

Downtrend in F_0 and P_{sb}

Helmer Strik and Louis Boves

University of Nijmegen, Department of Language and Speech, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

(Received 11th September 1993, and in revised form 10th October 1994)

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_1), 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

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 reiterant 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_{sb} 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_{sb} , 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_{sb} 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_{sb}) 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,

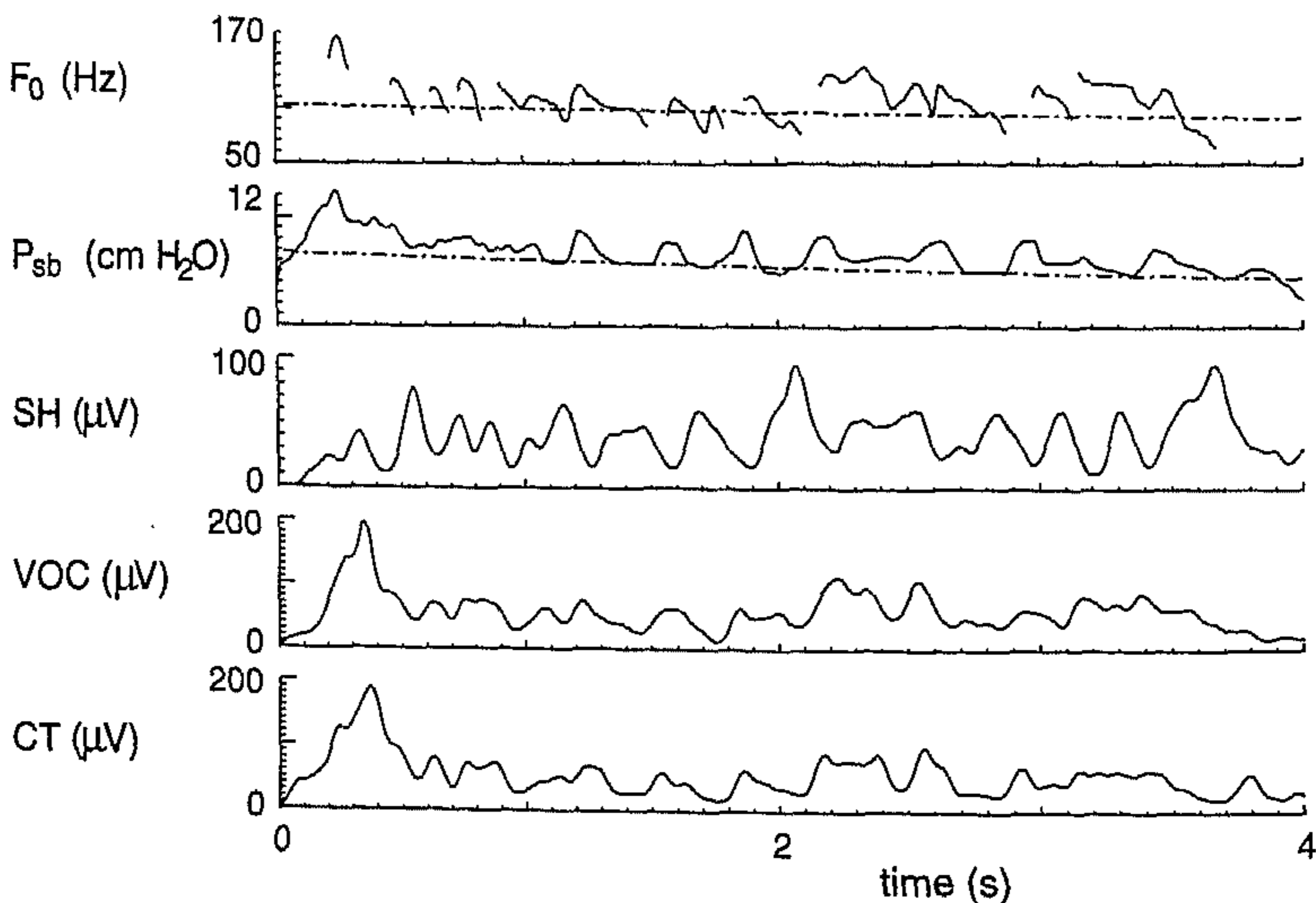


Figure 1. Average physiological signals for the Dutch utterance. "Piet slikte gisteren zijn vierentwintig gele pillen liever in stilte met bier" (LU1) spoken by subject LB. Also shown in the first and second panel are the global trend lines $F_{0,g}$ and $P_{sb,g}$, respectively (dashed-dotted lines).

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_l), nor any resets of F_0 or P_{sb} .

Our definition of a global component is a gradual change spanning the total duration of an utterance. Therefore, in our model P_{sb} and F_0 have a global component. The global component of F_0 and P_{sb} in our model will be called $F_{0,g}$ and $P_{sb,g}$ respectively. In this article, the terms $F_{0,g}$ and $P_{sb,g}$ 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_{sb} in a global and a local component is shown in Fig. 1. $P_{sb,g}$ was obtained by manually fitting an exponential function through most of the valleys of P_{sb} (Fig. 1). Because it is assumed that F_0 varies linearly with P_{sb} (Titze, 1989), $F_{0,g}$ was defined in the following way: $F_{0,g} = B_0 + B_1 * P_{sb,g}$. 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_{sb} 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_{sb} 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_{sb} 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_{sb} 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_{sb} 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_{sb} ($P_{sb,l} = P_{sb} - P_{sb,g}$) 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 $P_{sb,l}$) 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 H₂O (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 H₂O (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

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_{sb} values for first ($P_{sb}(T_1)$) and last ($P_{sb}(T_2)$) voiced sample in cm H₂O, total fall of P_{sb} ($dP_{sb} = P_{sb}(T_1) - P_{sb}(T_2)$) in cm H₂O, average rate of change of P_{sb} (dP_{sb}/T) in cm H₂O/s, $FPR_1 = dF_0/dP_{sb}$ in Hz/cm H₂O, and the regression coefficient between F_0 and P_{sb} (FPR_2) in a multiple regression equation, also in Hz/cm H₂O (for explanations, see also the text).

utt	Subject LB						Subject HB					
	Declarative utterances				Questions		Declarative utterances				Questions	
	SU1	SU2	LU1	LU2	SU3	LU3	SU1	SU2	SU3	SWC	SU4	LU4
N	234	226	558	524	222	490	314	342	288	680	260	435
T	1.42	1.41	3.46	3.40	1.31	3.18	1.66	1.78	1.54	3.62	1.39	2.40
$F_0(T_1)$	150	136	147	136	121	138	119	113	121	132	118	114
$F_0(T_2)$	65	67	66	79	167	169	102	106	102	104	200	188
dF_0	85	69	81	57	-46	-31	17	7	19	28	-82	-74
dF_0/T	60.1	49.1	23.4	16.7	-35.2	-9.7	10.2	3.9	12.3	7.7	-59.0	-30.8
$P_{sb}(T_1)$	9.58	9.92	11.64	11.82	8.44	10.95	6.13	6.47	6.29	5.86	5.83	6.04
$P_{sb}(T_2)$	3.44	3.50	4.82	4.57	4.36	5.10	2.33	1.64	1.42	1.77	4.10	3.96
dP_{sb}	6.14	6.42	6.82	7.25	4.08	5.85	3.80	4.83	4.87	4.09	1.73	2.08
dP_{sb}/T	4.34	4.57	1.98	2.13	3.12	1.84	2.29	2.71	3.16	1.13	1.24	0.87
FPR_1	13.9	10.8	11.9	7.87	-11.3	-5.30	4.47	1.45	3.90	6.84	-47.5	-35.5
FPR_2	3.97	7.63	2.30	4.58	6.48	4.42	3.20	3.02	4.79	3.78	62.5	4.12

calculated: $dF_0 = F_0(T_1) - F_0(T_2)$, $dP_{sb} = P_{sb}(T_1) - P_{sb}(T_2)$, $FPR_1 = dF_0/dP_{sb}$. The total fall in F_0 and P_{sb} from T_1 up to T_2 (dF_0 and dP_{sb} , respectively) is used to model the downtrend in F_0 and P_{sb} , 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_{sb} 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_{sb} is positive (Table I). For subject LB dP_{sb} is always larger than for subject HB. For both subjects, dP_{sb} for the interrogative utterances is smaller than dP_{sb} 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_{sb} 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 H₂O and one is smaller than 2 cm H₂O. 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

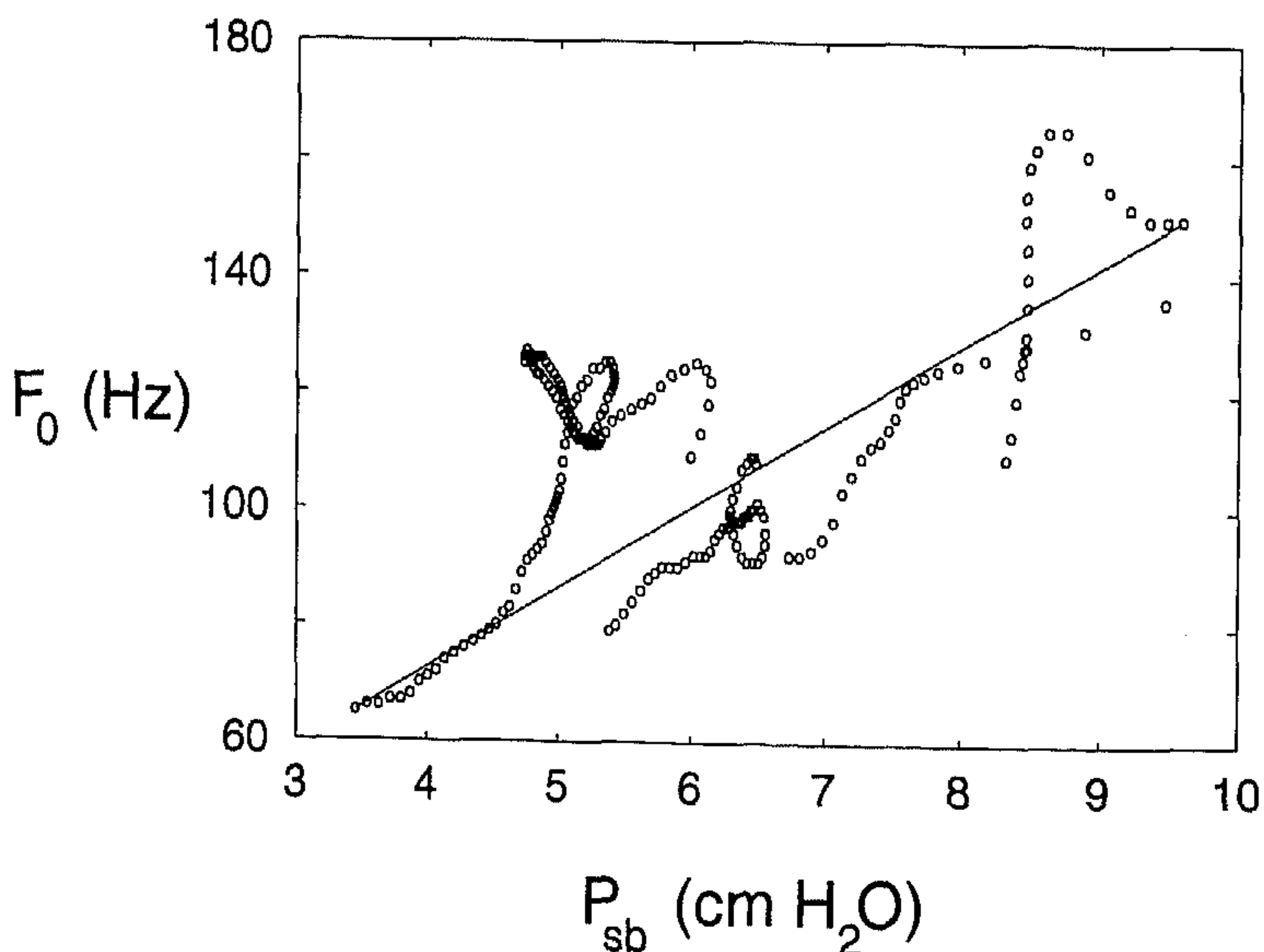


Figure 2. F_0 as a function of P_{sb} for the Dutch utterance "Piet slikte zijn pillen met bier" (SU1) spoken by subject LB. The straight line is the line connecting the first and the last voiced frame. FPR_1 is the slope of this line.

