Forschungsberichte des
Instituts für Phonetik und
Sprachliche Kommunikation
der Universität München
(FIPKM) 35 (1997) 101-106

TECHNIQUES FOR INVESTIGATING LARYNGEAL ARTICULATION AND THE VOICE-SOURCE

Section 1: Investigation of the devoicing gesture[1]

Philip Hoole

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 Munich, Germany
hoole@phonetik.uni-muenchen.de

ABSTRACT

This paper gives an overview of techniques for the investigation of the laryngeal devoicing gesture, discussing methodological issues relevant to the ensuring reliability and interpretability of the data. The emphasis is placed on fiberoptic filming and transillumination, as being the techniques most widely available in phonetics laboratories and probably most suitable for routine articulatory investigations. Brief consideration is also extended to laryngeal EMG and pulse-echo ultrasound.

INTRODUCTION

In this sub-chapter we will deal with techniques which permit investigation of two rather different aspects of laryngeal behaviour. The first, the articulatory aspect, is dealt with in Section 1, and specifically concerns those laryngeal adjustments of abduction and adduction involved in the production of voiceless segments, and which can conveniently be referred to by the term "devoicing gesture". The first section will accordingly look mainly at techniques for assessing the kinematics of this gesture.

The second aspect concerns the detailed acoustic analysis of phonation itself. A number of studies in recent years have highlighted the fact that even during what would be termed modal phonation there is considerable modulation of the voice source as a function of the prosodic and segmental content of utterances. Of particular interest here are exploratory studies which indicate that the phonatory quality of a voiced segment may be affected by laryngeal adjustments associated with an adjacent voiceless consonant, and which suggest that there may be striking cross-language differences. These phonatory variations may yield valuable insights into the nature and timing of laryngeal gestures associated with these stops. In Section 2 of this chapter, we give a brief outline of the techniques that may be used to analyse the voice source, concentrating especially on those used to obtain the data presented in the section on "Voice source variation in the vowel as a function of consonantal context" in the chapter entitled "Laryngeal Coarticulation".

---

[1]This work is to appear as a contribution to "*Coarticulation: Theoretical and Empirical Perspectives*", W.H Hardcastle & N. Hewlett (eds.), CUP. We here only reproduce Section 1, "*Investigation of the devoicing gesture*". Section 2, entitled "*Techniques for analyzing the voice source*", was contributed by Ailbhe Nì Chasaide and Christer Gobl, Dublin. However, the Introduction pertains to both sections of the chapter. This methodologically oriented chapter should be seen as a companion to the chapter on "*Laryngeal Coarticulation*" also reproduced in this volume of FIPKM.

SECTION 1

Investigation of the devoicing gesture

It can be assumed that the immediate aim of any investigation of devoicing is to obtain measurements of the amplitude, form and timing of the abductory-adductory cycle, either directly in terms of the time-course of the separation of the vocal processes of the arytenoid cartilages, or indirectly through the resulting transillumination signal, or through the underlying electromyographic activity. As discussed further in the section on "Coarticulatory investigations of the devoicing gesture" in the chapter entitled "Laryngeal Coarticulation" the wider aim of such analyses is to gain further insight into, firstly, the nature of laryngeal-oral coordination in different categories of consonants and secondly into the blending processes occurring in clusters of voiceless sounds. We will discuss how the data for such analyses can acquired most effectively using a combination of fiberoptics and transillumination, and will also look more briefly at EMG and pulse-echo ultrasound. The emphasis on the first technique is motivated by the fact that it is probably the one most accessible to laboratories involved in coarticulation research. The reader is referred to Löfqvist (1990) for very convenient illustrations of some typical voiceless sequences since transillumination and EMG signals are shown in parallel.

Fiberoptic endoscopy, transillumination, and their combination

In 1968 Sawashima and Hirose presented an endofiberscope that for the first time made it possible to routinely investigate laryngeal activity in running speech. For a description of fiberscope construction see Sawashima and Ushijima (1971). For a representative early study examining the voiced and voiceless consonants of English see Sawashima, Abramson, Cooper & Lisker (1970).

One of the main problems in analyzing laryngeal films (ceteris paribus, this also applies to transillumination) is caused by the fact that the distance between objective lens (i.e distal end of the fiberscope) and the glottis is not constant and not known. Differences in this distance can be caused, of course, by vertical movements of the larynx but also by the influence of velar movement on endoscope position. These influences must be minimized by careful choice of the speech material since methods that have been proposed to control for the varying distance between endoscope and glottis, such as radiographic monitoring (Kiritani, 1971) or stereoscopic procedures (Sawashima & Miyazaki, 1974, Fujimura, Baer & Niimi, 1979, Yoshioka, 1984), are unlikely to prove suitable for routine use, at least in coarticulatory studies. Thus most investigations must currently content themselves with relative rather than absolute measurements of glottal opening. Assuming suitable phonetic material is employed the stability actually achieved in a particular recording must essentially be judged by the experienced investigator.

One reason why a combination of fiberscopy and transillumination is attractive is that at standard video or cine frame rates the temporal resolution of an endoscopic film is rather low, particularly if it is desired to locate the offset and onset of voicing in the devoicing gesture (cf. Hirose and Niimi, 1987).

One of the main attempts to circumvent the above limitation involves the use of specialized digital video equipment (Honda, Kiritani, Imagawa & Hirose, 1987; Kiritani, Imagawa & Hirose, 1992). CCD cameras allow trade-offs between spatial and temporal resolution, so an adequate frame-rate is certainly achievable. Use of image-processing to extract parameters such as glottal area or distance between the vocal processes should be feasible. However, in view of the fact that such algorithms tend to be computationally expensive and that some compromises with regard to spatial resolution have to be made it seems that at the present time this represents essentially a rather roundabout way of arriving at something that is not very different from a straightforward transillumination signal, at least as far as the gross characteristics of the devoicing gesture are concerned. Nonetheless, for the investigation of the way in which patterns of vocal fold vibration are modified at transitions between vowels and consonants the technique already appears to offer much promise (it should also be ideally suited to investigating irregular or asymmetric phonatory phenomena such as creak or diplophonia). With further technological advances more routine application of this technique should become possible.

The transillumination technique initially developed independently of fiberoptic endoscopy. It essentially involves a light source and a phototransducer located on opposite sides of the glottis; the amount of light passing through the glottis, and accordingly the output voltage of the phototransducer amplifier, is modulated by the changes in the size of the glottal aperture occurring duing speech and respiraton. Over the years various different arrangements of these two basic components have been tried out (see e.g Sonesson, 1960; Malécot & Peebles, 1965; Lisker, Abramson, Cooper & Schvey, 1969; Ohala, 1966; Frøkjær-Jensen (1967), mainly regarding whether the light source is applied externally to the neck below the glottis, and with the phototransducer in the pharynx, or vice-versa. See Hutters (1976) for a valuable overview of methodological issues[2].

Once fiberoptic endoscopy started to become widespread in phonetic research, it was natural to employ an arrangement with the fiberscope functioning as light source in the pharynx and with the transducer applied externally to the neck[3]

[2]One aspect of Hutter's work that deserves to be more widely followed is a procedure to determine the relationship between change in output voltage of the PGG amplifier for defined changes in light-intensity.

[3]In contrast to custom-designed equipment such as the Frøkjær-Jensen photoelectroglottograph, the stability of standard endoscopic light-sources should not be taken for granted. Appreciable ripple at the first few harmonics of the mains frequency can be troublesome. Use of a DC power source may be advisable.

(Löfqvist & Yoshioka, 1980). The great advantage of this approach is, of course, that the endoscopic view allows the stability of the positioning of the light-source in the pharynx to be monitored, at least qualitatively.

There are essentially two positions in which the transducer can be applied to the neck, either between the cricoid and thyroid cartilage, or below the cricoid cartilage. Following Frøkjær-Jensen (1971) we can assume that devoicing mainly involves modulation of the width of the posterior glottis, while in phonation the main modulation of the width occurs at more anterior locations. It appears (and is anatomically plausible) that the lower transducer position weights the posterior devoicing activity more strongly, while the upper position weights phonatory activity more strongly[4]. Accordingly, the sub-cricoid position is preferable for studies of devoicing; in particular, a more stable (albeit sometimes rather weak) signal is obtained since in the upper position the signal can be strongly influenced by changes in laryngeal height and orientation, related, for example, to the intonation pattern of the utterance. Thus, Löfqvist & Yoshioka found a good correlation between the amplitude of the transillumination signal and the glottal width as measured from a fiberoptic cine film, but only when the lower transducer postion was used[5].

In our own implementation (Hoole, Schröter-Morasch & Ziegler), we became convinced that it is essential to record the whole experimental session on videotape (cf. Hirose, 1986) since if a permanent record of the endoscopic view is not available it is very easy for the human observer to overlook brief retractions of the tongue-root or epiglottis that have a massive influence on the transillumination signal. Explicit synchronization of the video and transillumination signals is also very important[6]. In order to further facilitate the recognition of such artefacts, it seemed to us that one obvious approach would be to record from **two** phototransistors simultaneously, one between cricoid and thyroid, and one below the cricoid cartilage. Based on the above discussion of transducer positioning one can expect that the reduction in light intensity caused by shadowing will in general either be greater at the upper transducer position or will at least start earlier there. In other words, when the two signals diverge, a departure from ideal recording conditions may have occurred[7].

In figure 1 typical traces obtained with the twin transducer system are displayed for the German sentence "Lies 'die Schiffe' bitte" ("Read 'the ships' please"). The sentence contains the voiceless sounds /s/, /ʃ/, /f/ and /t/. Each of these sounds is associated with a peak in the amplitude of the transillumination signals. Main stress in this sentence fell on the first syllable of "Schiffe", and it will be observed that much the largest amplitude of the devoicing gesture in this utterance is associated with the fricative /ʃ/ at the onset of this syllable, peak glottal opening occurring at about the temporal midpoint of the fricative. Vocal fold vibration is visible in the transillumination signal as a high-frequency modulation overlaid on the gross abductory and adductory movements. As is to be expected from the above discussion, this phonatory activity is a more salient feature of the signal labelled PGG1, which corresponds to the transducer position between thyroid and cricoid cartilages. A consistent feature of voiceless fricatives that can be easily followed here (especially for the stressed /ʃ/) is the hysteresis effect by which vocal fold vibration dies out rather gradually during glottal abduction for the fricative, but does not recommence for the following vowel until the glottis is almost completely adducted again.

Normally, as here, the relative amplitude of glottal abduction for the different voiceless sounds in the sentence is the same at both transducer positions, the two traces proceeding essentially in parallel. If, however, this is not the case, (for example, if abduction for /ʃ/ were to appear greater than for /f/ in the PGG 2 trace but smaller in the PGG 1 trace) then this is an indication that recourse should be had to the video film to check for shadowing of the glottis etc.

---

[4]cf. Baer, Löfqvist & McGarr (1983) for this conception of the transillumination signal as the weighted sum of the glottal width at a series of points along its length.

[5]It should not be overlooked, however, that the correlation is not based on two completely independent measurement methods; it will be probably be the case, for example, that changes in the distance between larynx and fiberscope will have a similar effect on the analysis of both the cine film and the transillumination signal (e.g smaller image, lower signal amplitude).

[6]We have found the following solution very convenient: a commercially available video-timer (FOR-A VTG33) was modified according to a technique developed by N.R. Petersen at the phonetics lab of Copenhagen University. In addition to its normal function, which is to insert date and time into the video signal with a resolution of 1 cs, pulses for the one-second intervals were led out in a form which allowed them to be recorded on one track of an instrumentation recorder that was also used to record the photoglottographic and audio signals. Alternatively, if the measurement signals are recorded online by computer, a straightforward solution would be to remote-control the video-timer using a digital i/o port, with the computer resetting and starting the timer at the start of each measurement sequence.

[7]This double transduction may also increase the utility of transillumination for studies of phonation, since vibratory parameters may differ along the front-back dimension of the glottis (Baer et al., 1983).

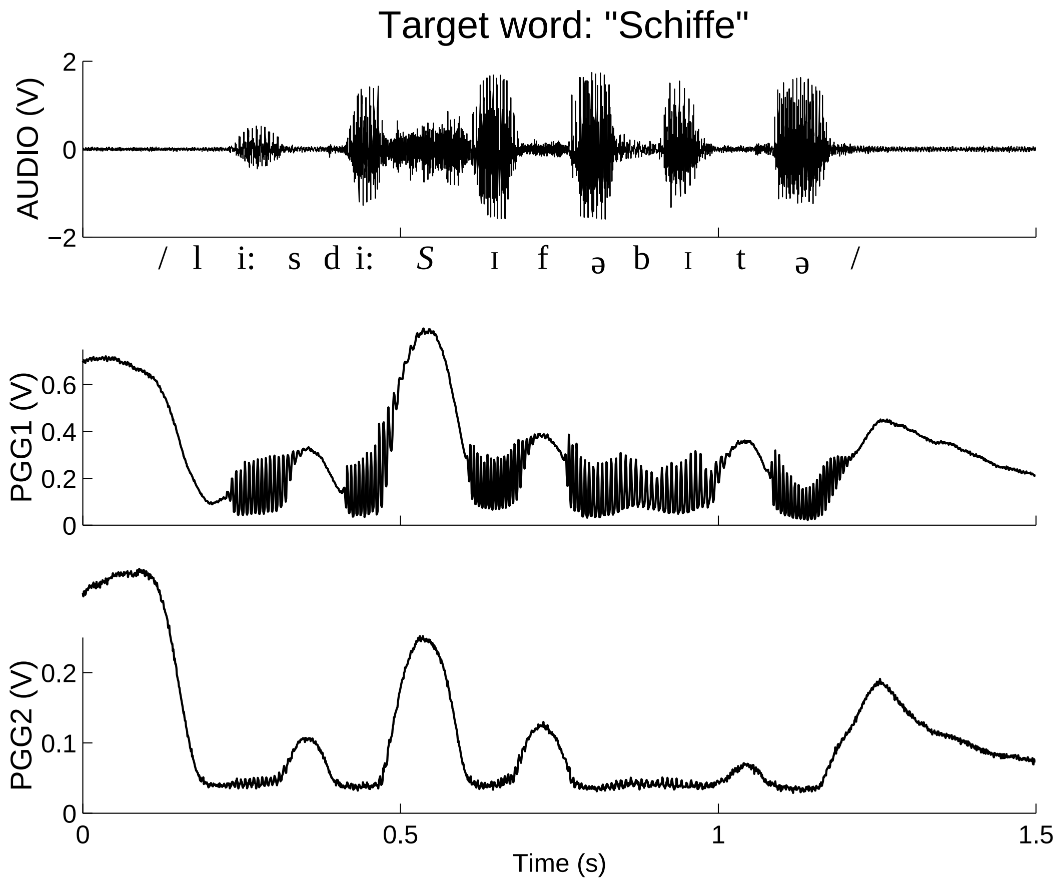

Fig. 1: *Example of two-channel transillumination signal for a German sentence ("Lies die Schiffe bitte") containg several voiceless sounds. The trace labelled PGG 1 (middle panel) is derived from a phototransistor applied externally to the neck between thryroid and cricoid cartilages. For PGG 2 (bottom panel) the phototransistor was located below the cricoid cartilage.*

Electromyography of the larynx

Laryngeal EMG has clearly contributed much to our understanding of articulatory functions of the larynx. See for example the work of Hirose (1975) in clarifying the status of the PCA (posterior cricoarytenoid) as a speech muscle, and in establishing the typical reciprocal pattern of activity of PCA and INT (interarytenoid) in the devoicing gesture (Sawashima & Hirose, 1983). The specific role with respect to the devoicing gesture of the other intrinsic muscles LCA (lateral cricoarytenoid), VOC (Vocalis) and CT (cricothyroid) appears to require further clarification (e.g Hirose & Ushijima, 1978).

The use of laryngeal EMG for articulatory investigations received its major stimulus from the development of hooked-wire electrodes (Basmajian & Stecko, 1962), since, in contrast to the concentric needle electrodes more common in clinical investigations, these electrodes allow the subject freedom of movement and cause little discomfort once the insertion (by means of hypodermic needle) has been completed (Hirano & Ohala, 1969; Shipp, Fishman, Morrissey & McGlone, 1970). The path of electrode insertion as further developed by Hirose (1971; summaries also in Hirose, 1979, and Hirano, 1981) involves a percutaneous approach for CT, LCA and VOC, and a peroral approach for PCA and INT. Further details of electrode construction, insertion and verification are to be found in the cited articles, as well as in Harris (1981) and Honda (1983).

The obvious drawback to the technique, unanimously alluded to by the above authors is the amount of practice required, including for example electrode insertions on cadaveric material. Hirano & Ohala also point out that the insertion of hooked-wire electrodes probably requires more skill than needle electrodes. A final consideration is the extent to which topical anaesthesia of the laryngeal mucosa and premedication of the subject (Hirose & Gay, 1972; Shipp et al., 1970) is necessary and compatible with the aims of a specific experiment.

Pulse-echo ultrasound

The ultrasound technique has provided much useful kinematic data on several articulatory systems (e.g. Keller & Ostry, 1983; Munhall, Ostry & Parush, 1985). Munhall &

Ostry (1985) report on its application to laryngeal kinematics. Briefly, the procedure is designed to measure the distance between an ultrasound transducer applied laterally to the thyroid cartilage and the moving vocal fold, sample rates of 1 kHz or more being possible. A potential advantage of the ultrasound technique is that, in contrast to transillumination, absolute values for the excursion of the vocal fold can be achieved. However, this may only be an apparent advantage because it is never possible to indicate precisely which point (or rather small area) on the vocal fold forms the basis of the measurement at any given moment or in any given session. Perhaps the main drawback currently with this interesting technique is that there are relatively few published investigations, so that little is known about possible sources of error; there are no studies of which we are aware which compare the technique with parallel recordings made with an alternative procedure such as fiberoptics (but see Hamlet, 1981, for extensive discussion of the utility of ultrasound in the analysis of phonation).

References

JASA = J. of the Acoustical Society of America

ABRILP = Annual Bulletin of the Research Institute of Logopedics and Phoniatrics

ARIPUC = Annual Report of the Institute of Phonetics of the University of Copenhagen

Baer, T., Löfqvist, A., & McGarr, N.S. (1983). "*Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques"*. JASA, 73, 1304-1308.

Basmajian, J.V & Stecko, G. (1962). "*A new bipolar hooked-wire electrode for electromyography"*. J. Applied Physiology, 17, 849.

Frøkjær-Jensen, B. (1967). "*A photo-electric glottograph"*. ARIPUC, 2, 5-19.

Frøkjær-Jensen, B., Ludvigsen, C. & Rischel, J. (1971). "*A glottographic study of some Danish consonants"*. In Hammerich, L., Jakobson, R. & Zwirner, E. (Eds.), Form and Substance. Copenhagen: Akademisk Forlag.

Fujimura, O., Baer, T. & Niimi, A. (1979). "*A stereoscopic fiberscope with magnetic interlens bridge for laryngeal observation"*. JASA, 65, 478-480.

Hamlet, S.L. (1981). "*Ultrasound assessment of phonatory function"*. In C.L. Ludlow & M. O'Connell Hart (eds.), Proc. of the conference on the assessment of vocal pathology (= ASHA reports 11) (pp.128-140).

Harris, K.S. (1981). "*Electromyography as a technique for laryngeal investigation"*. In C.L. Ludlow & M. O'Connell Hart (eds.), Proc. of the conference on the assessment of vocal pathology (= ASHA reports 11) (pp.70-87).

Hirano, M. & Ohala, J. (1969). "*Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles"*. J. Speech and Hearing Research, 12, 362-373.

Hirano, M. (1981). "*Clinical examination of voice"*. Springer Verlag: Wien, New York.

Hirose, H. (1971). "*Electromyography of the articulatory muscles: Current instrumentation and technique"*. Haskins SR 25/26, 75-86.

Hirose, H. & Gay, T. (1972). "*The activity of the intrinsic laryngeal muscles in voicing control - an electromyographic study"*. Phonetica, 25, 140-164.

Hirose, H. (1975). "*The posterior cricoarytenoid as a speech muscle"*. ABRILP, 9, 47-66.

Hirose, H. & Ushijima, T. (1978). "*Laryngeal control for voicing distinction in Japanese consonant production"*. Phonetica, 35, 1-10.

Hirose, H. (1979). "*Laryngeal EMG"*. ABRILP, 13, 13-22.

Hirose, H. (1986). "*Pathophysiology of motor speech disorders (dysarthria)"*. Folia Phoniatrica, 38, 61-88.

Hirose, H., & Niimi, S. (1987). "*The relationship between glottal opening and the transglottal pressure differences during consonant production"*. In Baer, T., Sasaki, C. & Harris, K.S. (Eds.), Laryngeal function in phonation and respiration (pp. 381-390). Boston: College Hill.

Honda, K. (1983). "*Electrical characteristics and preparation technique of hooked-wire electrodes for EMG recording"*. ABRILP, 17, 13-22.

Honda, K., Kiritani, S., Imagawa, H. & Hirose, H. (1987). "*High-speed digital recording of vocal fold vibration using a solid-stae image sensor"*. In Baer, T. et al (Ed.), Laryngeal function in phonation and respiration (pp. 485-491) Boston: College Hill.

Hoole, P., Schröter-Morasch, H. & Ziegler, W. (1997). "*Patterns of laryngeal apraxia in two patients with Broca's aphasia"*. Clinical Linguistics and Phonetics, 11, 429-442.

Hutters, B. (1976). "*Problems in the use of the photoelectric glottograph"*. ARIPUC, 10, 274-312.

Keller, E., & Ostry, D.J (1983). "*Computerized measurement of tongue-dorsum movement with pulsed-echo ultrasound"*. JASA, 73, 1309-1315.

Kiritani, S. (1971). "*X-ray monitoring of the position of the fiberscope by means of computer-controlled radiography"*.

ABRILP,5 , 35-39.

Kiritani, S., Imagawa, H. & Hirose, H. (1992). "*Vocal-cord vibration during consonants - High-speed digital imaging using a fiberscope"*. Proc. 1992 Int. Conf. on Spoken Language Processing, 1, 1661-1664.

Lisker L., Abramson, A.S, Cooper, F.S & Schvey, M. (1969). "*Transillumination of the larynx in running speech"*. JASA, 45, 1544-1546.

Löfqvist, A. & Yoshioka, H. (1980). "*Laryngeal activity in Swedish obstruent clusters"*. JASA, 63, 792-801.

Malécot, A. and Peebles, K. (1965). "*An optical device for recording glottal abduction-adduction during normal speech"*. Zschr. f. Ph., 18, 545-550.

Munhall, K. G., and Ostry, D. J. (1985). "*Ultrasonic Measurements of laryngeal kinematics"*. In Titze. I., and Scherer, R. (Ed.), Vocal fold physiology: Biomechanics, acoustics and phonatory control (pp. 145-162). Denver: Denver Center for the performing arts.

Munhall, K.G., Ostry, D.J. and Parush, A. (1985). "*Characteristics of velocity profiles of speech movements"*. J. Exp Psychology, Human Perception and Performance, 11, 457-474.

Ohala, J. (1966). "*A new photoelectric glottograph"*. UCLA Working Papers in Phonetics, 4, 40-54.

Sawashima, M. & Hirose, H. (1968). "*A new laryngopscopic technique by means of fiberoptics"*. JASA, 43, 168-169.

Sawashima, and Ushijima, T. (1971). "*Use of the fiberscope in speech research"*. ABRILP, 5, 25-34.

Sawashima, M., Abramson, A.S., Cooper, F.S. & Lisker, L. (1970). "*Observing laryngeal adjustments during running speech by use of a fiberoptics system"*. Phonetica, 22, 193-201.

Sawashima, M. & Miyazaki, S. (1974). "*Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes"*. ABRILP, 8, 7-12.

Sawashima, M. and Hirose, H. (1983). "*Laryngeal gestures in speech production"*. In MacNeilage, P.F. (Ed.), The Production of Speech (pp. 11-38). N.Y: Springer.

Shipp, T., Fishman, B.V., Morrissey, P. & McGlone, R.E. (1970). "*Method and control of laryngeal EMG electrode placement in man"*. J. Acoustical Soc. America, 48, 429-430.

Sonesson, B. (1960). "*On the anatomy and vibratory pattern of the vocal folds"*. Acta Otolarynologica, Supplement, 156, 1-80.

Yoshioka, H. (1984). "*Glottal area variation and supraglottal pressure change in voicing control"*. ABRILP, 18, 45-49.