THE FUNDAMENTAL FREQUENCY - SUBGLOTTAL PRESSURE RATIO

Helmer Strik and Louis Boves.

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

ABSTRACT

It is known that subglottal pressure (P_{sb}) is a major factor in the control of fundamental frequency (F_o) in speech. Yet, the details of this relation remain unclear. Estimates of the F_o to P_{sb} ratio (FPR) from speech and special phonation tasks yield values between 5 and 15 Hz/cmH_2O [1,2,3,4]. In another type of experiments pressure variations are induced externally, either subglottally or supraglottally. The FPR's measured in these experiments tend towards values of 2-5 Hz/cmH_2O [5,6,7,8]. There seems to be no a priori reason for the FPR to be different in both kinds of experiments. After all, the voice source is the same and why should it behave differently during both kinds of phonation tasks? Therefore we carried out experiments that aimed at resolving this discrepancy.

I. THE FPR IN EXPERIMENTS WITH INDUCED PRESSURE VARIATIONS

INTRODUCTION

The FPR in experiments with artificially induced pressure variations was studied first, because we had some ideas why estimates of the FPR in these experiments could be too low. These ideas are described below, and are formalized in three hypotheses.

Except for P_{sb} there are other factors that control F_o. If we want to know the effect of P_{sb} alone on F_o then we must check whether all other factors are constant. It is known that F_o is also controlled by the laryngeal muscles. Baer [5] studied the influence of the laryngeal muscles on the FPR in an experiment in which the subject is pushed on the chest to increase P_{sb}. He found a consistent increase in the EMG activity of vocalis (VOC) and interarytenoid 30-40 ms after each push. Even for the fastest laryngeal muscles it takes about 15-20 ms before a change in the activity of a muscle is followed by a change in F_o [9,10]. So the first 45-60 ms following a push the laryngeal muscles probably do not affect F_o. Baer calculated the FPR during the first 30 ms and found a value of 2-4 Hz/cmH_2O in the chest register, a value that did not deviate from the values reported earlier by others. We did not reexamine the effect of the laryngeal muscles on the FPR.

The first hypothesis:

a sudden rise in P_{sb} is followed by a rise in P_{sp}.

In most experiments either sub- or supraglottal pressure (P_{sp}) is measured and varied, while the other pressure signal (P_{sp} resp. P_{sb}) is not measured. During sustained phonation of a vowel the impedance of the glottis is high but finite. A change of the pressure on either side of the glottis could leak through the glottis. If this would happen the change in transglottal pressure (P_t) is smaller than the change in the measured pressure signal. Because it is really P_t that controls F_o [11], it is also the change in P_t that has to be related to a change in F_o. The effect would be that the estimated FPR is smaller than the ratio between change in F_o and P_t.

The second hypothesis:

a change in F_o lags a change in P_{sb}.

The scatter plots of F_o versus P_{sb} in Baer's article [5] exhibit hysteresis. The hysteresis is already visible during the first 45 ms, so before laryngeal muscle activity could influence F_o. This could be an indication that the F_o change lags the P_{sb} change. During the sustained vowel the vibratory system is in a steady state. When P_{sb} is changed it takes some time for the vocal folds to reach a new steady state. The time constant of this adaptation process depends on the total P_{sb} change. Furthermore, this lag would only show up if the time constant of the P_{sb} change is less than the time constant of the adaptation process. In speech the rate of P_{sb} change during an utterance of about 1-8 cmH_2O/s is probably slow enough for the vocal folds to adjust almost instantaneously to the new vibratory conditions. Both Ladefoged [6] and Baer [5] used short pushes to vary P_{sb}. During these pushes the estimated rate of P_{sb} change is substantially larger than the aforementioned rate of P_{sb} change in speech. If the changes in F_o would lag the changes in P_{sb}, then the duration of their pulsatile P_{sb} changes could be too short for the vocal folds to reach a new steady state. The result would be an underestimation of dF_o, and hence an underestimation of dF_o/dP_{sb}.

The third hypothesis:

the FPR is different in P_{sb} rising and lowering.

In utterances that exhibit declination both F_o and P_{sb} decrease during the course of the utterance. This is most clearly seen in declarative utterances with a single accent early in the utterance. In these cases the FPR is calculated for decreasing F_o and P_{sb}. On the other hand, in experiments where P_{sb} is changed by pushing on the chest the FPR is calculated for increasing F_o and P_{sb}. Differences between F_o rising and F_o falling have been reported and Breckenridge [12] summarizes them by stating that "it has been found that falling tones are more common in the world's languages than rising tones, can be produced faster, and furthermore fall more than rising tones rise." Maybe the FPR is different for P_{sb} rising and lowering, i.e. the ratio in lowering is higher.

In short, three hypotheses were postulated that could explain why estimates of the FPR in experiments with induced pressure changes are too low: 1. a rise in P_{sb} is followed by a rise in P_{sp} 2. a change in F_o lags the change in P_{sb} 3. the FPR is different in P_{sb}

rising and lowering These three hypotheses were tested with the data of an experiment.

METHOD

An experiment was carried out in which simultaneous recordings of acoustic signal, electroglottogram (EGG), P_{sb}, P_{sp} and sternohyoid (SH) were obtained while a subject sustained a vowel /a/ at a comfortable F_o and intensity level. During phonation he was pushed on the chest to increase P_{sb}, the chest was held down to keep P_{sb} high, and finally the chest was released again to lower P_{sb}. In normal speech the fall of P_{sb} during an utterance generally varies from 2 to 12 cmH_2O (see section II). In earlier experiments the magnitude of the induced pressure change was about 1-4 cmH_2O [5,7,8]. We tried to induce larger pressure variations. The P_{sb} changes were induced as fast as possible, in order to produce a large P_{sb} gradient.

All measured signals were stored on a 14-channel instrumentation recorder (TEAC XR-510). The signals are A/D-converted off-line at a 10 kHz sampling rate. F_o was calculated from the EGG signal with a frame rate of 200 frames/s. The pressure signals were low-pass filtered and downsampled to 200 Hz.

RESULTS AND DISCUSSION

The results of this experiment were used to test the three hypotheses. The hypotheses were tested in the same order as they are presented in the introduction above. In Figure 1 the F_o, P_{sb} and P_{sp} signals are shown for one of the pushes.

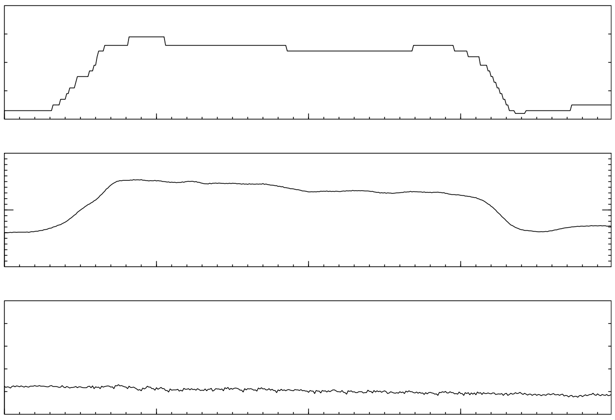

Figure 1. Fo, P_{sb} en P_{sp} during a chest push.

The fact that a change in P_{sb} is followed immediately by a change in F_o indicates that there must be a direct relation between these two variables. It is observed that we succeeded fairly well in keeping P_{sb} high for some time. In all cases P_{sb} decreased during the time that the chest was held down. This could be caused by a partial release of the chest, an adjustment of the respiratory muscles, or it could be a by-product of the decreasing lung volume. In the example in Fig. 1 the stepwise increase in P_{sb} was 9.2 cmH_2O, while the P_{sb} sudden decrease was 7.0 cmH_2O. This means that we also succeeded in inducing pressure variations of substantial magnitude. Both the average rate of change and the maximum rate of change are about the same during rising and lowering (±25 cmH_2O/s for the average resp. ±55 cmH_2O/s for the maximum). This value is much larger than the rate of P_{sb} change during speech utterances, that is known to be in the range of 3-8 cmH_2O/s (see section II), and therefore the P_{sb} changes seem fast enough to test whether there is a lag between F_o and P_{sb} changes. For three chest pushes the P_{sb} variation was as intended: the rise and fall are fast and large enough, and P_{sb} is kept high for some time. Particularly the data of these pushes are used to test the hypotheses. This is discussed below.

During P_{sb} rising and lowering no significant changes in P_{sp} were observed, as can be seen from the example in Fig. 1. This was the case for the three 'succesful' pushes mentioned above, but also for all other pushes. A P_{sb} rise was never followed by a P_{sp} rise, so our first hypothesis was rejected.

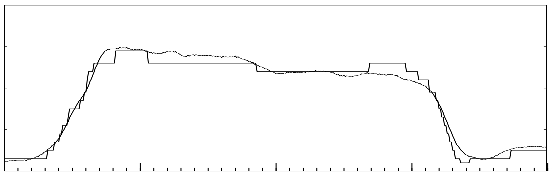

Figure 2. F_o and P_{sb} during a chest push.

The F_o and P_{sb} signals of Figure 1 are plotted together in Figure 2. F_o changes instantaneously with P_{sb}, even if the total P_{sb} change is ±9 cmH_2O and if the rate of P_{sb} change is ±55 cmH_2O/s. A lag between F_o and P_{sb} was not found. The voice source apparently is capable of adjusting very fast to changing phonatory conditions.

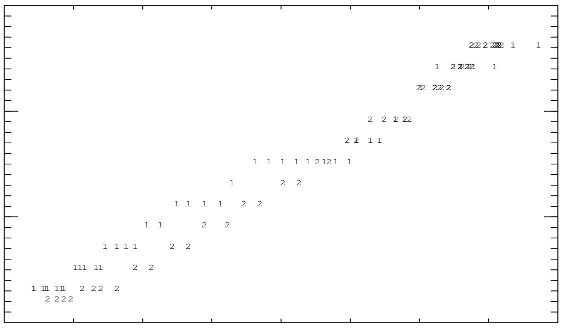

Figure 3. $F_o(P_{sb})$ during P_{sb} rising (∗) and lowering (+).

A scatter plot of F_o versus P_{sb} is shown in Figure 3. Shown are the data during P_{sb} rising (∗) and lowering (+). One can see that the FPR is almost the same during rising and lowering. A substantial difference in the FPR during rising and lowering was not observed.

CONCLUSIONS

All three postulated hypothesis were falsified. At the moment there seems to be no reason to doubt the values of the FPR found in the experiments with induced pressure variations. Therefore the values obtained from measurements on normal speech have to be questioned.

II. THE FPR IN SPEECH

INTRODUCTION

There are two mutually exclusive explanations why estimates of the FPR in speech utterances showing F_o declination are larger than estimates in experiments with induced pressure variations: the FPR is really larger in speech, or the estimates obtained from measurements in speech are wrong. The second explanation seemed more probable to us, so we first examined the methods that are used to calculate the FPR in the experiments on declination [1,2,3,4].

Usually the F_o and P_{sb} values are taken at two instants, one near or at the beginning (T_i) and one near or at the end (T_f) of an utterance. An estimate of the FPR is then calculated with these values:

$$FPR_1 = [F_o(T_i) - F_o(T_f)]/[P_{sb}(T_i) - P_{sb}(T_f)]$$

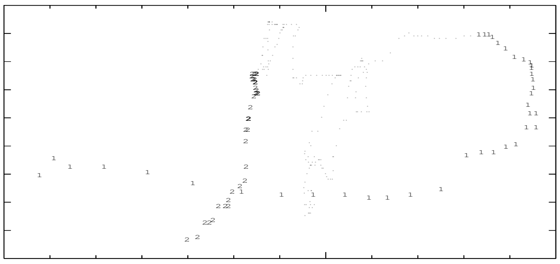

Figure 4. $F_o(P_{sb})$ during first 200 ms (∗), during last 200 ms (+). and during intermediate period (.).

In a plot of F_o as a function of P_{sb} FPR_1 is the slope of the line connecting the data measured at T_i and T_f. In Figure 4 a scatter plot of F_o versus P_{sb} is given for one of the sentences of this experiment. Shown are the first 40 voiced samples (∗), the last 40 voiced samples (+), and the intermediate samples (.). It can be seen that the value of FPR_1 strongly depends on the exact choice of T_i and T_f. Compared to the data in Figure 3 the data are much more scattered here because apart from P_{sb} there are many other physiological processes that influence F_o. This makes it hazardous to make an estimation based on the values at two instants only. It would just be a matter of coincidence if the influence of all other factors on F_o is the same at those two instants. A statistically better method would be to calculate the regression from P_{sb} on F_o. The slope of the regression line would be a better estimate of the FPR, because it takes into account all F_o,P_{sb} pairs, not just two of them. Define:

FPR_2 = regression coefficient between F_o and P_{sb}

The fact that the calculated FPR in experiments on declination (FPR_1) is almost always larger than 2-5 Hz/cmH$_2$O (the value obtained in experiments with induced pressure variations) was an indication that the other F_o regulating processes could participate in the decline of F_o, i.e. their influence on F_o could be such that the total fall of F_o is larger than the fall of F_o resulting from the fall of P_{sb} alone. If this is the case then the regression coefficient between F_o and P_{sb} is not a good measure of the FPR in speech. F_o first has to be corrected for the influence of other variables. This is achieved by partitioning out the effects of the additional factors from F_o. The regression coefficient between corrected F_o (F_o') and P_{sb} would then be a better estimate of the rate of F_o change resulting from a change in P_{sb} alone. Define:

FPR_3 = regression coefficient between F_o' and P_{sb}

Our hypothesis is that the true FPR is the same in 'normal speech' and sustained phonation with induced pressure variations. Estimates of the FPR in 'normal speech' (FPR_1) often are too high because other processes also participate in the decline of F_o. To test this hypothesis two experiments were carried out in which, apart from P_{sb}, also other physiological processes were measured that could control F_o.

METHOD

In the first experiment simultaneous recordings of the acoustic signal, EGG, P_{sb}, lung volume (V_l), and EMG activity of the cricothyroid (CT), vocalis (VOC) and SH were obtained while the subject performed several speech tasks, a.o. the repeated production of a short and a long Dutch sentence. The short sentence was also produced in reiterant form, using either the syllable /fi/ or /vi/. The sentences had to be produced with three different intonation contours, i.e. a 'flat hat pattern' (FH), two 'pointed hats' (PH) and question intonation (Q). Each of the 12 sentence-contour pairs (4 sentences x 3 intonation contours) was repeated at least five times to make averaging possible.

In the second experiment recordings of the supraglottal pressure (P_{sp}) were also made, but activity of the CT was not recorded. Near the end of the experiment the subject was asked to produce an utterance spontaneously. After he spoke this sentence, he was asked to repeat the same sentence 29 times.

Preprocessing of the data was done with the Haskins Laboratories EMG data processing system. The repetitions were time aligned using line-up points. A DTW algorithm was used to correct for the differences in the temporal structure between repetitions. Median values were then calculated for all variables. The exact procedure of data measurement and data processing is described in [11].

RESULTS AND DISCUSSION

The resulting signals were used to calculate the values given in Table I. The values for the utterances in which all syllables are replaced by /vi/ were deviating. In these sentences voicing starts well before the initial peak in F_o and P_{sb}. As a result the F_o and P_{sb} values are small for the first voiced sample, and dF_o and dP_{sb} are small too. We could have chosen another instant (T_i) to measure F_o and P_{sb}, but that is beyond the scope of this paper. The total fall in P_{sb} varied between 4.0 and 11.9 cmH$_2$O, and the overall rate of P_{sb} change varied between 1.7 and 8.1 cmH$_2$O/s.

			SHORT			/FI/			/VI/			LONG	
utt.	SU	FH	PH	Q	FH	PH	Q	FH	PH	Q	FH	PH	Q
N	297	233	225	220	147	129	131	262	255	254	553	515	487
T	2.3	1.4	1.4	1.3	1.2	1.2	1.2	1.3	1.3	1.3	3.5	3.4	3.2
dF0	30	85	69	-46	51	56	-26	29	30	-77	97	73	-27
dF0/T	13	61	49	-35	43	47	-22	22	23	-59	28	22	-8
dPsb	4.3	9.5	11.3	6.2	8.0	8.7	7.8	-2.9	-3.4	-2.1	11.9	10.9	10.0
dPsb/T	1.9	6.8	8.1	4.8	6.6	7.2	6.5	-2.2	-2.6	-1.6	3.4	3.2	3.1
FPR1	7.0	8.9	6.1	-7.4	6.4	6.5	-3.3	-10	-8.8	37	8.1	6.7	-2.7
FPR2	6.9	4.0	6.0	-0.2	4.9	5.0	1.1	2.2	4.8	0.3	4.7	4.8	3.0
FPR3	3.9	1.5	3.3	2.5	2.1	2.2	1.9	1.8	1.5	1.5	0.8	1.8	1.7

Table I. Listed from top to bottom are utterance type, the number of voiced samples (N), length of the utterance (T), total fall of F_0 from first voiced sample to last voiced sample (dF_0), average rate of change of F_0 per time unit (dF_0/T), total fall of P_{sb} from first voiced sample to last voiced sample (dP_{sb}), average rate of change of P_{sb} per time unit (dP_{sb}/T), dF_0/dP_{sb} (FPR_1), slope of regression line of F_0 and P_{sb} (FPR_2), and slope of regression line of F_0' and P_{sb} (FPR_3).

The values of FPR_1 and FPR_2 for the questions are not relevant, because F_o rises markedly near the end of these sentences. The other values of FPR_1 vary from 6.1 to 8.9 Hz/cmH$_2$O. This is in agreement with the results of previous studies [1,2,3,4], and therefore these sentences seem suitable to test our hypothesis.

The values of FPR_2 for non-questions always are smaller than the values of FPR_1, and vary between 4.0 and 6.9 Hz/cmH$_2$O. But one has to be careful in interpreting these values. The value of a regression coefficient is dependent on the value of the correlation coefficient, and therefore a smaller correlation coefficient would result in a smaller regression coefficient. In any case, the values of FPR_2 are stil higher than the FPR values obtained in experiments with artificially induced P_{sb} variations.

The results for one sentence (long-FH) are shown in Figure 5. In most sentences CT and VOC were especially active during the first syllable, and their activity was suppressed at the end. This effect can also often be observed in the data of previous experiments on declination in which muscle activity was measured [1,2,3,4]. The peak activity of these F_o raising muscles is much larger during a stressed syllable at the beginning than during a stressed syllable at the end. And if the first syllable is not stressed, then CT and VOC still show increased activity. On the average the F_o raising muscles CT and VOC are more active at the beginning than at the end of utterances.

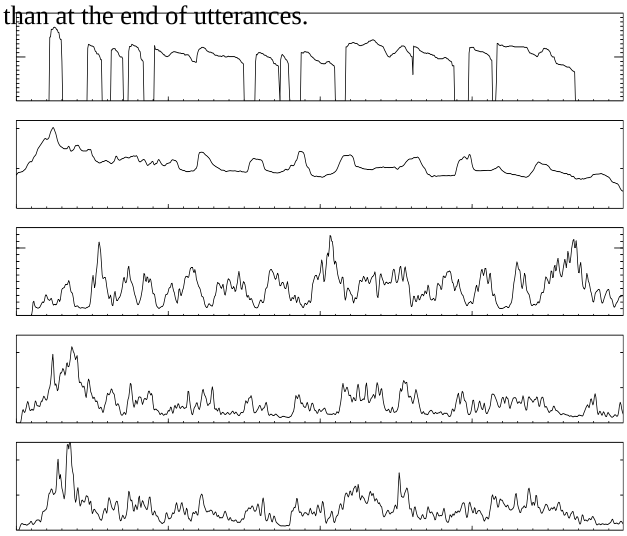

Figure 5. F_o, P_{sb}, SH, CT and VOC signals

It is often observed that the SH is especially active just before phonation [1,13,14], and it is assumed that the SH helps in preparing the larynx for the 'speech mode.' This was also observed in some of the utterances of this experiment. Usually SH activity has dropped to its base level when phonation starts. At the end of utterances F_o often falls abruptly (the so called final fall), and often this is accompanied by a rise of SH activity (and a lowering of the larynx). This is observed in the data of the present experiments, but also in the data of previous experiments [1,2,3,4]. On the average the F_o lowering muscle SH is more active at the end than at the beginning of utterances.

Thus it seems that the laryngeal muscles participate in the declination of F_o, so part of the decline in F_o is due to the activity of the laryngeal muscles. If we want to calculate the FPR we first have to correct F_o for these influences. This is done by calculating the regression equation between F_o and SH and VOC for the average signals of the spontaneous utterance, and the regression equation between F_o and SH and CT for the other 12 sentences. The value of FPR_3 is then calculated. Except for the long-FH-type the values vary between 1.5 and 3.3. Again we want to stress that regression coefficients do depend on the correlation between the variables. For instance in the long-FH-type the correlation was extremely low causing the value of FPR_3 to be very low. Still, if we compare the values of FPR_3 with those of FPR_2 we see that correction for the influence of two important laryngeal muscles resulted in a lowering of the estimate of the FPR in all non-questions.

For the questions the F_o rise at the end is mainly controlled by the combined activity of CT and VOC. The value of FPR_3 is corrected for this increase in CT activity, and therefore the value of FPR_3 is also relevant for questions. The values thus obtained are in the same range as the FPR_3 values for declarative utterances.

CONCLUSIONS

The data obtained in the two experiments described above do support our hypothesis that the FPR is the same in speech and in experiments with induced pressure variations. Our data, and data of previous experiments on declination, suggest that laryngeal muscles participate in the F_o declination during an utterance.

ACKNOWLEDGEMENTS

This research was supported by the foundation for linguistic reasearch, which is funded by the Netherlands Organization for the Advancement of Scientific Research N.W.O. Special thanks are due to Haskins Laboratories were one of the experiments was carried out; to dr. Thomas Baer who helped organizing and running the experiment at Haskins; to dr. Hiroshi Muta who inserted the EMG electrodes and the subglottal pressure sensor in the experiment at Haskins; and to dr. Philip Blok who inserted the EMG electrodes and the pressure catheter in the other two experiments.

REFERENCES

[1] Collier, R. (1975). Physiological correlates of intonation patterns. J. Acous. Soc. Am. 58: 249-255.

[2] Maeda, S. (1976). A characterization of American English intonation. Ph.D.thesis, MIT, Cambridge.

[3] Gelfer, C.; Harris, K.; Collier, R. and Baer, T. (1983). Is Declination Actively Controlled? In: Titze, I.R. and Scherer, C. (eds.), Vocal Fold Physiology. The Denver Center for the Performing Atrs, Inc., Denver, Colorado.

[4] Collier, R. and Gelfer, C.E. (1984). Physiological Explanations of F_o Declination. In: Van den Broecke, M.P.R. and Cohen, A. (eds.), Proc. of the tenth Int. Congres of Phonetic Sciences. Foris Publications Holland, Dordrecht.

[5] Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. J. Acoust. Soc. Am. 65: 1271-1275.

[6] Ladefoged, P. (1963). Some physiological parameters in speech. Language and Speech 6: 109-119.

[7] Rothenberg, M. and Mahshie, J. (1986). Induced transglottal pressure variations during voicing. J. of Phon. 14: 365-371.

[8] Baken, R.J. and Orlikoff, R.F. (1987). Phonatory Response to Step-Function Changes in Supraglottal Pressure. In: Baer, T.; Sasiki, C. and Harris, K. (eds.), Laryngeal Function in Phonation and Respiration. College-Hill Press, Boston, Massachusetts.

[9] Sawashima, M. (1974). Laryngeal research in experimental phonetics. In: Sebeok, T.A. et al (eds.), Current Trends in Linguistics, Vol. 12: 2303-2348. Mouton, The Hague.

[10] Atkinson, J.E. (1978). Correlation analysis of the physiological features controlling fundamental voice frequency. JASA 63: 211-222.

[11] Strik, H. and Boves, L. (1988). Averaging physiological signals with the use of a DTW algorithm. Proceedings SPEECH'88, 7th FASE Symposium, Edinburgh, Book 3: 883-890.

[12] Breckenridge, J. (1977). Declination as a phonological process. Bell Labatories Technical Memorandum, Murray Hill.

[13] Hirose, H. and Sawashima, M. (1981). Functions of the laryngeal muscles in speech. In: K.N. Stevens and M. Hirano (eds.), Vocal Fold Physiology. University of Tokyo press, Tokyo.

[14] Strik, H. and Boves, L. (1987). Regulation of intensity and pitch in chest voice. Proceedings 11th International Congres of Phonetic Sciences, Tallinn, Vol. VI: 32-35.