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 (Psb) is a major factor in the control of fundamental frequency (F0) in speech. Yet, the details of this relation remain unclear. Estimates of the F0 to Psb ratio (FPR) from speech and special phonation tasks yield values between 5 and 15 Hz/cmH2O [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/cmH2O [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 under 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 Psb there are other factors that control F0. If we want to know the effect of Psb alone on F0 then we must check whether all other factors are constant. It is known that F0 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 Psb. 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 F0 [9,10]. So the first 45-60 ms following a push the laryngeal muscles probably do not affect F0. Baer calculated the FPR during the first 30 ms and found a value of 2-4 Hz/cmH2O 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 Psb is followed by a rise in F0.

In most experiments either sub- or supraglottal pressure (Psp) is measured and varied, while the other pressure signal (Psp resp. Psb) 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 (Pt) is smaller than the change in the measured pressure signal. Because it is really Pt that controls F0 [11], it is also the change in Psb that has to be related to a change in F0. The effect would be that the estimated FPR is smaller than the ratio between change in F0 and Pt.

The second hypothesis:

a change in F0 lags a change in Psb.

The scatter plots of F0 versus Psb in Baer’s article [5] exhibit hysteresis. The hysteresis is already visible during the first 45 ms, so before laryngeal muscle activity could influence F0. This could be an indication that the F0 change lags the Psb change. During the sustained vowel the vibratory system is in a steady state. When Psb 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 Psb change. Furthermore, this lag would only show up if the time constant of the Psb change is less than the time constant of the adaptation process. In speech the rate of Psb change during an utterance of about 1-8 cmH2O/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 Psb. During these pushes the estimated rate of Psb change is substantially larger than the aforementioned rate of Psb change in speech. If the changes in F0 would lag the changes in Psb then the duration of their pulsatile Psb changes could be too short for the vocal folds to reach a new steady state. The result would be an underestimate of dF0/dPsb and hence an underestimation of δF0/δPsp.

The third hypothesis:

the FPR is different in Psb rising and lowering. In utterances that exhibit declination both F0 and Psb 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 Psb and Psb. On the other hand, in experiments where Psb is changed by pushing on the chest the FPR is calculated for increasing F0 and Psb. Differences between F0 rising and Psb 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 Psb 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 Psb is followed by a rise in F0. 2. a change in F0 lags the change in Psb. 3. the FPR is different in Psb rising and lowering. These three hypotheses were tested with the data of an experiment.
METHOD
An experiment was carried out in which simultaneous record-
ings of acoustic signal, electroglottogram (EGG), \( P_{ba}, P_{sp} \) and
sternohyoid (SH) were obtained while a subject sustained a
vowel /a/ at a comfortable \( F_0 \) and intensity level. During phona-
tion he was pushed on the chest to increase \( P_{ba} \), the chest was
held down to keep \( P_{ba} \) high, and finally the chest was released
again to lower \( P_{ba} \). In normal speech the fall of \( P_{ba} \) during an ut-
erance generally varies from 2 to 12 cmH\(_2\)O (see section II). In
earlier experiments the magnitude of the induced pressure
change was about 1-4 cmH\(_2\)O [5,7,8]. We tried to induce larger
pressure variations. The \( P_{ba} \) changes were induced as fast as possible,
in order to produce a large \( P_{ba} \) gradient.

All measured signals were stored on a 14-channel instrumen-
tation recorder (TEAC XR-510). The signals are A/D-converted
off-line at a 10 kHz sampling rate. \( F_0 \) was calculated from the
EGG signal with a frame rate of 200 frames/s. The pressure sig-
nals were low-pass filtered and downsampled to 200 Hz.

RESULTS AND DISCUSSION

The results of this experiment were used to test the three hy-
potheses. The hypotheses were tested in the same order as they
are presented in the introduction above. In Figure 1 the \( F_0, P_{ba} \)
and \( P_{sp} \) signals are shown for one of the pushes.

![Figure 1. \( F_0, P_{ba} \) and \( P_{sp} \) during a chest push.](image)

The fact that a change in \( P_{ba} \) is followed immediately by a change
in \( F_0 \) indicates that there must be a direct relation between these
two variables. It is observed that we succeeded fairly well in
keeping \( P_{ba} \) high for some time. In all cases \( P_{ba} \) 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 \( P_{ba} \) was
9.2 cmH\(_2\)O, while the \( P_{ba} \) sudden decrease was 7.0 cmH\(_2\)O. 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\(_2\)O/s for the average resp. ±55 cmH\(_2\)O/s for the
maximum). This value is much larger than the rate of \( P_{ba} \)
change during speech utterances, that is known to be in the range of
3-8 cmH\(_2\)O/s (see section II), and therefore the \( P_{ba} \) changes
seem fast enough to test whether there is a lag between \( F_0 \) and
\( P_{ba} \) changes. For three chest pushes the \( P_{ba} \) variation was as in-
tended: the rise and fall are fast and large enough, and \( P_{ba} \) is kept
high for some time. Particularly the data of these pushes are used
to test the hypotheses. This is discussed below.

During \( P_{ba} \) 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 'successful' pushes mentioned above, but
also for all other pushes. A \( P_{ba} \) rise was never followed by a \( P_{sp} \)
rise, so our first hypothesis was rejected.

![Figure 2. \( F_0 \) and \( P_{ba} \) during a chest push.](image)

The \( F_0 \) and \( P_{ba} \) signals of Figure 1 are plotted together in Figure
2. \( F_0 \) changes instantaneously with \( P_{ba} \), even if the total \( P_{ba} \)
change is ±9 cmH\(_2\)O and if the rate of \( P_{ba} \) change is ±55
cmH\(_2\)O/s. A lag between \( F_0 \) and \( P_{ba} \) was not found. The voice
source apparently is capable of adjusting very fast to changing
phonatory conditions.

![Figure 3. \( F_0(P_{ba}) \) during \( P_{ba} \) rising (1) and lowering (2).](image)

A scatter plot of \( F_0 \) versus \( P_{ba} \) is shown in Figure 3. Shown are
the data during \( P_{ba} \) rising (1) and lowering (2). One can see that
the FPR is almost the same during rising and lowering. A sub-
stantial 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 \( F_0 \) 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 declina-
tion [1,2,3,4].

Usually the \( F_0 \) and \( P_{ba} \) 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 ut-
terance. An estimate of the FPR is then calculated with these
values:

\[ FPR_1 = \frac{[F_0(T_i) - F_0(T_f)]}{[P_{ba}(T_i) - P_{ba}(T_f)]} \]
Figure 4. F0 (Pab) during first 200 ms (1), during last 200 ms (2), and during intermediate period (3).

In a plot of F0 as a function of Pab, FPR1 is the slope of the line connecting the data measured at T1 and T2. In Figure 4 a scatter plot of F0 versus Pab is given for one of the sentences of this experiment. Shown are the first 40 voiced samples (1), the last 40 voiced samples (2), and the intermediate samples (3). It can be seen that the value of FPR1 strongly depends on the exact choice of T1 and T2. Compared to the data in Figure 3 the data are much more scattered here because apart from Pab there are many other physiological processes that influence F0. 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 F0 is the same as those at those instants. A statistically better method would be to calculate the regression from Pab on F0. The slope of the regression line would be a better estimate of the FPR, because it takes into account all F0-Pab pairs, not just two of them. Define:

\[ \text{FPR2} = \text{regression coefficient between } F_0 \text{ and } P_{ab} \]

The fact that the calculated FPR in experiments on declination (FPR1) is almost always larger than 2-5 Hz/cmH2O (the value obtained in experiments with induced pressure variations) was an indication that the other F0 regulating processes could participate in the decline of F0, i.e. their influence on F0 could be such that the total fall of F0 is larger than the fall of F0 resulting from the fall of Pab alone. If this is the case then the regression coefficient between F0 and Pab is not a good measure of the FPR in speech. F0 first has to be corrected for the influence of other variables. This is achieved by partitioning out the effects of the additional factors from F0. The regression coefficient between corrected F0 (F0') and Pab would then be a better estimate of the rate of F0 change resulting from a change in Pab alone. Define:

\[ \text{FPR3} = \text{regression coefficient between } F_0' \text{ and } P_{ab} \]

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' (FPR1) often are too high because other processes also participate in the decline of F0. To test this hypothesis two experiments were carried out in which, apart from Pab also other physiological processes were measured that could control F0.

METHOD

In the first experiment simultaneous recordings of the acoustic signal, EGG, Pab, lung volume (Vl), 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 sentences consisted of 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 (P0) 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 (SU). 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 1. 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 F0 and Pab. As a result the F0 and Pab values are small for the first voiced sample, and dF0 and dPab are small too. We could have chosen another instant (T1) to measure F0 and Pab, but that is beyond the scope of this paper. The total fall in Pab varied between 4.0 and 11.9 cmH2O, and the overall rate of Pab change varied between 7.1 and 3.1 cmH2O/s.

<table>
<thead>
<tr>
<th>Sentence</th>
<th>SU</th>
<th>FH</th>
<th>PH</th>
<th>Q</th>
<th>PH</th>
<th>Q</th>
<th>PH</th>
<th>Q</th>
<th>FH</th>
<th>Q</th>
<th>PH</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>297</td>
<td>233</td>
<td>225</td>
<td>220</td>
<td>147</td>
<td>129</td>
<td>131</td>
<td>162</td>
<td>262</td>
<td>255</td>
<td>254</td>
<td>353</td>
</tr>
<tr>
<td>T</td>
<td>2.3</td>
<td>1.4</td>
<td>1.3</td>
<td>1.7</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>dF0</td>
<td>50</td>
<td>85</td>
<td>36</td>
<td>51</td>
<td>56</td>
<td>26</td>
<td>29</td>
<td>30</td>
<td>77</td>
<td>97</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>dPab</td>
<td>13</td>
<td>61</td>
<td>49</td>
<td>-35</td>
<td>45</td>
<td>47</td>
<td>22</td>
<td>23</td>
<td>59</td>
<td>28</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>dPab/T</td>
<td>4.3</td>
<td>9.5</td>
<td>11.3</td>
<td>6.2</td>
<td>8.0</td>
<td>7.8</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
<td>7.3</td>
<td>7.3</td>
<td>7.3</td>
</tr>
<tr>
<td>FPR1</td>
<td>7.0</td>
<td>8.9</td>
<td>7.1</td>
<td>-7.4</td>
<td>6.4</td>
<td>6.5</td>
<td>-3.3</td>
<td>-10</td>
<td>-8.8</td>
<td>37</td>
<td>8.1</td>
<td>7.7</td>
</tr>
<tr>
<td>FPR2</td>
<td>6.9</td>
<td>4.0</td>
<td>6.0</td>
<td>-0.2</td>
<td>4.9</td>
<td>5.0</td>
<td>1.1</td>
<td>2.2</td>
<td>4.8</td>
<td>0.3</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>FPR3</td>
<td>3.9</td>
<td>1.5</td>
<td>3.3</td>
<td>2.5</td>
<td>2.1</td>
<td>2.2</td>
<td>1.9</td>
<td>1.8</td>
<td>1.5</td>
<td>1.5</td>
<td>0.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>

DISCUSSION

The values of FPR1 and FPR2 for the questions are not relevant, because F0 rises markedly near the end of these sentences. The other values of FPR1 vary from 6.1 to 8.9 Hz/cmH2O. 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 FPR2 for non-questions always are smaller than the values of FPR1, and vary between 4.0 and 6.9 Hz/cmH2O. 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 FPR2 are still higher than the FPR values obtained in experiments with artificially induced Pab 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 F0 raising muscles is much
larger during a stressed syllable at the beginning than during a stressed syllable at the end. If the first syllable is not stressed, then CT and VOC still show increased activity. On the average the F0 raising muscles CT and VOC are more active at the beginning than at the end of utterances.

Figure 5. F0, Pb, 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 F0 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 F0 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 F0, so part of the decline in F0 is due to the activity of the laryngeal muscles. If we want to calculate the FPR we first have to correct F0 for these influences. This is done by calculating the regression equation between F0 and SH and VOC for the average signals of the spontaneous utterance, and the regression equation between F0 and SH and CT for the other 12 sentences. The value of FPR3 is then calculated. Except for the long-FH-type the values vary between 1.5 and 3.5. 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 FPR3 to be very low.

Still, if we compare the values of FPR3 with those of FPR2 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 F0 rise at the end is mainly controlled by the combined activity of CT and VOC. The value of FPR3 is corrected for this increase in CT activity, and therefore the value of FPR3 is also relevant for questions. The values thus obtained are in the same range as the FPR2 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 F0 declination during an utterance.

ACKNOWLEDGEMENTS
This research was supported by the foundation for linguistic research, 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