Proceedings of The Institute of Acoustics

AERODYNAMIC ASPECTS OF VOICE PRODUCTION

Bert Cranen and Louis Boves

Institute of Phonetics, Nijmegen University, The Netherlands

INTRODUCTION

Detailed measurement data on pressure/flow relations during speech production are notoriously scarce. Over the years a number of techniques have been developed for measuring subglottal pressure (henceforward Psb). Among these are the indirect measurement of Psb by means of a gass-filled balloon inserted into the oesophagus, direct measurement via a tracheal puncture and direct measurement via a thin hollow catheter inserted into the trachea via the glottis. It has been shown that the frequency range of these measurements is mostly limited to approximately 300 Hz, due to the low-pass effect of the tubing [1]. Recently ultra-miniature semiconductor pressure sensors have become available, which can be mounted in a very thin catheter, allowing them to be introduced into the trachea via the glottis [2,3,4]. These transducers are specified to have a flat frequency response up to at least 10 kHz. The high resolution has, however, to be paid for with a large temperature sensitivity. This makes an absolute calibration of the semiconductor pressure sensors quite cumbersome. Also it is impossible to prevent the sensors from inadvertent tapping at the wall of the trachea due to slight movements of the subject, especially if a sensor is located at the very tip of the catheter. Contact of a sensor with the wall will cause artifacts that cannot be guaranteed to be obvious on visual inspection of a pressure recording on its own.

In this paper we will describe a procedure that allows us to obtain broadband Psb recordings which are both accurately calibrated and free of artifacts. The procedure is based on the simultaneous registration of the pressure at two equidistant points in the trachea and in the pharynx. This brings us in a position where we cannot only calculate transglottal pressure, but also pressure gradients in the trachea and in the pharynx. We will argue that it is possible to obtain accurate estimates of the time course of the air flow in the trachea and in the pharynx by fairly straightforward processing of the pressure gradient signals.

SIMULTANEOUS MEASUREMENTS OF GLOTTAL ACTION

General experimental setup

In our experiment the following signals were simultaneously recorded on an instrumentation recorder (Philips ANALOG 14)

(1-2) pressure at two locations in the trachea (5 cm apart)
(3-4) pressure at two locations in the pharynx (5 cm apart)
(5) speech signal about 10 cm in front of the mouth
(6) photoglottogram (PGG) (Frøkjaer-Jensen LG600 with custom photo-transistor)
(7) electroglottogram (EGG) (Fourcin-Abberton laryngograph)

The experiment was carried out with two male subjects, who had to perform two tasks, the first of which is part of the calibration procedure to be described below. The second task consisted of reading lists of two-syllable nonsense utterances at a comfortable pitch and intensity level. These lists contained

AERODYNAMIC ASPECTS OF VOICE PRODUCTION

Dutch pseudowords of the form ´VCV, VC´V, C´VCV and CV´CV where V denotes a vowel of the set /a,o,u,y,i,e,ə/ (with ´V in stressed and V in unstressed position) and where C denotes a consonant from the set /p,b; t,d; s,z; f,v; k,ɣ; h/. The reading of the word lists took about 15 minutes.
After each recording session, part of the signals were A/D converted (effective sample frequency = 10 kHz/channel; amplitude resolution = 12 bits) and stored on a disk of a digital minicomputer (Data General Eclipse/S200).

Pressure measurements
The four pressure signals were obtained by means of a Millar catheter with four miniature pressure transducers (type PC-784[K]) which was inserted via the nasal passage through the glottis into the trachea. The transducers are situated at equidistant intervals of 5 cm, and the first transducer is located at the tip of the catheter [5]. After the catheter has found its proper position in the posterior commissure it hardly influences phonation.
Each pressure sensor forms a Wheatstone bridge with two semiconductor strain gauges. The reference pressure is atmospheric pressure, obtained by a vent tube which connects the back of the sensor to the open air. The sensitivity of the sensors varies between 1.2 and 2.5 mV/V/mm Hg. With a bridge excitation voltage of 5 V(DC) this is equivalent to 4.4 - 9.2 µV/mm aq. The long-term drift of the transducers is specified to be less than 6 mm Hg (≈82 mm aq.) in 12 hours.
The noise levels which were measured at the outputs of the home built 4-channel pressure amplifier were less than .95 mV(eff). The amplifiers all have a gain of 668, so that the equivalent transducer/amplifier noise at the input is on the order of 1.42 µV. Thus, if we define S/N-ratios of 6 dB as reliable measurements, we may conclude that with this equipment pressures can be measured with a resolution of 0.65 mm aq.
The temperature error band of the sensors is specified as a worst case shift of 2 mm Hg (or 2.8 cm aq.) when the temperature is raised from 28 to 38 °C. This corresponds to an error of more than 10% of full-scale if the transducers are employed in speech work. The temperature error is the more troublesome because the sensor outputs behave completely unpredictably after they have been exposed to a sudden temperature change.
Due to different offsets and sensitivities of the transducers, it is mandatory that they all be calibrated simultaneously. A rescaling procedure which accounts for different sensitivities should be an integral part of the calibration procedure if the outputs of two or more sensors are to be compared.

CALIBRATION OF THE PRESSURE SENSORS

Procedure
In the calibration procedure three phases can be discerned:
First, the subject blows into a U-tube manometer, exerting increasing pressures which are maintained for a couple of seconds. The experimenter reads out the manometer deflection and his comments are recorded on the audio channel of the FM recorder. Second, the signals (after A/D conversion) are rescaled in order to compensate for the differences in transducer sensitivity, amplifier gains and/or offsets, etc. Third, the numerical values of the rescaled signals must be mapped onto the actual pressures as they were called out by the

experimenter in order to allow absolute interpretation.
The second step of this calibration procedure is the most critical one. As long as all signals to be scaled represent the static pressure in one and the same cavity, linear regression analysis can be used to obtain correction terms for the offsets and gains. This implies that one has to be certain that the glottis is open and that no artifacts are present in the signals. This can only be done by visual inspection of a signal, preferably in combination with another signal which should behave in a similar way. Use of signal segments where one of the sensors has been mechanically excited or is still recovering from a large temperature change may disturb the linear relation between the transducer outputs and thus cause gross errors.

Results
Attempts to calibrate recordings made in four independent measurement sessions show that the error observed in the gains generally does not exceed 5%. As to the offsets, two effects can be discerned viz. a passive long-term drift and a temperature drift related to the subjects' breathing. In two recording sessions the long-term drift was nearly zero for all transducers, while in two others drifts up to 3 cm aq. were observed over a time interval of 45 minutes. In order to check how large the drift in the zero level due to temperature effects was during the sessions, we took a closer look at the DC components of each pair of transducers in the same cavity. We defined a situation as 'temperature stable' when two transducers in the same cavity did not differ by more than 1 cm aq. After an inhalation differences of 4 - 6 cm aq. were sometimes observed, which most of the time took 1 - 4 seconds to return to the stable condition. Both drift effects illustrate the necessity to establish the zero pressure level more often during a recording.

All effects that hinder calibration of the pressure sensors are, of course, also present in actual recordings during speech production. In practice, a measurement accuracy of 5% can be obtained in the absolutely scaled pressure recordings, at least as far as their AC-components is concerned. The short-term drift of the zero level results in an additional error. It can be kept below 1 cm aq., however, if the subject applies only superficial breathing. If only high pass filtered versions of the pressure signals are used, as will be done in the section on flow calculation, only the 5% error remains.

RESULTS

Broadband pressure registrations
In the left column of Fig. 1 a microscopic view of a portion of a vowel /a/ is shown for speaker LB. Moments of glottal closure (marked with an asterisk) and opening (marked with circles) which are automatically detected by means of peak picking in the differentiated EGG waveform are indicated. The waveforms which are shown are simultaneous recordings of EGG, PGG, sub-, supra-, and transglottal pressure. Transglottal pressure (henceforward Ptr) has been calculated by subtracting supraglottal from subglottal pressure after suitable rescaling. It can be shown that the absolute error in the DC-component is about the same as for the pressure signals of individual sensors, i.e. 1 cm aq.

Moreover, because the absolute level of Ptr is of the same order of magnitude as the subglottal pressure the relative error in the AC-component is 5%.
From Fig. 1 it becomes clear that Psb is by no means constant. AC peak-to-peak variations occur which are of the same order of magnitude as the DC level . The variations in Ptr may be even twice as large. Furthermore, it can be observed that glottal closure nearly always precedes the maximum in the subglottal pressure (which coincides with the minimum in the supraglottal pressure). This is in contradiction with the observations of Koike and Hirano [6], who observed that the peaks in Psb coincided with the moment of glottal opening. Note that a pressure peak at glottal closure does not fit a model which is based on stationary flow and which is reigned by the Bernoulli equation. Pressure peaks at glottal closure can only be explained by assuming acoustical excitation of the subglottal cavities.

Pressure gradient and flow waveforms

In this section we will first show that the use of two pressure transducers in the trachea and in the pharynx enables us to estimate the volume flow waveforms in both cavities. In order to estimate flow it is assumed that uni-directional plane wave propagation takes place in the cavities both below and above the glottis. Then, with the viscosity of the fluid neglected, the linearized equation of momentum becomes:

$$\frac{\partial \vec{u}}{\partial t} = -\rho_o \frac{\partial p}{\partial x} \tag{1}$$

Here $\vec{u}$ denotes the particle velocity vector, p the pressure, and ρ_o the mass density of the fluid at rest. The X-axis is taken in the direction of the flow. Because we have recorded subglottal and supraglottal pressure at two locations in the trachea and pharynx respectively, we can obtain an estimate of the pressure gradient ($\partial p/\partial x$) as follows:

$$\left.\frac{\partial p}{\partial x}\right|_{X_o} \simeq \frac{p(X_o + dX/2) - p(X_o - dX/2)}{dX} \tag{2}$$

with Xo a fixed location ca. 5 cm under or above the glottis and dX = 5 cm.
For frequencies below 1500 Hz, dX is less than a quarter wavelength and generally equation (2) will not be too bad an estimate of the pressure gradient. In other words, integrating the difference signal of two pressure transducers in the same cavity with respect to time will yield a signal which is proportional to the particle velocity. Provided that the cross-sectional area of the tubes in the neighborhood of Xo does not vary too abruptly as a function of x, particle velocity can be considered as a scaled version of volume velocity (U).
Some results of this flow estimation procedure are shown in the right column of Fig. 1 for the same vowel portions for which the pressure recordings are shown in the left column. The shown waveforms represent bandpass filtered pressure gradient in trachea ($\partial p/\partial x|sb$) and pharynx ($\partial p/\partial x|sp$) (the difference signals were filtered by means of a digital linear phase filter with cut-off frequencies at 40 and 1500 Hz), plus the corresponding particle velocities ($u|sb$ and $u|sp$). The particle velocity waveforms have been set to an arbitrary zero-level which was chosen so that the moments of glottal closure correspond to zero flow.
The flow waveforms obtained by integrating the pressure gradient in the trachea

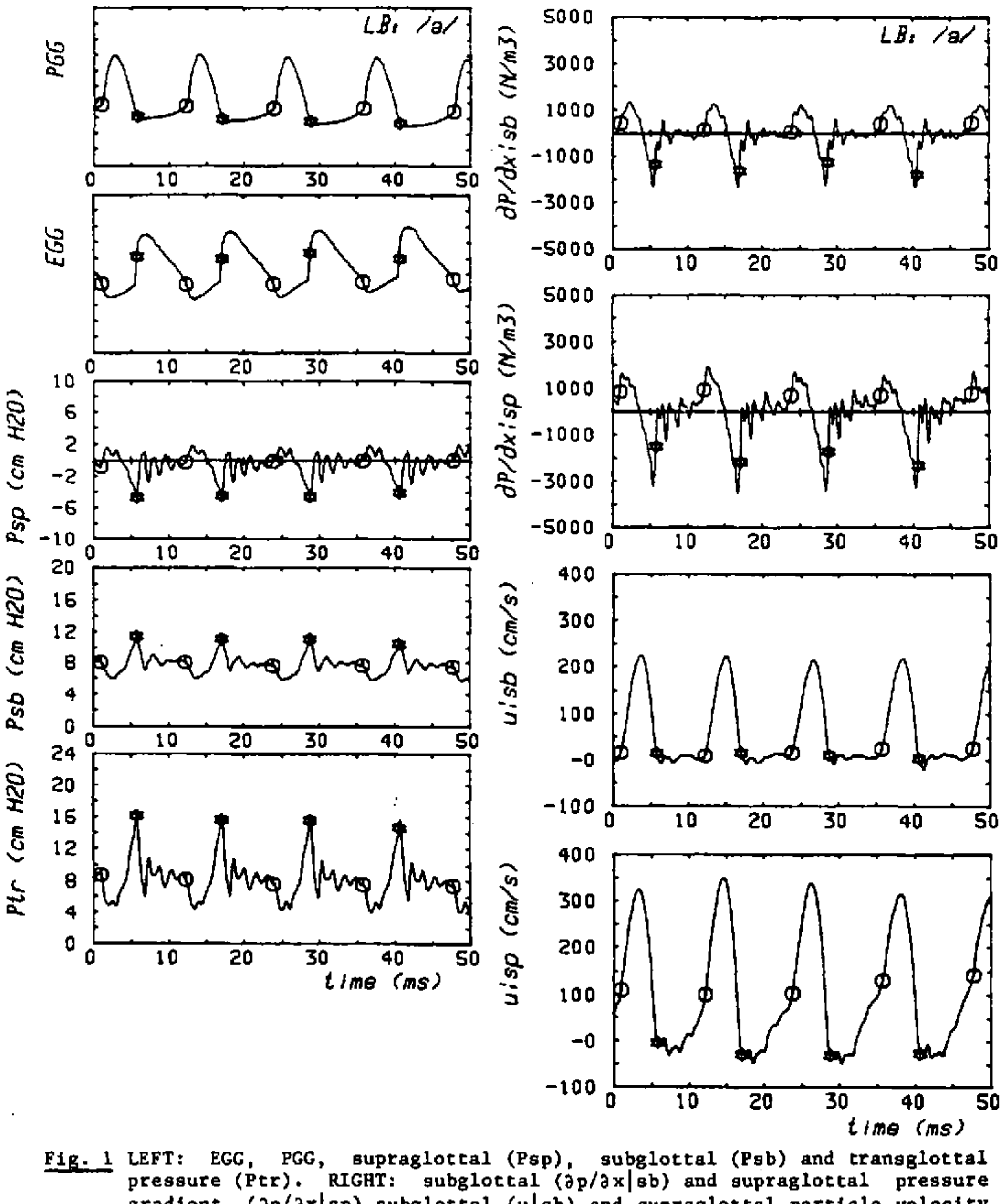

Fig. 1 LEFT: EGG, PGG, supraglottal (Psp), subglottal (Psb) and transglottal pressure (Ptr). RIGHT: subglottal ($\partial p/\partial x|sb$) and supraglottal pressure gradient ($\partial p/\partial x|sp$) subglottal ($u|sb$) and supraglottal particle velocity ($u|sp$). Vowel: /a/; Speaker: LB.

look very convincing. Assuming a cross-sectional area of about 3 cm2 the peak volume flow in the trachea amounts to 600 cc/s, a value which compares well with data in the literature [7,8]). The peak-to-peak amplitude of the pressure gradient signal in the trachea is 4 cm aq. Given a measurement error of 5% in the individual pressure signals, the measurement error of a 4 cm aq. peak-to-peak signal will not exceed 25%, a value which is deemed acceptable.
The results of the flow calculations in the pharynx, on the other hand, look less convincing. Presently we can think of three factors which may have contributed to this effect. First, the assumption of pure plane wave propagation may not hold in the lower part of the pharynx due to a jet expansion right above the glottis. Second, the ripples in the reconstructed pharyngeal flow during the closed glottis interval may be due to interactions with formant oscillations or to vertical movements of the folds after closure. Last, the results of the calculations may suffer from a relatively large measurement error. The peak-to-peak amplitude of the pressure gradient signal in the pharynx varies from 6 cm aq. for open vowels to a low 1 cm aq. for the closed vowels /i,u/, which is close to the noise level in the difference signals, which will typically assume values of .5 to 1 cm aq.
Research is under way that attempts to answer the questions formulated above. Here we will confine ourselves to the conclusion that the measurements secured by means of the set-up described in this paper appear to enable and to stimulate both experimental and theoretical research into the physiology of phonation and voice source - vocal tract interaction. Theoretical studies appear to profit from the availability of measurement data that can serve to guide thinking and that provide testing grounds for assessing the predictive power of models.

Slow variations in transglottal pressure

In Fig. 2 Ptr signals are shown for a reading of ´VCV nonsense word lists by subject LB. The signals are obtained by subtracting rescaled sub- and supraglottal pressure signals and then low pass filtering the result by means of a 251 point linear phase filter with cut-off frequency at 30 Hz.
The first observation that can be made from the figure is that even in a situation in which the subject is asked to breath superficially there is hardly any indication of a gradual decline of Ptr as the lung volume diminishes towards the end of a breath period. Most likely this is due to the fact that the subjects know that stressed syllables keep coming until the very end of the utterance.
A second observation is that the unstressed syllables tend to show a Ptr that is lower than in the stressed syllables. There is also a strong tendency for Ptr to drop during the unstressed syllables, but this is believed to be an ´end of utterance´ effect, because the subjects were asked to produce each word as an independent unit.
Finally, it can be seen that Ptr drops to zero very rapidly in the voiceless consonants (/p,t,s,f,k/), and remains at a level of at least a couple of cm aq. during the voiced consonants (/b,d,z,v/). The largest pressure drops for voiced consonants occurs in the realizations of the phoneme /v/, where Ptr reaches values as low as 1 cm aq. This observation is explained by the fact that for this subject (as for most speakers of standard Dutch) the voiced character of the phoneme /v/ is very weak, with the natural result that /VvV/ and /VfV/ pairs are very difficult to distinguish.

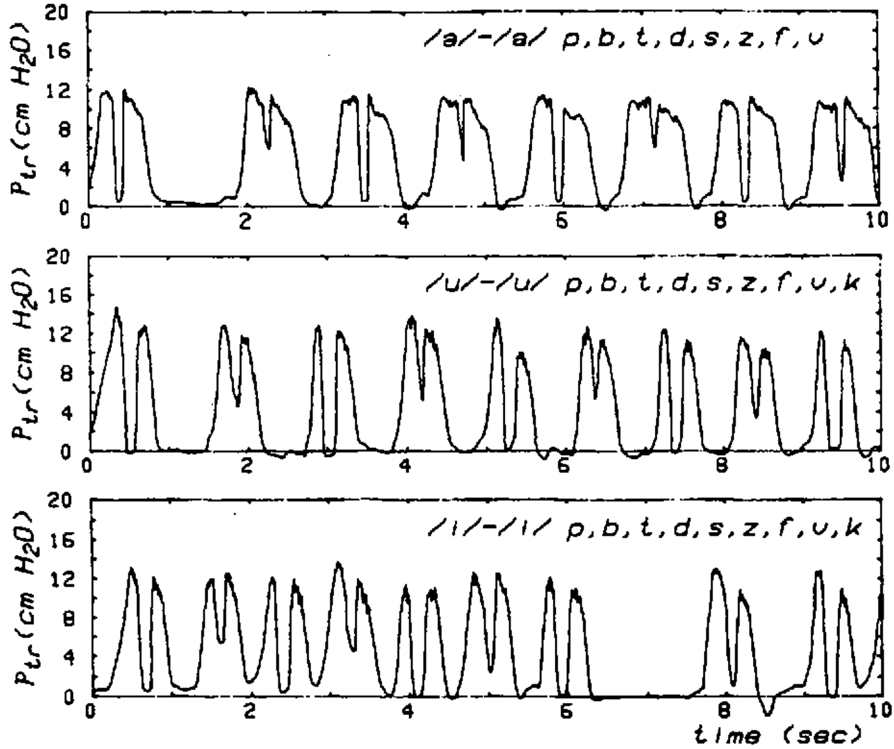

Fig. 2 Low pass filtered (cut-off frequency 30 Hz) and down sampled transglottal pressure signals (obtained by subtracting supra-glottal pressure from subglottal pressure). Nonsense word lists of the form 'VCV

CONCLUSIONS

From our work it appeared that temperature instabilities are a major obstacle hindering the use of semiconductor strain gauge pressure transducers in speech research. We have, however, described a simple and reliable technique for an 'in vivo' calibration of the transducers.

By comparing the outputs of two such transducers in the same cavity it was found that the cooling caused by a fairly deep inhalation suffices to make the transducer zero level drift by several centimeters of water. Also, the transducers prove to be extremely susceptible to chance mechanical excitation. Most artifacts can, however, be detected by comparing the outputs of two

transducers in the same cavity or by relating the output of a single transducer to simultaneously recorded laryngeal activity or to the speech wave.
Thanks to the high frequency resolution of the pressure transducers the dynamic behavior of the air mass around the vocal folds can be studied quantitatively. We have shown that pressure difference signals within one cavity are a useful basis for the calculation of particle velocity, which in its turn can be converted to volume flow. Simple linear filtering of the high resolution pressure signals yields very accurate recordings of the slow variations in Ptr associated with the production of stressed syllables and voiced or voiceless consonants.
At the moment the extremely large database collected during four recording sessions serves as the raw material for empirical and theoretical studies of various aspects of phonation, including the voiced/voiceless distinction and voice source - vocal tract interactions.

ACKNOWLEDGEMENT

This research was supported by the Foundation for Linguistic Research which is funded by the Netherlands Orginization for the Advancement of Pure Research, ZWO. Drs. Ph. Blok of the E.N.T. clinic of Nijmegen University Hospital assisted in making the recordings.

REFERENCES

[1] T.D. Edmonds, D.J. Lilly & J.C. Hardy, 'Dynamic characteristics of air-pressure measuring systems used in speech research', J. Acoust. Soc. Am., Vol. 50, 1051-1057, (1971).
[2] Y. Koike, 'Sub- and supraglottal pressure variation during phonation', in: K.N. Stevens & M. Hirano (Eds.) 'Vocal Fold Physiology', Tokyo: Univ. of Tokyo Press, 1981, 181-192.
[3] P. Kitzing & A. Lofqvist, 'Subglottal and oral pressure during phonation - Preliminary investigation using a miniature transducer system', Med. Biol. Eng., Vol. 13, 644-648, (1975).
[4] L. Boves, 'The phonetic basis of perceptual ratings of running speech', Dordrecht: Foris Publ., 1984.
[5] B. Cranen & L. Boves, 'Absolute pressure measurements during speech production using semiconductor pressure transducers: Impacts on models of speech production', to appear in J. Acoust. Soc. Am.
[6] Y. Koike & M. Hirano, 'Glottal area time function and subglottal pressure variation', J. Acoust. Soc. Am., Vol. 54, 1618-1627, (1973).
[7] J.W. van den Berg, J.T. Zantema & P. Doornenbal, 'On the air resistance and the Bernoulli effect of the human larynx', J. Acoust. Soc. Am., Vol. 29, 626-631, (1957).
[8] J.L. Flanagan, 'Speech analysis, synthesis, and perception', Berlin: Springer Verlag, 1972.