A physiological model of intonation

Helmer Strik and Loe Boves

Introduction

There are many articles dealing with the relation between fundamental frequency (F_o) and the underlying physiological processes, which show that subglottal pressure (P_{sb}) and the activity of the cricothyroid (CT), vocalis (VOC), and sternohyoid (SH) muscles are important factors in the control of F_o (Rubin, 1963; Shipp & McGlone, 1971; Collier, 1975; Baer, Gay, & Niimi, 1976; Maeda, 1976; Atkinson, 1978; Shipp, Doherty, & Morrissey, 1979; Hirose & Sawashima, 1981; Gelfer, 1987). However, most of the data described in these papers concern singing and sustained phonation. Moreover, in many studies measurements were obtained for either the respiratory system or the activity of the laryngeal muscles. There are relatively few studies in which simultaneous registrations of respiratory and laryngeal activity were made for running speech. This may be an important reason why it is not completely clear yet how these factors cooperate in the regulation of F_o for running speech.

The purpose of the research reported in this paper is to clarify the relation between F_o and the physiological mechanisms for running speech. To this end simultaneous measurements of laryngeal and respiratory activity were made for two subjects. Our main goal is to propose a comprehensive model for the physiological control of intonation. From our own data, as well as from the literature (Gay et al., 1972; Ladefoged, 1967), it appears that there are differences between subjects in the physiology underlying intonation. To overcome the problems deriving from this variability, we made ample use of data available in the literature. Especially the data found in Ladefoged (1967), Lieberman (1967), Collier (1975), and Gelfer (1987) proved to be useful.

From the outset we did not want to subscribe to one of the extant models of intonation. Instead, we wanted to analyse the data as objectively as possible, i.e. our methodology is essentially data driven. As a consequence, we try to avoid theory-laden terms like 'declination' and 'baseline' as much as possible.

The outline of the article is as follows. Section 1 describes the material and the method used in our research. The results for running speech are presented in section 2. The physiological model resulting from our investigations is given in section 3. Finally, in section 4 we discuss our findings and their relation to previous research.

1. Material and method

For two Dutch male subjects recordings were made of the audio signal, electroglottogram, lung volume (V_l), P_{sb}, SH, and VOC. In addition to these signals the CT was also measured for subject LB, and oral pressure (P_{or}) for subject HB. In the latter case transglottal pressure (P_{tr}) was calculated by taking the difference of P_{sb} and P_{or}.

The measurements were made while the subjects produced sustained vowels and meaning-ful Dutch sentences with different intonation patterns. The subjects repeated each sentence 5 to 8 times. The signals of these repetitions were used to calculate average signals. The method of non-linear time-alignment and averaging (Strik & Boves, 1991) was used to average the signals. An advantage of this method is that it also yields an average F_o contour. All signals shown in the present article are average signals, which are time-aligned with the audio signal. The procedures used for recording and processing the data are described in more detail in Strik & Boves (1992).

2. Running speech

A speaker can use many different physiological mechanisms to control F_o. Therefore, it is remarkable that the within-subject variation between the signals of repetitions of the same utterance is relatively small. This suggests that speakers have a good notion of the manner in which they want to produce an utterance, and that they have good control over these mechan-isms. Consequently, meaningful averaging of the data is possible (Strik & Boves, 1991).

Although there are some individual differences, consistent behaviour between subjects can be observed in the data. Ladefoged (1967) also noted that for his subjects "the results obtained so far are sufficiently consistent to suggest the general pattern of the relationships involved". This consistency led to the discovery of general patterns in the behaviour of P_{sb}, CT, VOC, and SH, that we want to describe in this section.

From our data it appears that both F_o and the physiological signals have two components, viz. a global and a local one. This was also found by Maeda (1980), and Gelfer (1987). In Strik & Boves (1992) we showed that this qualitative observation has a quantitative statistical basis: on a global level P_{sb} explains most of the observed variance of F_o, while on a local level the laryngeal muscles become more important. Our treatment focuses on the linguistically significant aspects of F_o, especially those connected with stress, phrasing and the question-statement distinction. Initial rise and final lowering of F_o are treated separately for reasons that are explained in section 2.2.5.

2.1. Global level

From the recordings shown in Figures 1 and 2 it is apparent that P_{sb} has a global and a local component. The global pattern of P_{sb} will be called $P_{sb,g}$. In most sentences there is a gradual lowering of $P_{sb,g}$. This can also be seen in the data of Lieberman (1967), Collier (1975), and Gelfer (1987).

The shape of $P_{sb,g}$ differs among speakers. For subject LB the shape is concave (see e.g. Figure 1b): the slope is steep initially, and it gradually becomes more flat towards the end. The same pattern is observed for the two subjects in the studies of Collier (1975) and Gelfer (1987), and for speaker 2 in the study of Lieberman (1967). However, for subject HB the pattern is more convex (see e.g. Figure 2a): $P_{sb,g}$ decreases slowly in the beginning, and more rapidly near the end. The same pattern is also found for speakers 1 and 3 in the study of Lieberman (1967). In spite of the differences between subjects, the behaviour is relatively consistent within subjects.

The global reference level of CT, VOC, and SH seems to be constant (see e.g. Figure 1f). In other words, these laryngeal muscles usually do not have a global component. Consequently, the global behaviour of F_o is generally determined by $P_{sb,g}$. The global component of F_o will be called $F_{o,g}$. A downtrend in $P_{sb,g}$ will result in a downtrend in $F_{o,g}$. This downtrend in $F_{o,g}$ has been observed in declarative utterances of many languages.

2.2. Local level

F_o, P_{sb}, and the laryngeal muscles have a local component: if there are local variations in F_o, then local variations in P_{sb} and in the laryngeal muscles are often observed (see e.g. Figures 1f, 2d, 2f). This can also be seen in the data of Ladefoged (1967), Lieberman (1967), Collier (1975) and Gelfer (1987).

2.2.1. Cricothyroid

The conclusion of many studies was that of all physiological factors known to affect F_o, the CT shows the most consistent relation to F_o (Collier, 1975; Maeda, 1976; Atkinson, 1978; Shipp, Doherty, & Morrissey, 1979; Erickson, Baer, & Harris, 1983; Gelfer, 1987). Also in our data we see that for local variations in F_o there usually is a local variation in the activity of the CT. At the moment there is no doubt that the CT is an important factor in the control of F_o. The local variation in CT explains (at least) part of the local variation in F_o.

2.2.2. Vocalis

For local F_o movements we usually observe a covariance of F_o and VOC in our data. This relation was also studied by Maeda (1976) and Atkinson (1978) for sentences with various intonation patterns. No direct relation between VOC and the F_o movements was found by Maeda for a single subject. However, Atkinson did find a positive correlation between VOC and F_o for his subject. VOC is used to control F_o for sustained phonation and singing (Rubin, 1963; Sawashima, Gay, & Harris, 1969; Shipp & McGlone, 1971; Gay et al., 1972; Shipp, Doherty, & Morrissey, 1979). Hirose & Gay (1972) observed an increase in the activity of CT and VOC for stressed vowels in isolated words. Probably there is a synergism of CT and VOC in the control of F_o.

2.2.3. Subglottal pressure

Both measurements (Ladefoged, 1967; Lieberman, 1967; Collier, 1975; Baer, Gay, & Niimi, 1976; Atkinson, 1978; Baken & Orlikoff, 1987; Gelfer, 1987) and modelling (Titze, 1989) have shown that a change in P_{sb} will affect F_o, ceteris paribus. During local F_o movements a covariation of F_o and P_{sb} is often observed in our data, and in the data of Ladefoged (1967), Lieberman (1967), Collier (1975) and Gelfer (1987). Part of this local P_{sb} variation might be due to a change in the impedance of the glottis which, in turn, results from changes in the activity of the laryngeal muscles (e.g. the changes in CT and VOC as noted above). However, part of the P_{sb} variation could also be due to changes in pulmonic activity. For instance, increased activity of the respiratory muscles for stressed syllables was found by Ladefoged (1967) and van Katwijk (1974). Whatever the cause of a P_{sb} variation, the result is a change in F_o.

2.2.4. Sternohyoid

The function of the SH in the control of F_o is not completely understood. Erickson & Atkinson (1976), Maeda (1976) and Erickson, Baer, & Harris (1983) postulated that F_o falls are initiated by a relaxation of the CT, which is followed by increased activity of the SH. Collier (1975) argued that SH cannot be the primary effector of an F_o fall. Atkinson (1978) found a high negative correlation between SH and F_o, while Erickson, Liberman, & Niimi (1977) concluded that the SH has "a slightly negative relation to F_o". For some sentences in our data there is also a small negative correlation between SH and F_o. However, this negative correlation is mainly brought about by the increase in SH and the lowering of F_o at the end of many utterances (the so-called final lowering, see section 2.2.5.). Of course, final lowering will affect the correlation coefficient to a greater extent if the utterances are short, like those used by Atkinson (1978). The SH is probably used in some F_o lowerings, but it is also used for articulatory gestures such as jaw lowering, tongue lowering and retraction. Therefore, the relation between the SH and F_o is probably complex. This is illustrated in Figure 2c. During the F_o lowering there is a peak in the activity of SH, and in this case the SH could have assisted in lowering F_o. But similar peaks can be observed also when F_o increases or remains steadily high. Anyhow, no consistent, transparent relation can be found in our data nor in the data of Collier (1975) and Gelfer (1987).

2.2.5. Initial rise and final lowering

High values of F_o, CT, VOC, and P_{sb} are often observed at the beginning of utterances, both in our data and in the data of Collier (1975), Maeda (1976), and Gelfer (1987). This effect shows up more prominently in the utterances of subject LB, especially in the longer ones, while it is less evident in the utterances of subject HB. In questions this initial rise is slightly reduced compared to the statements.

Towards the end many utterances F_o and P_{sb} often decrease substantially, while there is a marked increase in the SH activity. Final lowering has also been observed by Collier (1975) and Maeda (1976). Increased SH activity and the large drop in P_{sb} usually take place before phonation has stopped. However, in interrogative sentences both changes are often delayed till after the utterance. Furthermore, the small drop in P_{sb}, which sometimes remains, is counter-balanced by a large increase in the activity of CT and VOC. Therefore, the final lowering of F_0 is rarely observed in questions.

The initial rise probably is the result of laryngeal adjustments that are needed to start phonation (prephonatory tuning), while the final lowering could be a preparation for the next inhalation (Wyke, 1983). Both kinds of local F_o variations could therefore be seen as the by-product of physiological manoeuvres that are necessary for speech production. Initial rise and final lowering are not generally used to signal stress, but still they could be linguistically significant.

Prosody plays an important role in communication. It is used, among other things, to mark the boundaries between phrases (Breckenridge, 1977; Cooper & Sorensen, 1981). Pierrehumbert (1979) suggested that the downtrend in F_o and IL may be important in the perception of phrasing. The F_o fall that results from the downtrend in F_o, is often enlarged by

initial rise and final lowering. Consequently, both effects could assist in the signalling of boundaries. In interrogative utterances the indications of a linguistic control of both phenomena are especially clear. In these utterances initial rise and final lowering were often reduced. Of course, a high F_o at the beginning, and especially a lowering of F_o at the end of an utterance would interfere with the desired rising intonation.

From our data it is not manifest whether initial rise and final lowering are linguistically controlled variables, or if they are primarily the by-product of physiological gestures that are needed in speech production. That is the reason why these local F_o movements are treated separately from the other local F_o movements which obviously do have a linguistic purpose.

3. A physiological model of intonation

3.1. The model

In this section we propose a qualitative model of F_o control in running speech. It describes consistent behaviour of P_{sb}, CT, VOC, and SH that was observed in the data of various subjects. Although the SH is consistently used in final lowerings, no transparent relation was found between the SH and other local F_o variations. Therefore, in our model the SH does not play a role in the control of the latter type of local variations.

Intonation and its physiological control take place at two levels, viz. a global and a local level.

CT, VOC, and SH generally do not seem to have a global component. Therefore, the global component of F_o ($F_{o,g}$) is determined by $P_{sb,g}$. $P_{sb,g}$ has a tendency to decline, which could be due to an economic principle. The downtrend in $P_{sb,g}$ will lead to a downtrend in $F_{o,g}$.

At the beginning of utterances CT, VOC, and P_{sb} may have extra high values (initial rise). At the end of utterances SH often shows an increase while P_{sb} drops sharply. If these effects occur during voiced sounds at the end of the utterance, final lowering is observed. Alternatively, SH activity and P_{sb} release may be delayed until after the last voiced sound, in which cases final lowering is absent. The initial rise and final lowering of F_o will add to the F_o fall that results from the downtrend in F_o.

Besides initial rise and final lowering, other local variations in F_o often occur. These local variations in F_o are generally caused by variations in CT, VOC, and P_{sb}. F_o can be raised by increasing CT, VOC and P_{sb}, and F_o can be lowered by decreasing P_{sb} and relaxing CT and VOC.

3.2. Some remarks

Compatible behaviour has been found in the data of Dutch, British English and American English subjects. Our model describes the behaviour, that seems to be shared by many speakers. Individual differences were found, though, and it is always possible that an individual uses a different strategy to control intonation.

The SH is usually involved in final lowering of F_o. In our model the other F_o lowerings are brought about by a relaxation of CT, VOC, and P_{sb}, i.e. the same mechanisms used to raise F_o are also used to lower it. According to our data and the data of Collier (1975) and Gelfer (1987), no separate mechanism (like SH) seems to be needed to produce these low tones. The strap muscles are probably used to produce very low tones, as during final lowering. It is possible that these extra low tones do not occur often in those parts of utterances that precede final lowering. This would imply that the role of the SH in the control of F_0 in running speech is limited.

The reference line of a laryngeal muscle is the activity observed when the muscle is not active. Consequently, the activity of the CT and VOC can only be lowered if it has been raised previously. For local variations of P_{sb} it is also observed that P_{sb} is first raised, relative to $P_{sb,g}$, and then it is lowered again. Thus it seems that a local lowering of CT, VOC, and P_{sb} is always preceded by a local rise. The question is what happens if a sentence starts with a high F_o that is part of the intonation contour proper (i.e., it is not an initial rise). As there is no such intonation pattern in our data nor in the data of Collier (1975) and Gelfer (1987), we can only speculate on the answer. In this case we would expect CT, VOC, and P_{sb} to rise before phonation has started, and to remain high until the first F_o lowering.

In previous intonation studies the term baseline was used regularly. In general it is defined as a line "drawn near or through the low values of F_o occurring in an utterance" (Cooper & Sorensen, 1981). This baseline will resemble $F_{o,g}$, although they are not identical. In our model $F_{o,g}$ is the global component of F_o, i.e. the component that remains after all local effects have been removed. Initial rise, final lowering, and the rise at the end of questions are considered to be local effects, and thus are not part of $F_{o,g}$. According to the definition given above, they probably are part of the baseline. The baseline also differs from $F_{o,g}$ when F_o is lowered by F_o-lowering mechanisms (e.g., the strap muscles). In that case the baseline will drop below $F_{o,g}$.

4. Discussion

Our physiological model of intonation is based on our own data, and on the data of Lieberman (1967), Ladefoged (1967), Collier (1975) and Gelfer (1987). However, some of the conclusions that were expressed in these articles are different from our conclusions.

Lieberman (1967) made measurements of P_{sb}, but he did not measure the activity of the laryngeal muscles. He observed a resemblance in the behaviour of F_o and P_{sb}, except at the end of interrogative utterances. At the end of questions there was an increase in F_o, while P_{sb} generally did not increase. His assumption was that the activity of the laryngeal muscles increased at the end of questions, but remained relatively steadily otherwise. Based on this assumption he concluded that, apart from questions, F_o is a function of P_{sb} alone. This conclusion can easily be verified by calculating the frequency-to-pressure ratio in his data. The rate of F_o changes that result from a change in P_{sb} alone should be in the range 2-7 Hz/cm H_2O (e.g. Ladefoged, 1967; Baer, 1979). According to Lieberman (1967: 97) this ratio is about 20 Hz/cm H_2O in his data, while Ohala (1990) claims that it is even larger. In any case, P_{sb} alone cannot explain all the variation in F_o, and other mechanism must have been involved. It is likely that the laryngeal muscles were involved, not only at the end of questions but also

in other parts of the utterances. Although we do not agree with his conclusion, our model fits the general pattern in his data: $P_{sb,g}$ gradually declines, and local variations in P_{sb} explain part of the local variations in F_o.

The conclusion of Ladefoged (1967) that both vocal cord tension and P_{sb} contribute to stress is in agreement with our model. He presents data for utterances with stress on the last word, and part of the utterances is also produced with a rising intonation (questions). In his data it can be seen that P_{sb} has a local component, for P_{sb} generally increases for stressed words and at the end of questions. This is also in line with our model. As these P_{sb} increases are present at the end of most of his utterances, P_{sb} of these short utterances is about level or slightly increases. This seems to be in contradiction with our claim that $P_{sb,g}$ is generally decreasing. However, to study the behaviour of $P_{sb,g}$ the local variations in P_{sb} have to be removed. After this has been done, it is likely that $P_{sb,g}$ will decline, also in Ladefoged's data.

The conclusions of Collier (1975) are based on the data of one subject. The majority of the physiological data presented in Gelfer (1987) concern the same subject, while she also shows data for one other subject. Although they do not offer an explicit model, their main conclusions are similar: P_{sb} controls the gradual falling baseline, while local F_o movements are controlled by the CT. They both observed local variations in CT and P_{sb} for local F_o movements, and found that the frequency-to-pressure ratio for these movements is higher than the expected 2-7 Hz/cm H_2O. They argued that as P_{sb} cannot explain all the variation in F_o, it must be the CT that is the most important factor in the control of F_o. However, one can calibrate the F_o-P_{sb} ratio, but it is almost impossible to calibrate the F_o-EMG ratio for a laryngeal muscle. An important reason is that the magnitude of an EMG signal depends on many factors that are difficult to control (for instance, the magnitude is dependent on the exact place of the electrode in the muscle). The conclusion is that one can check whether P_{sb} explains all of the variance in F_o, but the same check cannot be made for a laryngeal muscle. Besides CT other factors could be involved. In fact, P_{sb} and VOC (and probably other factors) are usually involved in the local F_o movements. Because it is difficult to calibrate the F_o-EMG ratio, it is hardly possible to decide on quantitative grounds which factor is most important.

References

Atkinson, J.E. (1978) Correlation analysis of the physiological features controlling fundamental voice frequency. JASA 63: 211-222.
Baer, T. (1979) Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. JASA 65: 1271-1275.

Baer, T.; Gay, T. & Niimi, S. (1976) Control of fundamental frequency, intensity and register of phonation. HLSRSR 45/46: 175-185.

Baken, R.J. & Orlikoff, R.F. (1987) Phonatory responses to step-function changes in supraglottal pressure. In: Laryngeal function in phonation and respiration (T. Baer, C. Sasaki and K.S. Harris, editors), pp. 273-290. Boston: College-Hill Press.

Breckenridge, J. (1977) Declination as a phonological process. Bell Labatories Technical Memorandum, Murray Hill, New Jersey.

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

Cooper, W.E. & Sorensen, J.M. (1981) Fundamental frequency in sentence production. Springer-Verlag, New York.

Erickson, D. & Atkinson, J.E. (1976) The functions of the strap muscles in speech. HLSRSR SR-45/46: 205-210.

Erickson, D.; Baer, T.; Harris, K.S. (1983) The role of the strap muscles in pitch lowering. In: D.M. Bless & J.H. Abbs (eds.), Vocal Fold Physology. College-Hill press, San Diego.

Erickson, D.; Liberman, M. & Niimi, S. (1977) The geniohyoid and the role of the strap muscles. HLSRSR SR-49: 103-110.

Gay, T.; Hirose, H.; Strome, M. & Sawashima, M. (1972) Electromyography of the intrinsic laryngeal muscles during phonation. Annals of otology, rhinology and laryngology 81 (8): 401-409.

Gelfer, C.E. (1987) A simultaneous physiological and acoustic study of fundamental frequency declination. Ph.D. dissertation, Univ. of New York.

Hirose, H. & Gay, T. (1972) The activity of the intrinsic laryngeal muscles in voicing distinction. Phonetica 25: 140-164.

Hirose, H. & 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.

van Katwijk, A. (1974) Accentuation in Dutch: An experimental study. Ph.D. dissertation, Utrecht Univ.

Ladefoged, P. (1967) Three areas of experimental phonetics. Oxford University Press, Oxford.

Lieberman, P. (1967) Intonation, Perception and Language. The M.I.T. Press, Cambridge, Massachusetts.

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

Ohala, J.J. (1990) Respiratory activity in speech. In: Speech production and speech modelling (W.J. Hardcastel & A. Marchal, eds.), pp. 23-53, Kluwer Academic Publishers, Netherlands.

Pierrehumbert, J.B. (1979) The perception of fundamental frequency declination. J. Acoust. Soc. Am. 66: 363-369.

Rubin, H.J. (1963) Experimental studies on vocal pitch and intensity in phonation. The Laryngoscope 8: 973-1015.

Sawashima, M.; Gay, T.J. & Harris, K.S. (1969). Laryngeal muscle activity during vocal pitch and intensity changes. HLSRSR 19/20: 211-220.

Shipp, T. & McGlone, R.E. (1971) Laryngeal dynamics associated with voice frequency change. J. of Speech and Hearing Research 14: 761-768.

Shipp, T.; Doherty, E.T. & Morrissey, P. (1979) Predicting vocal frequency from selected physiologic measures. JASA 66: 678-684.

Strik ,H. & Boves ,L. (1991) A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech. Journal of Phonetics 19, pp. 367-378.

Strik ,H. & Boves ,L. (1992) Control of fundamental frequency, intensity and voice quality in speech. Journal of Phonetics 20, pp. 15-25.

Titze ,I. (1989) On the relation between subglotal pressure and fundamental frequency in phonation. JASA 85 (2), pp. 901-906.

Wyke, B. (1983) Neuromuscular control systems in voice production. In: Vocal Fold Physiology (D.M. Bless & J.H. Abbs, editors), pp. 71-76. San Diego: College-Hill press.