

Title:

Control of fundamental frequency, intensity and voice quality in  
speech

Running title:

Control of  $F_0$ , IL and voice quality

Helmer Strik and Louis Boves,  
University of Nijmegen, Dept. of Language and Speech,  
P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

### **Abstract**

In this paper the control of fundamental frequency ( $F_0$ ), intensity level of the radiated acoustic signal (IL), and voice quality is studied in normal conversational speech. It is shown that the physiological factors that best explain measured features of the speech wave depend on the part of the utterance taken into account. Also, it appears that in speech transglottal pressure ( $P_{tr}$ ) is more important than subglottal pressure ( $P_{sb}$ ). We conclude that currently available mathematical models that describe the waveform of glottal volume flow ( $U_g$ ) lack a number of parameters necessary for a better understanding of the physiological control of the speech parameters investigated in this study.

## 1. Introduction

The relation between  $P_{sb}$  and laryngeal configurations on the one hand, and  $F_0$ , IL, and  $U_g$  on the other is extremely complex. Moreover,  $P_{sb}$  and especially laryngeal configurations and the ways in which they are brought about are difficult to measure. Perhaps due to the measurement problems most investigations of laryngeal control and its effects on the radiated acoustic signal have dealt with sustained vowels produced in widely different ways, rather than with "normal" speech production. In many studies  $F_0$  was varied over several octaves and  $P_{sb}$  over a range from approximately 5 cm aq. to well above 30 cm aq. There are several reasons why the results obtained in those studies may not directly be applied to speech production. In "normal neutral" speech the ranges are much smaller. This may imply that some of the control mechanisms needed to span the wide ranges in 'phonation experiments' are much less important in speech. Also, in sustained vowels oral pressure ( $P_{or}$ ) may be considered equal to atmospheric pressure. But in speech production, where non-negligible constrictions of the vocal tract occur,  $P_{or}$  is much more important. In the present study we have looked into the relation of laryngeal characteristics to IL and  $F_0$  in normal speech. We will touch upon some methodological aspects of the research. Also, we will pay due attention to the role of  $P_{or}$  and  $P_{tr}$ .

## 2. Material and methods

### 2.1. Experimental procedure

The subject in this study was a male native speaker of Dutch, with no experience in phonetics or linguistics and with no history of respiratory or laryngeal dysfunction. During the production of various utterances (sustained vowels, sentences with different intonation patterns) simultaneous recordings of the acoustic signal, electroglottogram (EGG), lung volume (Vl),  $P_{sb}$ ,  $P_{or}$ , and electromyographic (EMG) activity of the sternohyoid (SH) and vocalis (VOC) muscles were obtained. Near the end of the recording session he was asked to produce an utterance spontaneously. He replied by saying (in Dutch): "*Ik heb het idee dat mijn keel wordt afgeknepen door die band*" (I have the feeling that my throat is being

pinched off by that band). After he spoke this sentence, he was asked to repeat it 29 times.

## *2.2. Data recording*

The speech signal was transduced by a condenser microphone (B&K 4134) placed about 10 cm in front of the mouth, and amplified by a measuring amplifier (B&K 2607).

The EGG was recorded with a Fourcin-Abberton laryngograph (Fourcin, 1974).

The pressure signals were recorded using a Millar(R) catheter with four miniature pressure transducers, in the way described by Cranen and Boves (1985). The catheter was introduced into the pharynx via the nose, and then into the trachea via the posterior commissure. It did not have a noticeable effect on phonation (Boves, 1984).

The EMG signals were recorded using hooked-wire electrodes (Hirose, 1971). The electrodes were inserted percutaneously, and correct electrode placement was confirmed by audio-visual monitoring of the signals during various functional manoeuvres.

The perimeter of chest and abdomen were measured with mercury filled strain-gauge wires (Strik and Boves, 1988).

All signals were recorded on a 14-channel instrumentation recorder (TEAC XR-510), using a bandwidth of 5 kHz.

## *2.3. Processing of the data*

All signals were A/D converted off-line at a 10 kHz sampling rate. The files were stored on a microVAX computer. Because of the sluggishness of the articulators it seems sufficient to use a sampling frequency of 200 Hz. Therefore, the goal of preprocessing is to derive physiological signals which all have a sampling rate of 200 Hz.

$F_0$  and IL were calculated with the SIF program of ILS. Both values were calculated every 5 ms, resulting in  $F_0$  and IL signals sampled at a 200 Hz rate.

Pressure signals, chest and abdomen signals were low-pass filtered and downsampled to 200 Hz. Lung volume was calculated from the low-pass filtered chest and abdomen signals.

The integrated rectified EMG was calculated in the way described by Basmajian (1967): first the signal is full-wave-rectified, and then it is integrated over successive periods of 5 ms. The integrator is reset after each integration. Finally, the signal is smoothed by convolving it with a triangular function (base length 35 ms).

There is a time delay between the change of the electric potential of a muscle and the resulting effect in the acoustic signal (Atkinson, 1978). To overcome this delay, all EMG signals were shifted forward over their mean response times.

After preprocessing median signals were calculated with the method of non-linear time-alignment that is described in Strik and Boves (1991), in which the fifth repetition was used as a reference.

#### *2.4. The parameters of the glottal volume flow*

There are a number of different ways to parameterize the glottal volume velocity waveform (Klatt & Klatt, 1990; Fant, 1986; Cranen & Boves, 1987). Although we do not believe that it is the best model from a physiological point of view, we will use the Liljencrants-Fant (LF) model in this paper, mainly because it seems to be the model used in most recent studies. It seems that many of the features of this model are motivated from a perceptual point of view, i.e. by the ease with which they allow one to approximate or explain (spectral) characteristics of the speech wave that are important from a perceptual point of view. However, most of the parameters can also be related to what is known about the physiology of phonation. Specifically, the LF-model allows one to describe the maximum amplitude of the flow during the open glottis interval ( $U_0$ ), the duty cycle of the flow pulses, the amount of skewing of the pulses, the amplitude of  $dU_g/dt$  at the moment of glottal closure ( $E_e$ ) and the time delay between the moment of major vocal tract excitation and the instant where the glottal flow becomes quasi-constant ( $T_a$ ).

### 2.5. Calculation of glottal volume flow

Of course, no direct recordings of the glottal volume flow were made; this signal is derived from the speech waveform by means of inverse filtering. Closed Glottis Interval Covariance LPC was used to estimate the parameters of the inverse filter. In de Veth, Cranen, Strik, and Boves (1990) it was shown that this procedure outperforms more complicated ones that attempt to estimate the parameters of the inverse filter by means of Robust ARMA analysis.

Inverse filtering yields an estimate of  $dU_g/dt$ . Integration of this signal gives the flow signal. For the present article we only wanted to measure peak glottal flow  $U_0$ , excitation strength  $E_e$ , and  $P_{tr}$  for each voiced period. The value of  $E_e$  is obtained by taking the minimum of the differentiated flow in each pitch period. Likewise  $U_0$  is found by looking for the maximum of the flow signal.  $P_{tr}$  is measured at the moment of maximum glottal flow. Its value is obtained from a low-pass filtered pressure signal.

Inverse filtering was done on the fifth utterance, because that is the one that is used as a reference in the method of non-linear time-alignment. Inverse filter results were obtained for all voiced periods, including vowels and voiced consonants.

## 3. Results

### 3.1. Control of fundamental frequency

The relation between  $P_{sb}$  and  $F_0$  has been the object of quite a number of experimental (e.g. Atkinson, 1978; Collier, 1975; Maeda, 1976; Strik and Boves, 1989) and modelling studies (e.g. Ishizaka and Flanagan, 1972; Titze and Talkin, 1979). Yet, the details of this relation remain unclear. Estimates of the  $F_0$  to  $P_{sb}$  ratio from speech and special phonation tasks resulted in values between 5 and 15 Hz/cm aq. (Collier, 1975; Maeda, 1976; Strik and Boves, 1989). In another type of experiment pressure variations are induced externally. The  $F_0$  to  $P_{sb}$  ratios measured in these experiments tend towards values of 2-5 Hz/cm aq. (Baer, 1979; Strik and Boves, 1989). Strik and Boves (1989) showed that the ratio of an  $F_0$  change

resulting from a  $P_{sb}$  change alone probably is the same in both experiments, viz. 2-5 Hz/cm aq. In "normal" speech there are other factors that control  $F_0$ , especially the laryngeal muscles. Due to the simultaneous operation of these factors the ratio of total  $F_0$  change to  $P_{sb}$  change in utterances is often larger than 2-5 Hz/cm aq. The latter ratio is in agreement with the ratio of 2-3 Hz/cm aq. that was found by Ishizaka and Flanagan (1972) for their self-oscillating two-mass model.

Furthermore, it seems that in most experiments, and therefore in most presently existing models, the effects of  $P_{or}$  on  $F_0$  are not sufficiently taken into account. Probably this is due to the fact most experiments were done with sustained vowel phonation, in which the variation in  $P_{or}$  is much smaller than in normal speech. Strik and Boves (1988) studied the relation of  $F_0$  to  $P_{sb}$ ,  $P_{or}$ , and  $P_{tr}$  in connected speech. The median signals for the 29 sentence repetitions of this experiment, obtained with the method of non-linear time-alignment, are shown in Fig. 1. These signals were used to calculate correlations between the variables of interest. In Table I the correlations are given for a long voiced interval, while Table II contains the same correlations for all voiced frames.

The most important conclusion that can be drawn from the data in Tables I and II is that the pattern of correlations between  $P_{sb}$ ,  $P_{or}$ ,  $P_{tr}$ , and  $F_0$  depends very much on the part of the utterance over which the measurements are taken. If measurements are limited to a single voiced interval,  $P_{tr}$  (and  $P_{or}$ ) are much better predictors of  $F_0$  than  $P_{sb}$  (see Table I). When measured over a complete utterance, however,  $P_{sb}$  and  $P_{tr}$  explain essentially the same proportion of the variation in  $F_0$  (see Table II). This is due to the fact that the range of  $P_{sb}$  in individual voiced intervals is rather small (see Table I). The range spanned by  $P_{tr}$ , on the other hand, is much wider, because of the fact that  $P_{or}$  varies between  $P_{sb}$  in voiceless stops and zero in open vowels. In a complete declarative utterance, on the other hand, the correlation between  $P_{sb}$  and  $F_0$  is much enhanced by the fact that both show some amount of declination. The data in the Tables were obtained from a single subject and therefore should be verified on a larger population. Yet, from a physiological point of view (as well as on statistical grounds) they seem to be quite plausible.

Our results show that one must be very cautious in interpreting the outcomes of experiments on physiological control of  $F_0$  (and all other speech parameters, for that matter). Such caution is, of course, the more necessary with respect to single subject studies, like our present study. One must be especially cautious in generalizing the results of experiments to other situations than those under which they were obtained. In fact, only results that can be explained by a fairly comprehensive model may be generalized to situations where a similar model can be assumed, operating in the same regime. We are confident that the conclusion of our investigation are supported by a sufficiently complete model.

### *3.2. Control of Intensity and Voice Quality*

Even if the relation between  $P_{sb}$  and  $F_0$  has received some attention in the literature, one still must be aware that the effects of  $P_{sb}$  and laryngeal configurations are not limited to  $F_0$ ; on the contrary, factors like the acoustic power generated at the glottis and the waveshape of the glottal volume flow pulses are also affected. These relations are much less studied. That may, at least in part, be due to the assumption that voice intensity and voice timbre are of less importance from a linguistic point of view. However, if it comes to a better understanding of the fundamentals of phonation and of para-linguistic phenomena like voice quality and its variations, radiated intensity and details of the glottal volume velocity waveform become of crucial importance. In the present study we contribute some measurement data related to control of IL and voice quality obtained from connected speech and show how these data can fit in with modelling research.

#### *3.2.1. The relation between IL and Pressure*

It has long been known that there must be a relation between  $P_{sb}$  and IL, if only because  $P_{sb}$  is the major source of phonatory energy (cf. Rubin, 1963). Most measurement data on the relation between  $P_{sb}$  and IL seem to stem from in vitro experiments, however, or at best from experiments where sustained vowels were produced with intensity and pressure variations spanning a range larger than usually found in speech (e.g. Bouhuys, Mead, Proctor, and Stevens, 1968; Cavagna and Margaria, 1968; Isshiki, 1964; Tanaka and Gould, 1983).



In our own investigation of the best predictor of IL in the production of voiced speech sounds, we found that  $P_{tr}$  outperforms  $P_{sb}$  by far (Strik and Boves, 1988). The result is true both on a local (i.e. within words or voiced intervals) and on a global level (i.e. looking over complete utterances). In both situations the correlation between  $P_{tr}$  and IL exceeds 0.92, while the correlation with  $P_{sb}$  is at most 0.49 (when measured over a complete sentence, see Tables I and II). So, at least for this subject, it seems that  $P_{tr}$  is more important in the control of IL than  $P_{sb}$ .

Bouhuys et al. (1968), Cavagna and Margaria (1968), Isshiki (1964), and Tanaka and Gould (1983) all found high correlations between IL and the logarithm of  $P_{sb}$  when subjects produced sustained vowels. For sustained vowel phonation  $P_{or}$  is almost constant and close to zero and, as a result,  $P_{tr}$  is almost equal to  $P_{sb}$ . In our data  $P_{or}$ ,  $P_{tr}$ , and IL vary quickly and considerably, while  $P_{sb}$  decreases slowly during the course of the utterance (Fig. 1). This explains why in our data the relation between IL and  $P_{sb}$  is rather weak.

Except for the correlation the regression coefficient is also of importance, because it predicts the amount of change in IL due to a given change in  $P_{tr}$ . In order to be able to compare our findings with previous results we calculated the regression equation between IL and the logarithm of  $P_{tr}$ . Based on the 293 voiced frames of the median signals of the current experiment (see Fig. 1) we found the following relation

$$IL = 41.6 + 30.3 \cdot \log(P_{tr}) \quad (N = 293, R = 0.90)$$

Or, in other words, the intensity (I) of the radiated speech wave is proportional to  $P_{tr}$  to the power 3.03. Interestingly enough the value of the power in the resulting relation between I and  $P_{tr}$  is quite comparable to results reported in the literature about the relation between I and  $P_{sb}$ . For sustained vowel phonation Cavagna and Margaria (1968) found a value of  $3.0 \pm 1.0$ , Isshiki (1964) found a value of  $3.3 \pm 0.7$ , and Tanaka and Gould (1983) found a value of 3.18; while Bouhuys et al. (1968) reported a value of 3.0 for singing.

At a first glance it seems strange that comparable regression equations are found for different relations (IL and  $P_{tr}$  vs. IL and  $P_{sb}$ ), obtained for different kinds of speech (normal conversational speech vs.

sustained phonation) and different ranges of IL and pressure (2-7 cm aq. vs. 2-60 cm aq.). But closer inspection reveals that both relations are not really different. For sustained vowel phonation, and singing of constant tones,  $P_{or}$  usually is close to zero and  $P_{sb}$  and  $P_{tr}$  are almost equal. Therefore, for these modes of phonation, the relations between IL and  $P_{tr}$  and between IL and  $P_{sb}$  are very similar. The conclusion is that the relation between IL and  $P_{tr}$  obtained by Bouhuys et al. (1968), Cavagna and Margaria (1968), Isshiki (1964), and Tanaka and Gould (1983) for sustained phonation and large ranges of IL and  $P_{sb}$  is comparable to the relation obtained in this experiment for normal conversational speech.

It may still be that  $P_{sb}$  is an important factor in the control of IL, certainly if it is varied over ranges that are much wider than normally found in conversational speech but that are not unusual in singing or in very loud speech. But our data suggest that the faster variations of IL related to articulatory manoeuvres are primarily determined by variations in  $P_{or}$  that cause similar variations in  $P_{tr}$ , whereas the gradual decrease of IL observed during many (declarative) utterances in a large number of languages is caused by a gradual decrease in  $P_{sb}$ . The finding that IL is mainly controlled by  $P_{tr}$  makes it interesting to further investigate the detailed way in which IL is influenced by  $P_{tr}$  via the characteristics of the glottal volume flow.

### 3.2.2. Flow waveform characteristics and $P_{tr}$

From the literature it is known that the parameters  $E_e$  and  $U_0$  in the LF-model have most effect on IL (Gauffin and Sundberg, 1989). Thus we measured  $E_e$  and  $U_0$  for all 181 pitch periods of the fifth repetition, for which reliable inverse filter results could be obtained. Most of these periods pertained to vowels, but a substantial part comes from voiced consonants. We wanted to examine the relation of  $E_e$  and  $U_0$  to  $P_{tr}$ .

The relation between  $E_e$  and  $P_{tr}$  is shown in Fig. 2a. It seems as if this relation shows three different regimes. The bulk of the samples (148 out of a total of 181) falls into the category of, what we call, steady phonation. For the data of this category an exponential fit ( $R = 0.79$ , see Fig. 2a) is slightly better than a linear fit ( $R = 0.73$ ). The second category consists of the pulses in V-UV transitions (i.e. both V→UV and UV→V transitions). For this category  $E_e$  is often relatively lower, com-

pared to steady phonation, especially at the beginning of voicing. On the other hand, for the vowel /a/ from the very last syllable of the utterance,  $E_e$  is relatively higher (the reasons for taking the utterance final syllable apart are more fully explained in section 3.3).

The correlation between  $U_0$  and  $P_{tr}$  is depicted in Fig. 2b. Again, for the data of the category 'steady phonation', an exponential fit ( $R = 0.72$ , see Fig. 2b) is somewhat better than a linear fit ( $R = 0.65$ ). The data for the vowel /a/ of the last syllable still deviate considerably from the regression line, while the data for V-UV transitions are scattered on both sides of the regression line.

From a look at the spectra of the glottal flow waves it is immediately apparent that the spectral slopes in the three regimes are quite different. In the pitch pulses taken from vowel onsets and from the final stressed syllable the spectral tilt is much steeper than in the central pitch periods of the vowels taken from the beginning and middle of the utterance. The slope difference is more than large enough to have clear perceptual consequences. Thus, the observed effects in voice quality are of sufficient interest to take them into account in the description of speech production and to model them in high quality speech synthesis.

Although  $E_e$  and  $U_0$  appear as separate parameters in the LF-model, they are not unrelated themselves, since  $E_e$  is  $dU_g/dt$  at the moment of major excitation. Thus, if  $U_0$  increases,  $E_e$  should also increase, everything else being equal. Therefore, we looked at the relation between  $E_e$  and  $U_0$ , which is shown in Fig. 2c. For steady phonation the correlation between  $E_e$  and  $U_0$  is very high (viz. 0.80). Apparently the effect of other parameters (like  $T_0$ , skewness, and duty cycle) on this relation is not large in "normal" speech.

### 3.3. *The utterance final syllable*

Towards the end of the utterance  $F_0$ ,  $IL$ ,  $P_{tr}$ , and  $P_{sb}$  decrease substantially, while there is a marked increase in the SH activity during the last syllable (see Fig. 1). This phenomenon, the so called final fall, is observed more often (Collier, 1975; Maeda, 1976; Strik and Boves, 1989). Presumably, the larynx returns to its rest position, and the lowering of the larynx already starts before phonation has stopped (Maeda,

1976). One would expect that these gross changes in the posture of the larynx should affect the mode of vibration of the vocal folds. This observation motivated a separate study of the glottal flow pulses in the utterance final vowel.

The fact that the characteristics of the vowel in the utterance final syllable deviate from those of the preceding vowels was also observed by Klatt and Klatt (1990). For the last syllable they found increased noise in the F3 region of the spectrum, indicating a greater glottal airflow. But they also found a weaker first harmonic (relative to the amplitude of the second harmonic) in an utterance final syllable, indicative of a pressed voice with a slightly smaller open quotient. Therefore they introduced a novel breathy-laryngealized mode of vibration.

We tried to verify their hypothesis by comparing the data of the (stressed) vowel /a/ of the last syllable, with the data of the first (unstressed) vowel /a/ in the utterance. The spectrum of the utterance final vowel indeed showed increased noise at frequencies above roughly 1.4 kHz. But the amplitude of the first harmonic (relative to the amplitude of the second harmonic) was about 1.5 dB stronger, and the open quotient was approximately 50% in both vowels. Consequently, there is evidence for a breathy mode of phonation at the end of this utterance, but not for laryngealization.

Generally, everything else being equal, a decrease in  $P_{tr}$  would lead to a decrease in  $U_0$  (Ishizaka and Flanagan, 1972).  $P_{tr}$  decreases from 5.5 cm aq. for the first vowel /a/, to 4.2 cm aq. for the last vowel /a/, but the amplitude of the AC-component of glottal flow ( $U_0$ ) increases with roughly 6%. Substantial differences in the degree of adduction are not likely, since the open quotient is about 50% in both vowels. Presumably in the utterance final vowel the vocal folds are slackened, either to facilitate the maintenance of voicing with decreased  $P_{tr}$ , or due to a general relaxation of muscular activity and a preparation for breathing at the end of an utterance.

Comparing both vowels /a/ it is observed that there is a decrease in  $E_e$  of approximately 14% in the last vowel, even though  $U_0$  increases slightly. We found that, generally, the effect of other parameters (like  $T_0$ , skewness, and duty cycle) on the relation between  $E_e$  and  $U_0$  is not very

large for the data of the present experiment (see Fig. 2c). But for the last vowel /a/  $T_0$  is substantially larger than  $T_0$  of the first vowel /a/. After correction for this temporal difference, i.e. when the same number of flow pulses are plotted on the same horizontal scale for both vowels, no major differences in the shape of the glottal volume flow are observed. Consequently, the change in  $U_0$  (+6%) combined with the change in  $F_0$  (-20%) determines the change in  $E_e$  (-14%). The fact that, apart from time-stretching, no major differences were found in duty cycle and shape of the glottal pulses between both vowels /a/, also indicates that the degree of adduction has not changed substantially.

#### 4. Conclusion

In this paper we have shown that the control of  $F_0$ , IL and voice quality in normal speech may be somewhat different from what is known from the literature on studies based on sustained vowels or singing. In speech  $P_{tr}$  seems to be more important than  $P_{sb}$ , mainly because  $P_{or}$  cannot be considered as constant and negligible. Also, it was shown that the relative importance of physiological parameters that affect  $F_0$ , IL and voice quality depends very much on the nature of the speech from which they are derived. Although the results are based on a single subject study, they fit in very nicely with current models of the physiology of phonation.

Especially from the results on the control of IL and voice quality it became clear that descriptive mathematical models of the glottal flow waveform do not allow one to make the step from description to explanation. High correlations were found for  $E_e$  and  $P_{tr}$ , and for  $U_0$  and  $P_{tr}$ . But in the LF-model there is no relation between  $E_e$  and  $P_{tr}$ , or between  $U_0$  and  $P_{tr}$ , for the simple reason that  $P_{tr}$  does not figure in the model. Thus, the LF-model will never allow one to explain these relations, or why several different regimes should exist in the relation between  $P_{tr}$  and basic parameters in the model. One will have to take recourse to models that have a firm physiological basis, like the ones proposed in Titze (1984) and Cranen (1990).

### **Acknowledgements**

This research was supported by the Foundation for linguistic research, which is funded by the Netherlands Organization for Scientific Research, N.W.O. Special thanks are due to Philip Blok M.D. who inserted the EMG electrodes and the catheter in the present experiment.

## References

- Atkinson, J.E. (1978) Correlation analysis of the physiological features controlling fundamental voice frequency, *Journal of the Acoustical Society of America*, **63**, 211-222.
- Basmajian, J.V. (1967) *Muscles Alive, their functions revealed by electromyography (second edition)*. Baltimore: The Williams & Wilkins company.
- Baer, T. (1979) Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes, *Journal of the Acoustical Society of America*, **65**, 1271-1275.
- Bouhuys, A.; Mead, J.; Proctor, D.F. and Stevens, K.N. (1968) Pressure-Flow Events during Singing. In *Annals of the New York Academy of Sciences* (M. Krauss, M. Hammer, & A. Bouhuys, editors), Vol. 155, pp. 165-176.
- Boves, L. (1984) *The phonetic basis of perceptual ratings of running speech*, pp. 73-78. Dordrecht: Foris Publications.
- Cavagna, G.A. and Margaria, R. (1968) Pressure-Flow Events during Singing. In *Annals of the New York Academy of Sciences* (M. Krauss, M. Hammer & A. Bouhuys, editors), Vol. 155, pp. 152-164.
- Collier, R. (1975) Physiological correlates of intonation patterns, *Journal of the Acoustical Society of America*, **58**, 249-255.
- Cranen, B. (1990) Simultaneous modeling of EGG, PGG and Glottal Flow. To appear in *Proceedings of the sixth Vocal Fold Physiology Conference, Stockholm*.
- Cranen, B. and Boves, L. (1985) Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production, *Journal of the Acoustical Society of America*, **77**, 1543-1551.
- Cranen, B. and Boves, L. (1987) The acoustic impedance of the glottis: modeling and measurements. In *Laryngeal function in phonation and respiration* (Th. Baer, C. Sasaki and K. Harris, editors), pp. 203-218. Boston: College-Hill.
- Fant, G. (1986) Glottal flow: Models and interaction, *Journal of Phonetics*, **14**, 393-400.

- Fourcin, A.J. (1974) Laryngographic examination of vocal fold vibration. In *Ventilatory and phonatory control systems* (B. Wyke, editor), pp. 315-326. London: Oxford University Press.
- Gauffin, J. and Sunberg, J. (1989) Spectral correlates of glottal voice source waveform characteristics, *Journal of Speech and Hearing Research*, **32**, 556-565.
- Hirose, H. (1971) Electromyography of the Articulatory Muscles: Current Instrumentation and Techniques, *Haskins Laboratory Status Report on Speech Research*, **SR-25/26**, 73-86.
- Isshiki, N. (1964) Regulatory mechanisms of voice intensity variation, *Journal of Speech and Hearing Research*, **7**, 17-29.
- Ishizaka, K. and Flanagan, J.L. (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords, *Bell Systems Technical Journal*, **51**, 1233-1268.
- Klatt, D.H. and Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, *Journal of the Acoustical Society of America*, **87**, 820-857.
- Maeda, S. (1976) *A characterization of American English intonation*, Ph.D. thesis, MIT, Cambridge.
- Rubin, H.J. (1963) Experimental studies on vocal pitch and intensity in phonation, *The Laryngoscope*, **8**, 973-1015.
- Strik, H. and Boves, L. (1988) Data processing of physiological signals related to speech. In *Proceedings of the Dept. of Language and Speech, Phonetics Section, Nijmegen University*, pp. 41-56.
- Strik, H. and Boves, L. (1989) The fundamental frequency - subglottal pressure ratio. In *Proceedings of EUROSPEECH-89*, Vol. 2, 425-428.
- Strik, H. and Boves, L. (1991) A DP algorithm for time-aligning physiological signals related to speech. This issue.
- Tanaka, S. and Gould, W.J. (1983) Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects, *Journal of the Acoustical Society of America*, **73**, 1316-1321.
- Titze, I.R. (1984) Parameterization of glottal area, glottal flow, and vocal fold contact area, *Journal of the Acoustical Society of America*, **75**, 570-580.



Titze, I.R. and Talkin, D.T. (1979) A theoretical study of various laryngeal configurations on the acoustics of phonation, *Journal of the Acoustical Society of America*, **66**, 60-74.

Veth, J. de; Cranen, B.; Strik, H.; and Boves, L. (1990) Extraction of control parameters for the voice source in a text-to-speech system. In *Proceedings of ICASSP-90*, paper 21.S6a.2

### Figure captions

Figure 1. Median physiological signals, obtained by the method of non-linear time-alignment. Plotted are, from top to bottom,  $F_0$ , IL,  $P_{tr}$ ,  $P_{or}$ ,  $P_{sb}$ ,  $V_1$ , SH, and VOC.

Figure 2. Scatterplots of respectively (a)  $E_e$  and  $P_{tr}$ ; (b)  $U_0$  and  $P_{tr}$ ; and (c)  $E_e$  and  $U_0$ . Given are regression lines for exponential or linear fits, and the correlation coefficients for the fits for the data of the category 'steady phonation'.  $E_e$  and  $U_0$  values are given relative to the maximum observed value for each quantity.

### Tables

Table I. Correlation matrix, means and standard deviations of the median physiological signals for a voiced interval (N=66,  $|R|>0.315$  for  $p<0.01$ ).

	F <sub>0</sub>	IL	P <sub>tr</sub>	P <sub>or</sub>	P <sub>sb</sub>	mean	SD
F <sub>0</sub>	1.000	0.808	0.851	-0.783	0.478	118.58	3.70
IL		1.000	0.960	-0.983	0.111	63.23	3.38
P <sub>tr</sub>			1.000	-0.968	0.274	5.42	0.88
P <sub>or</sub>				1.000	-0.054	1.16	0.91
P <sub>sb</sub>					1.000	6.35	0.16

Table II. Correlation matrix, means and standard deviations of the median physiological signals for all voiced frames (N=293,  $|R|>0.151$  for  $p<0.01$ ).

	F <sub>0</sub>	IL	P <sub>tr</sub>	P <sub>or</sub>	P <sub>sb</sub>	mean	SD
F <sub>0</sub>	1.000	0.667	0.729	-0.153	0.772	115.87	8.59
IL		1.000	0.923	-0.663	0.492	62.20	4.25
P <sub>tr</sub>			1.000	-0.638	0.612	4.95	1.17
P <sub>or</sub>				1.000	0.211	0.89	0.95
P <sub>sb</sub>					1.000	5.65	0.90