
<th>utt</th>
<th>Declarative utterances</th>
<th>Questions</th>
<th>Declarative utterances</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SU1</td>
<td>SU2</td>
<td>LU1</td>
<td>LU2</td>
</tr>
<tr>
<td>N</td>
<td>234</td>
<td>226</td>
<td>558</td>
<td>524</td>
</tr>
<tr>
<td>T</td>
<td>1.42</td>
<td>1.41</td>
<td>3.46</td>
<td>3.40</td>
</tr>
<tr>
<td>$F_0(T_1)$</td>
<td>150</td>
<td>136</td>
<td>147</td>
<td>136</td>
</tr>
<tr>
<td>$F_0(T_2)$</td>
<td>65</td>
<td>67</td>
<td>66</td>
<td>79</td>
</tr>
<tr>
<td>$dF_0$</td>
<td>85</td>
<td>69</td>
<td>81</td>
<td>57</td>
</tr>
<tr>
<td>$dF_0/T$</td>
<td>60.1</td>
<td>49.1</td>
<td>23.4</td>
<td>16.7</td>
</tr>
<tr>
<td>$P_{sh}(T_1)$</td>
<td>9.58</td>
<td>9.92</td>
<td>11.64</td>
<td>11.82</td>
</tr>
<tr>
<td>$P_{sh}(T_2)$</td>
<td>3.44</td>
<td>3.50</td>
<td>4.82</td>
<td>4.57</td>
</tr>
<tr>
<td>$dP_{sh}$</td>
<td>6.14</td>
<td>6.42</td>
<td>6.82</td>
<td>7.25</td>
</tr>
<tr>
<td>$dP_{sh}/T$</td>
<td>6.8</td>
<td>4.57</td>
<td>1.98</td>
<td>2.13</td>
</tr>
<tr>
<td>FPR$_1$</td>
<td>13.9</td>
<td>10.8</td>
<td>11.9</td>
<td>7.87</td>
</tr>
<tr>
<td>FPR$_2$</td>
<td>3.97</td>
<td>7.63</td>
<td>2.30</td>
<td>4.58</td>
</tr>
</tbody>
</table>

calculated: $dF_0 = F_0(T_1) - F_0(T_2)$, $dP_{sh} = P_{sh}(T_1) - P_{sh}(T_2)$, FPR$_1 = dF_0/dP_{sh}$. The total fall in $F_0$ and $P_{sh}$ from $T_1$ up to $T_2$ ($dF_0$ and $dP_{sh}$, respectively) is used to model the downtrend in $F_0$ and $P_{sh}$, respectively. Basing $dF_0$ on two $F_0$ values is error prone. In some studies the $F_0$ values are obtained from a trend line (e.g., the baseline in Maeda, 1976), while in other studies the $F_0$ values are taken from a single, representative $F_0$ contour (e.g., Collier, 1975; Gelfer, Harris, Collier and Baer, 1983; Collier, 1987). Our data processing procedure allowed us to average the $F_0$ curves of all repetitions of a given sentence, therewith making the estimation procedure more reliable. In previous studies various choices of $T_1$ and $T_2$ have been made, based on different motives (see, e.g., Gelfer et al., 1983). In this study, $T_1$ is the first voiced frame, and $T_2$ the last voiced frame of each utterance. These instants of $T_1$ and $T_2$ were mainly chosen because the values of $F_0$ and $P_{sh}$ at these time-points can be determined very easily for each utterance. Given this choice of $T_1$ and $T_2$, all relevant values were calculated for the twelve utterances of subjects LB and HB (see Table I).

In all utterances, $dP_{sh}$ is positive (Table I). For subject LB $dP_{sh}$ is always larger than for subject HB. For both subjects, $dP_{sh}$ for the interrogative utterances is smaller than $dP_{sh}$ for the declarative utterances. At the end of each question there is a marked increase in $F_0$, and consequently $dF_0$ is negative for the questions. But for all declarative utterances $dF_0$ is positive. For the declarative utterances, $dF_0$ of subject LB is always larger than $dF_0$ of subject HB. Partly this is because $dP_{sh}$ is larger for subject LB, as noted above. In addition, for subject LB the CT and VOC often show increased activity at the beginning of an utterance, which causes an initial rise in $F_0$, and the SH is increased at the end of the utterance during the final
lowering of $F_0$. Both effects will cause $dF_0$ to be larger than the fall in $F_0$ resulting from $dP_{sb}$ alone, i.e., both $P_{sb}$ and the laryngeal muscles participate in $dF_0$.

The values of $FPR_1$ can be seen in Table I. Only three of the twelve $FPR_1$ values are within the accepted $FPR$-range. $FPR_1$ for the four questions is negative because $dF_0$ is negative, four of the eight values of $FPR_1$ for the statements are larger than 7 cm H$_2$O and one is smaller than 2 cm H$_2$O. Based on these $FPR_1$ values one could conclude that the downtrend in $P_{sb}$ cannot explain all the downtrend in $F_0$, and thus other factors should contribute to the downtrend in $F_0$. If downtrend is defined in this way, then this conclusion is correct. After all, $dF_0$ does depend on both $dP_{sb}$ and the activity of the laryngeal muscles (especially for subject LB, as explained above).

The $FPR$-range is obtained from experiments with externally induced pressure variations (e.g., Ladefoged, 1967; Baer, 1979). The goal of these experiments was to determine the $FPR$ for $F_0$ changes that result from $P_{sb}$ changes along, i.e., one tried to keep other processes that influence $F_0$ (like the laryngeal muscles) constant (see e.g., Baer, 1979). In these studies, the points in a scatterplot for $F_0$ as a function of $P_{sb}$ could usually be fitted reasonably by a straight line. In Fig. 2, an $F_0$-$P_{sb}$ scatterplot is given for a short utterance of subject LB. Clearly, in this scatterplot the points are not grouped around a straight line. The reason is that during this utterance the other factors which influence $F_0$ are not constant. Drawn in Fig. 2 is the straight line that connects the first and the last voiced frame. $FPR_1$ is the slope of this line. In Fig. 2, one can see that the $FPR$ obtained in this way depends heavily on the exact choice of $T_1$ and $T_2$. To sum up, method 1 has two important drawbacks:

1. Other factors that can affect $F_0$ are not constant over the course of an utterance; and

![Figure 2. $F_0$ as a function of $P_{sb}$ for the Dutch utterance “Piet slikte zijn pillen met bier” (SU1) spoken by subject LB. The straight line is the line connecting the first and the last voiced frame. $FPR_1$ is the slope of this line.](image-url)
2. Because the other factors are not constant it is hazardous to make estimates of the FPR which are based on the values of \( F_0 \) and \( P_{sb} \) at two instants only.

**Method 2.** In method 2, a multiple regression analysis is used, in which \( F_0 \) is the criterion and \( P_{sb} \), VOC, and SH are the predictors. The outcomes of the regression analysis are the coefficients \( A_i \) of the regression equation: \( F_0 = A_0 + A_1 P_{sb} + A_2 \text{VOC} + A_3 \text{SH} \). The FPR is the regression coefficient between \( F_0 \) and \( P_{sb} \): \( \text{FPR}_2 = A_1 \). This method does not have the drawbacks of method 1 because a correction is made for some important other factors which influence \( F_0 \), and the regression coefficient is based on the data of all voiced frames.

The multiple regression analysis decomposes \( F_0 \) into four components: \( A_0 \), \( A_1 P_{sb} \), \( A_2 \text{VOC} \), and \( A_3 \text{SH} \). The first component is the constant \( A_0 \). VOC and SH do not have a global component either (Section 4.1), and thus in this statistical model, the downtrend in \( F_0 \) is due to the downtrend in \( P_{sb} \) alone. This is in line with the physiological model presented in Section 4.1., except for one essential difference. In method 2, \( P_{sb} \) is not decomposed into a global and a local component. However, because these are no reasons to assume that the FPR is different on a global and a local level, this does not seem to be a problem. Consequently, the \( P_{sb} \) component in the regression analysis (\( A_1 P_{sb} \)) contains both the slow downtrend in \( F_0 \) and the part of the local variations in \( F_0 \) which is due to the local variations in \( P_{sb} \). The other part of the local variations \( F_0 \) is in the VOC and SH component (\( A_2 \text{VOC} \) and \( A_3 \text{SH} \)), respectively.

Instead of using the multiple regression analysis we could have based our estimates of the FPR on the global trend lines \( P_{sb,g} \) and \( F_{0,g} \). To that end, \( P_{sb,g} \) and \( F_{0,g} \) should have been determined in the way described in Section 4.1., i.e., by making manual fits for all utterances. This is certainly possible, but we prefer to use objective, statistical methods (like the multiple regression analysis described in this section), instead of more subjective methods in which trend lines are fitted manually.

For all voiced frames of the twelve utterances a multiple regression analysis was performed in which \( F_0 \) was the criterion and \( P_{sb} \), VOC and SH were the predictors. The resulting \( \text{FPR}_2 \) values (i.e., the \( A_1 \) values) can be seen in Table I. The resulting values of \( A_0 \), \( A_2 \), and \( A_3 \) were not used for further analysis. Of the 12 \( \text{FPR}_2 \) values, 11 are in the FPR-range, and one is slightly larger than the maximum of the FPR-range. If the CT had been used as a predictor instead of the VOC for subject LB, then \( \text{FPR}_2 \) would have been 6.44 Hz/cm H\(_2\)O for this utterance, and thus it would have been within the FPR-range. Also, for the interrogative utterances \( \text{FPR}_2 \) is always within the FPR-range, while this was never the case for \( \text{FPR}_1 \). The rise of \( F_0 \) at the end of questions is usually due to an increase of CT, VOC, and \( P_{sb} \). In method 2, a correction is made for the increase in VOC, and the result is that the \( \text{FPR}_2 \) is within the FPR-range. The rapid increase in \( P_{sb} \) at the end of the questions is part of \( P_{sb} \), and will also explain part of the end rise in \( F_0 \).

To conclude this section, comparison of \( \text{FPR}_1 \) and \( \text{FPR}_2 \) values for sentences has shown that the actual values obtained are crucially dependent on the way in which the \( F_0-P_{sb} \) ratio is defined. In our opinion \( \text{FPR}_1 \), which has been used to refute the

---

1 For subject LB the correlation between CT and \( F_0 \) is generally larger than the correlation between VOC and \( F_0 \), and thus CT is a better predictor of \( F_0 \). But because the behaviour of CT and VOC is almost identical for subject LB, and because the activity of the CT was not measured for subject HB, we have chosen the VOC as a predictor in the regression analysis for both subjects.
above-mentioned hypothesis, is not a fair measure because it isolates $P_{sb}$, but at the same time ignores all other factors affecting $F_0$. If some important additional influences are factored out of $F_0$ by means of a multiple regression analysis, as is done with $\text{FPR}_2$, a completely different picture emerges, which is compatible with the hypothesis that the downtrend in $P_{sb}$ explains the downtrend in $F_0$. Even though the way in which the influence of the laryngeal muscles on $F_0$ is modelled is extremely crude (the true relation between the activity of the laryngeal muscles and $F_0$ is very likely to be non-linear) $\text{FPR}_2$ is a much fairer measure than $\text{FPR}_p$. According to this measure, the variation in $P_{sb}$ can explain all the variation in $F_0$, and no additional mechanisms are necessary. Therefore, too large a total $F_0$ drop does not seem a reason to reject the hypothesis. Also, and perhaps even more important, arguments about the relation between $F_0$ and $P_{sb}$ depend fully on the way in which the two downtrends are modelled. As long as the model of $F_0$ downtrend does not partition out effects not related to $P_{sb}$, it may remain a valid definition of its own, but it should no longer be used in arguments involving $P_{sb}$.

4.3 Control of $F_0$ and $P_{sb}$

4.3.1 Counter-argument 2

At the basis of the second counter-argument is the idea that the laryngeal muscles can be controlled linguistically, while this is not possible for the respiratory muscles and thus the downtrend in $P_{sb}$ is a passive process. Subsequently, this idea is used as an argument against the above-stated hypothesis: because the downtrend in $F_0$ is (at least partially) linguistically controlled it cannot result from an automatic process like the downtrend in $P_{sb}$. The fact that some authors use this argument in the discussion about the physiological causes of declination was also noted by Cohen, Collier & ’t Hart (1982).

The second argument against the hypothesis is expressed most clearly by Breckenridge (1977). She states that declination is part of the linguistic system, and therefore it must be controlled by the laryngeal muscles just as other linguistically significant aspects of $F_0$ are. A similar line of reasoning is used by Ohala (1978, 1990). In Ohala (1978, 1990), three possible causes for declination are mentioned: (1) tracheal pull (Maeda, 1976); (2) downtrend in $P_{sb}$ (Collier, 1974, 1975); and (3) graded activity in the laryngeal muscles. According to Ohala the first two causes are automatic, non-purposive physiological causes. Because declination is not automatic but controlled, he argues that a model in which linguistic aspects of $F_0$ are completely determined by actions of the laryngeal muscles is much more likely than a two-component model in which respiratory and laryngeal factors interact.

Clear opinions about the control of the downtrend in $P_{sb}$ can also be found in Gelfer et al. (1983), Ladd (1984), and ’t Hart, Collier and Cohen (1990). Gelfer et al. (1983) studied whether declination is actively controlled. They noted a similar downtrend in $F_0$ and $P_{sb}$. They argue that if the declination in $F_0$ is due to the declination in $P_{sb}$, then this would suggest that declination is a passive phenomenon. In Ladd (1984) three physiological causes of declination are discussed: (1) the downtrend in $P_{sb}$ (Collier, 1975); (2) the tracheal pull (Maeda, 1976); and (3) $F_0$ rises are harder to produce than $F_0$ falls (Ohala and Ewan, 1973). According to Ladd, the downtrend in $P_{sb}$ and the tracheal pull are automatic mechanisms. Finally,
according to 't Hart et al. (1990), the muscular activity involved in the regulation of $V_f$ and $P_{sb}$ is subject to an automatic control system. In their view, declination should be seen mainly as an automatic by-product of respiration.

The examples given above clearly illustrate that there seems to be a widespread notion that the downtrend in $P_{sb}$ is an automatic process. If the downtrend in $P_{sb}$ is a completely passive process, then this could indeed be used as a counter-argument against the above-mentioned hypothesis, because there are many indications that declination is under linguistic control, at least so some extent. However, it is not certain that the downtrend in $P_{sb}$ is a passive mechanism. On the contrary, there are many reasons to believe that $P_{sb}$ is controlled. This will be discussed in the next section.

4.3.2 Respiratory system

There are three factors which may affect $P_{sb}$ (see, e.g., Ladefoged, 1967):

1. Passive forces, like elastic recoil and gravitational forces;
2. Active forces, resulting from contractions of respiratory muscles; and
3. The resistance to the air-stream, both at the glottis and in the vocal tract ($Z_g$).

The pressure that results from passive forces alone is generally called the relaxation pressure ($P_{rel}$), while the pressure change brought about by active muscle contractions is called the muscular pressure. For a speaker who remains in the same position (usually upright), the gravitational forces are roughly constant and thus $P_{rel}$ would depend on $V_f$ alone. If expiration during speech production were a truly passive process, then the muscular pressure should be zero and $P_{sb}$ should be a function of $V_f$ and $Z_g$ alone. Several observations reveal that this is not the case:

- Our data show that for repetitions of the same sentence the amount of inspiration before the utterance was not always the same. Consequently, the $V_f$ traces run essentially parallel (see e.g., Fig. 3), while $Z_g$ can be assumed to be reasonably constant. Although the differences in $V_f$ are large, the $P_{sb}$ contours are very much alike (Fig. 3).
- Some of the sentences were also produced in réitérant form, using either syllable /fi/ or /vi/. The slopes of the $V_f$ traces of these two types of utterances are different, but also in this case the $P_{sb}$ contours showed much resemblance (see, e.g., Fig. 4). This was also found by Gelfer (1987).
- Speakers can keep their $P_{sb}$ constant during the production of a long sequence of /ma/ syllables (Collier, 1987), and during sustained phonation (Section 3.). In both cases, the activity of the measured laryngeal muscles also remained constant, so $Z_g$ was probably constant. The fact that speakers can keep $P_{sb}$ constant while $V_f$ is decreasing also proves that $P_{sb}$ is not simply a function of $V_f$ and $Z_g$ alone.
- During phonation $P_{sb}$ should not become smaller than a threshold value below which phonation is not possible (the so-called phonation threshold pressure, see Titze, 1992). Furthermore, the loudness of the speech is determined to a large extent by $P_{sb}$, and thus $P_{sb}$ should be kept within a certain range to produce speech with the desired loudness. After inspiration at the beginning of an utterance $P_{rel}$ is often larger than the desired $P_{sb}$, while at the end of an utterance $P_{rel}$ is often lower than the desired $P_{sb}$ (see, e.g., Ladefoged, 1967). If the
Figure 3. $F_0$, $P_{sb}$, and $V_1$ signals for two repetitions of a spontaneous sentence spoken by subject HB. The average difference for $V_1$ is 470 cc, and for $P_{sb}$ it is 0.05 cm H$_2$O.

respiratory muscles were not used, then $P_{sb}$ and the loudness would decrease rapidly; soon $P_{sb}$ would be smaller than the phonation threshold pressure and phonation would stop. To prevent this, the inspiratory muscles are used at the beginning of an utterance to keep $P_{sb}$ lower than $P_{rel}$, while expiratory muscles are used when $P_{rel}$ is lower than the desired $P_{sb}$ (Ladefoged, 1967).

Figure 4. Average $F_0$, $P_{sb}$, and $V_1$ signals for two utterances produced with réitérant speech: /vi/ (— —) and /fi/ (———).
The arguments given above force one to assume that the respiratory muscles are used to control $P_{sb}$ during speech production. The following question then arises: How are the respiratory muscles used to control $P_{sb}$? According to Ladefoged (1967) and Ohala (1990) the amount of control is limited, i.e., they claim that these muscles are only used to keep $P_{sb}$ reasonably constant above some minimal level. However, many measurements show that in general $P_{sb}$ is not constant but has a tendency to decline, both in sustained phonation (Section 3.) and in running speech (Lieberman, 1967; Ohala, 1970; Collier, 1974, 1975; Atkinson, 1978; Gelfer, 1987; Strik and Boves, 1993). Furthermore, $P_{sb}$ contours for repetitions of a sentence appear to be very similar in shape as well as in amplitude (see, e.g., Fig. 3), too similar to assume that $P_{sb}$ has just a convenient (more or less random) value above its minimum.

If the respiratory muscles are under voluntary control, then they can be used to control $P_{sb}$ during speech production. Active control of the respiratory muscles and $P_{sb}$ in speech production seems likely, given the following arguments:

- The way the respiratory muscles are used during speech production differs from the way they are used in normal breathing. In normal breathing the duration of inhalations and exhalations is about equal, while in speech production the inspiratory phase is much shorter. Furthermore, it has been observed that the posturing of the respiratory system for speech production (the prephonatory posturing of the chest wall) is different from the posturing for normal breathing (Hixon, Goldman and Mead, 1973; Baken, Cavallo and Weismann, 1979; Baken and Cavallo, 1981).

- Breathing pauses occur mainly at major constituent breaks (Winkworth et al., 1984). Breathing pauses can also occur at minor constituent boundaries, but as speaking rate increases they are eliminated from these minor breaks (Grosjean and Collins, 1978). Grosjean and Collins (1978) conclude that “it would appear that breathing in speech depends to a large extent on the speaker’s preplanned pause patterns”, and thus breathing would be linguistically controlled.

- The amount of air inspired and the $V_t$ at the beginning of sentences was found to be significantly larger for longer utterances compared to shorter ones, and for major syntactic breaks compared to more minor ones (Winkworth, Davis, Ellis and Adams, 1994). According to Winkworth et al. (1994), these findings indicate that speakers pre-plan their $V_t$ and the volume inspired. It should be noted that this study concerned reading, and therefore their results suggest that the respiratory muscles are under linguistic control during reading.

- Indications of extra respiratory activity (i.e., increased lung volume decrement) for stressed syllables were found by Ohala (1977), while Ladefoged (1967) and van Katwijk (1974) actually measured increased activity of respiratory muscles for stressed syllables. Although not all stressed syllables are probably accompanied by extra activity of the respiratory muscles, these results indicate that linguistic control of the respiratory muscles is possible, at least at a local level. If active control of the respiratory muscles is possible at a local level, then it is likely that it is also possible at a global level.

- Loudness is a prosodic, i.e., a linguistic variable. If speakers are asked to increase loudness, they tend to initiate speech at higher lung volumes (Hixon, et al. 1973). Winkworth et al. (1994) also found that louder utterances within the
"comfortable loudness" range are generally associated with higher lung volumes. According to Weismer (1985) it is more efficient to start at higher lung volumes for loud speech, because larger values of \( P_{sb} \) are needed to generate loud speech. So, not only is this an example of linguistic control of the respiratory muscles, it is also an indirect indication of linguistic control of \( P_{sb} \). But there are also more direct indications of voluntary control of \( P_{sb} \).

- In addition to \( P_{sb} \), a speaker can use many different physiological mechanisms to control \( F_0 \), and thus a given \( F_0 \) contour could be produced in various ways. Still, the amount of variation between physiological signals (including \( P_{sb} \)) of repetitions of the same utterance is relatively small (Strik and Boves, 1991, 1993). The finding that the inter-repetition variation in \( P_{sb} \) and the other physiological signals is small suggests that speakers have a notion of the manner in which they want to produce an utterance, and that they have a good control over \( P_{sb} \) and the other mechanisms.

- Another indication that \( P_{sb} \) is actively controlled can be seen in Fig. 5. In the middle of a spontaneous utterance, subject HB made a swallowing gesture, probably because the pressure catheter was bothering him. During this interruption \( P_{sb} \) suddenly drops to about 5 cm H\(_2\)O. For subject HB phonation with such a level of \( P_{sb} \) is possible, because comparable and even lower values of \( P_{sb} \) were found at the beginning of many voiced intervals of the repetitions of the same utterance. If the subject's only intention was to provide a \( P_{sb} \) above some minimal level at which phonation is possible, he could have kept \( P_{sb} \) at approximately 5 cm H\(_2\)O. However, before he resumed phonation, \( P_{sb} \) was raised to approximately the value it had before the interruption, and from that point it started declining again.

- Finally, after the two subjects in our study had received instructions they were able to keep \( P_{sb} \) fairly constant at different levels (Section 3.), i.e., their \( P_{sb} \) was under voluntary control.

![Figure 5. \( F_0 \), \( P_{sb} \), and \( V_1 \) signals for a spontaneous utterance spoken by subject HB. The arrow marks the interruption of about 0.5 s.](image-url)
The conclusion of this section is that there are several reasons to believe that the respiratory muscles and $P_{sb}$ are actively controlled. Furthermore, if this is the case, then the second counter-argument (specified above) cannot be used to refute the hypothesis that the lowering $F_{0,g}$ is generally due to a decrease in $P_{sb,g}$.

5. Discussion

In this paper, we have argued in favour of a major role for $P_{sb}$ in the control of the ubiquitous downtrend in $F_{0}$ contours. The role of $P_{sb}$ has been called into question by a number of authors, and for a number of different reasons. The two most important counter-arguments center around the claim that the total $F_{0}$ fall in most published data seems to exceed the range that should be expected from the fall in $P_{sb}$ and the claim that the respiratory system is not suited for so precise a control as needed for the linguistic, communicative purpose served by $F_{0}$ downtrend. These counter-arguments have been discussed in Sections 4.2. and 4.3., respectively.

Before proceeding to a summary of these discussions, we would like to address one additional argument. Ohala (1990) claims that there are examples in the literature that show a gradual downtrend of the activity of CT. It appears that these examples are limited to the contours 11 and 15 in Collier (1974). In these registrations, a gradual decline of CT activity can indeed be seen, but only in the second half of the utterances. To the best of our knowledge, there are no data showing a gradual variation of CT, VOC, or SH over complete utterances. But there are numerous examples of $P_{sb}$ decline that span a complete utterance. Thus, we fully acknowledge the possibility that laryngeal muscles contribute to the total fall of $F_{0}$ over the course of an utterance, but the available data more or less force us to accept the conclusion that the contribution of $P_{sb}$ to the control of $F_{0}$ downtrend (as the concept is defined in our model) is much more important. For this reason, we think that the physiological validity of the models proposed by Öhman (1968) and Fujisaki (1991), which do not acknowledge a role for $P_{sb}$, is debatable. Speakers can exploit a large array of physiological means to reach a certain goal, and it would be surprising if some of these means would never be exploited. After all, there is no valid reason to suppose that all subjects should always behave in exactly the same way. But individual examples attesting a possible way of control should not be generalized. For the time being, the data speak in favour of $P_{sb}$.

Coming back to the arguments related to the $F_{0}$-$P_{sb}$ ratio, it must be concluded that fair estimates of that ratio are extremely difficult to obtain from sentence material. In all naturally produced utterances laryngeal muscles affect $F_{0}$ in addition to $P_{sb}$. In order to obtain a fair estimate of FPR these additional contributions must be factored out. That is certainly not done by defining $dF_{0}$ and $dP_{sb}$ as the difference between the values observed at the beginning and at the end of an utterance, not even when these values are averaged over a large number of tokens, simply because the $F_{0}$ values are affected by laryngeal muscle activity.

A fundamental problem in studying the physiological causes of downtrend is that the literature abounds with definitions of $F_{0}$ downtrend. Downtrend, declination or downdrift have been used to denote the tendency of $F_{0}$ to decrease during the course of an utterance. This qualitative definition can be interpreted in many different ways, and is hardly suitable for studying the relation between physiology and $F_{0}$ downtrend. Therefore, a more precise definition of downtrend is needed.
Some of the definitions used in the literature are illustrated in Fig. 6. Fig. 6 shows hand-fitted estimates of a top line, a bottom line, a line connecting the first and last voiced sample in addition to $F_{0,g}$, which was derived in the way described in Section 4.1. (this is the same trend line as the one shown in Fig. 1). It can easily be seen that the slopes of these lines differ considerably. There is less literature on the definition of downtrend in $P_{sb}$. Yet, it is clear that the existence of several essentially different definitions or models of $F_{0}$ downtrend makes it impossible to discuss "the" relation between downtrend in $P_{sb}$ and $F_{0}$: the outcome of such a discussion is certain to depend on the exact definition of downtrend that is assumed.

According to our definition of a global component, $F_{0}$ and $P_{sb}$ do have a global component while CT, VOC, and SH generally do not have a global component. The quantitative statistical analysis has shown that, after correcting for the influence of VOC and SH, the variation in $P_{sb}$ can explain all the variation in $F_{0}$ (i.e., the FPR is usually within the correct range). Consequently, in our physiological two-component model the downtrend in $F_{0}$ can be explained completely by the downtrend in $P_{sb}$. However, it is always possible that other (unknown) factors also contribute to the downtrend in $F_{0}$. That is a possibility which cannot be ruled out.

This physiological two-component model was chosen because it seems to be the model which best describes the physiological data. If, for some reasons, someone prefers another definition of the global component, like for instance the top- or bottomline in Fig. 6, the conclusion should indeed be that the downtrend in $F_{0}$ cannot be determined entirely by the downtrend in $P_{sb}$, because top- and bottomline are determined to a large extent by the activity of the laryngeal muscles.

To sum up, in our model the downtrend in $F_{0}$ could be entirely due to the downtrend in $P_{sb}$. For other definitions of downtrend this does not have to be the case, i.e., these downtrend could be determined partially by the activity of the laryngeal muscles. However, the downtrend in $P_{sb}$ will always explain part of the downtrend in $F_{0}$.

Ideally, trend lines should not be determined by means of hand fitting, but instead by means of formal, mathematical procedures. However, each and every mathematical fit procedure requires the definition of an error (or cost) function, to quantify the discrepancy between the observed data and the model curve. For the time being, such an error function is almost impossible to define, because it is not
possible to reach agreement on the weight of details in the deviations. To a considerable extent, these weights depend on one's theoretical opinions about which details in $F_0$ curves are linguistically relevant and which are not. Another factor complicating the construction of a completely quantitative model of the control of $F_0$ in running speech is to do with the lack of knowledge about the relation between EMG activity of the laryngeal muscles and elastic properties of laryngeal tissue. In our own models we have assumed a simple linear relationship, but that is not more than a very crude first approximation. Thus, we have to be content with models that contain non-quantitative or non-realistic quantitative components for some time to come.

6. Conclusions

In this paper, we have investigated the relation between downtrend in $F_0$ and $P_{sb}$, an issue that has been undecided despite considerable discussion in the recent literature. The most important conclusion of our own experiments and a detailed analysis of data published in the literature is that the issue is genuinely not decidable, unless there is agreement about the way in which downtrend in $F_0$ and $P_{sb}$ are defined. In our model of $F_0$ control presented in this paper we take the view that $F_0$ and $P_{sb}$ both have a global component, and that these components are related by definition. Other models or definitions of $F_0$ downtrend, like a line fitted through the $F_0$ peaks (the topline), include effects of other factors affecting $F_0$ besides $P_{sb}$; therefore, these definitions (or models) of $F_0$ downtrend do not allow a direct link with downtrend in $P_{sb}$. Also, we have presented data and arguments from our own experiments and from the literature in favour of a tight and precise control of $P_{sb}$ and the underlying respiratory system. Therefore, the phonetic implementation component of any intonation model should include a role for $P_{sb}$.

This research was supported by the Foundation for Linguistic Research, which is funded by the Netherlands Organization for Scientific Research (N.W.O.). Special thanks are due to Haskins Laboratories where one of the experiments was carried out, especially to Dr. Thomas Baer who made this possible. I also express my gratitude to Dr. Hiroshi Muta and Dr. Philip Blok, who inserted the EMG electrodes and the pressure catheter in the experiments in New Haven and Nijmegen, respectively.

References


