Reprinted from

939

SPEECH COMMUNICATION

Speech Communication 16 (1995) 49–66

Vocal tract aerodynamics in /aCa/ utterances: Measurements

Anders Löfqvist *, Laura L. Koenig, Richard S. McGowan

Haskins Laboratories, 270 Crown Street, New Haven, CT, USA

Received 3 May 1994; revised 17 October 1994

N·H
ELSEVIER

Scope. *Speech Communication* covers all aspects of speech communication processes between humans as well as between human and machines. *Speech Communication* features original research work, tutorial and review articles dealing with the theoretical, empirical and practical aspects of this scientific field. Special emphasis is given to material containing an interdisciplinary point of view.

Editorial policy. *Speech Communication* is an interdisciplinary Journal whose primary objective is to fill the need for the rapid dissemination and thorough discussion of basic and applied research results. In order to establish frameworks within which results from our different fields may be interrelated, emphasis will be placed on viewpoints and topics able to induce transdisciplinary approaches. The editorial policy and the technical content of the Journal are the responsibility of the Editors and the Advisory Committee. The Journal is self-supporting from subscription income. Advertisements are subject to the prior approval of the Editors.

Subject coverage. Subject areas covered in this journal include:
(1) Basics of oral communication and dialogue: modelling of production and perception processes, phonetics and phonology, syntax, semantics and pragmatics of speech communication, cognitive aspects.
(2) Models and tools for language learning: functional organisation and developmental models of human language capabilities, acquisition and rehabilitation of spoken language, speech and hearing defects and aids.
(3) Speech signal processing: analysis, coding, transmission, enhancement, robustness to noise.
(4) Models for automatic speech communication: speech recognition, language identification, speaker recognition, speech synthesis, oral dialogue.
(5) Development and evaluation tools: monolingual and multilingual data bases, assessment methodologies, specialised hardware and software packages, field experiments, market development.

Subscription information. *Speech Communication* (ISSN 0167-6393) is published in two volumes (eight issues) a year. For 1995 Volumes 16–17 are scheduled for publication. Subscription prices are available upon request from the publishers. Subscriptions are accepted on a prepaid basis only and are entered on a calender year basis. Issues are sent by surface mail except to the following countries where air delivery (S.A.L. – Surface Air Lifted) is ensured: Argentina, Australia, Brazil, Canada, China, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For the rest of the world, airmail and S.A.L. charges are available upon request. Claims for missing issues will be honoured free of charge within six months after the publication data of the issues. Mail orders and inquiries to: Elsevier Science B.V., Journals Department, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. For full membership information of the Associations, possibly combined with a subscription at a reduced rate, please contact: EURASIP, P.O. Box 134, CH-1000 Lausanne 13, Switzerland; ESCA, BP 7, B-1040, Brussels 40, Belgium.

Published eight times a year 0167-6393/95/$09.50 Printed in The Netherlands

ELSEVIER

Speech Communication 16 (1995) 49–66

SPEECH COMMUNICATION

Vocal tract aerodynamics in /aCa/ utterances: Measurements

Anders Löfqvist *, Laura L. Koenig, Richard S. McGowan

Haskins Laboratories, 270 Crown Street, New Haven, CT, USA

Received 3 May 1994; revised 17 October 1994

Abstract

This paper examines air flow patterns at vowel-consonant and consonant-vowel transitions. Oral air flow was recorded in six speakers of American English producing reiterant speech. The air flow signal was inverse filtered to obtain an estimate of the glottal pulse. Measurements were made of peak and minimum flow, open quotient, pulse area and fundamental frequency. The results show that at the transitions between vowels and voiceless consonants the pulse properties show large variations. In particular, the source is characterized by a breathy mode of phonation. Breathiness was indexed by large values of peak and minimum flow, and an open quotient close to 1. The observed variations can be accounted for by the laryngeal adjustments that are made for voiceless consonants, in particular the glottal opening movement and its phasing with the oral articulatory events. Individual differences suggest that speakers vary in their use of the longitudinal tension of the vocal folds in controlling voicelessness.

Zusammenfassung

In diesem Artikel werden typische Formen des Luftstroms bei Vokal-Konsonant und Konsonant-Vokal-Übergängen untersucht. Der orale Luftstrom von sechs amerikanischen Englischsprechern, die reiterative Sprache erzeugten, wurde aufgezeichnet. Das Luftstromsignal wurde invers gefiltert, um so die Anzahl der Glottisschläge annähernd bestimmen zu können. Die Messungen betrafen den Gipfel und das Minimum des Luftstroms, den Öffnungsgrad, den Impulsbereich und die Grundfrequenz. Die Ergebnisse zeigen, daß an den Übergängen von Vokal zu stimmlosem Konsonant die Eigenschaften der Impulse stark variieren. Insbesondere ist die Quelle durch eine gehauchte Aussprache gekennzeichnet. Diese Aspiration wird durch große Gipfel- und Minimumwerte des Luftstroms und durch einen Öffnungsgrad nahe 1 belegt. Die beobachteten Variationen können durch die Anpassung des Kehlkopfs bei der Erzeugung stimmloser Konsonanten erklärt werden, vor allem durch die Bewegung zur Glottisöffnung und die Synchronisation an die vokalischen Artikulationsbegebenheiten. Die individuellen Variationen lassen darauf schließen, daß die Sprecher für die Kontrolle der Stimmlosigkeit unterschiedlichen Gebrauch von der Längsspannung der Stimmritze machen.

* Corresponding author. Tel.: (203) 865-6163; Fax: (203) 865-8963; E-mail: LOFQUIST@YALEHASK (Bitnet), LOFQUIST@HASKINS.YALE.EDU (Internet).

SSDI 0167-6393(94)00049-2

Résumé

Cet article examine les formes d'écoulement d'air lors des transitions voyelle-consonne et consonne-voyelle. L'écoulement d'air buccal a été enregistré chez 6 locuteurs anglais-américains répétant des énoncés. Un filtrage inverse du signal a été effectué pour obtenir une estimation des impulsions glottiques. On a mesuré les écoulements maximum et minimum, le quotient d'ouverture, la zone des impulsions et la fréquence fondamentale. Les résultats montrent de grandes variations dans les caractéristiques des impulsions à la transition entre voyelles et consonnes sourdes. En particulier, la source est caractérisée par un mode de phonation rauque. Cette raucité est associée à de larges variations de l'écoulement et à une valeur du quotient d'ouverture proche de 1. Les variations observées peuvent être dues aux ajustement laryngés effectués pour les consonnes sourdes, en particulier le mouvement d'ouverture de la glotte et sa relation de phase avec les événements articulatoires buccaux. Les différences individuelles observées suggèrent que les locuteurs diffèrent dans leur façon d'utiliser la tension longitudinale des cordes vocales pour contrôler le non-voisement.

Keywords: Articulatory timing; Speech aerodynamics; Voice source properties

1. Introduction

In the production of stop and fricative consonants, several different valves in the vocal tract are operated to regulate air pressure and flow. At the laryngeal level, the glottal valve is opened for voiceless consonants by movements of the arytenoid cartilages. At the pharyngeal level, the velar port is closed by elevation of the soft palate. In the oral cavity, a complete closure or a narrow constriction is made by actions of the jaw, the tongue and/or the lips. All these valving operations contribute to creating the acoustic properties of obstruent consonants: a transient or continuous noise source and a period of voicelessness for voiceless consonants. The timing of the opening and closing of these valves, in particular the laryngeal and oral valves, is critical for producing contrasts of voicing and aspiration in stop consonants, e.g. (Abramson, 1977; Dixit, 1989; Löfqvist, 1980, 1992).

The acoustic cues for voiceless consonants are complex, spread over time, and involve variations in the sound source and the spectral composition of the signal. For example, in the production of a voiceless fricative in a vocalic environment, the sound source changes from periodic to aperiodic and then back to periodic. A voiceless (post) aspirated stop in the same environment is associated with the following sequence of source changes: periodic voicing during the preceding vowel, silence during the closure, transient noise at the release, aspiration noise, and periodic voicing during the following vowel. While there exists a large body of work on the acoustic properties of stops and fricatives, aerodynamic studies of obstruent production have mostly been concerned with peak values of intraoral air pressure and oral air flow, e.g. (Isshiki and Ringel, 1964; Subtelny et al., 1966; Arkebauer et al., 1967), and only a few have dealt with pressure and flow profiles, e.g. (Klatt et al., 1968; Scully et al., 1992a, 1992b). In addition, detailed investigations of the variations in the voice source at transitions between vowels and obstruents and between obstruents and vowels have only recently been performed, e.g. (Gobl, 1988; Gobl and Ní Chasaide, 1988; Nittrouer et al., 1990; Koreman et al., 1992; Löfqvist and McGowan, 1992; Palmer and House, 1992).

Löfqvist and McGowan (1992) used recordings of oral-nasal air flow to infer properties of the voice source, such as peak and minimum flow and the open quotient, during individual glottal pulses. The speech material consisted of VCV sequences, with an open vowel and the consonant one of the set /p, b, s, v, m, h, ʔ sp/. The flow was first inverse filtered, and then low-pass filtered to obtain an estimate of the glottal pulse. The results suggested that there were considerable influences on the source from the consonant both before the end of the preceding vowel and after the onset of the following vowel. These influences were most marked for voiceless consonants, where the source is breathy at both vowel offset and vowel onset. Indices of breathiness were large values of peak and minimum flow, and an open quotient close to 1. Analysis of laryngeal articulatory movements obtained simultaneously with the flow data suggested that these variations can be largely accounted for by the patterns of coordination between the laryngeal and oral valves. In particular, in anticipation of the glottal opening movement for an upcoming voiceless consonant, the degree of glottal abduction begins to increase during the vowel. This results in a gradual change of voice quality from modal to breathy. Similarly, at the release of a voiceless aspirated stop, the vocal folds are in the process of being adducted. At the onset of vibrations, the source pulses are characterized by a high rate of air flow and an open quotient close to 1 as a result of the incomplete glottal closure. During the first part of the vowel, the mode of glottal vibration changes from a breathy to a modal one. At the acoustic level, the change from a more breathy to a modal type of phonation is accompanied by a decrease in the noise components of the source and a decrease in the tilt of the source spectrum; see (Nittrouer et al., 1990) for acoustic analyses of vowels in different consonantal contexts.

The present study is a continuation and an extension of the work reported by Löfqvist and McGowan (1992). Their work was limited in the number of subjects investigated and in the technical procedure, in particular the use of a static inverse filter. Therefore, the present study examines six subjects in order to assess further variations between speakers. It also uses a time-varying software filter for the inverse filtering of the oral flow.

While it is thus possible to draw conclusions about the articulatory activities that generate different aerodynamic patterns, it is important to remember that the oral flow is not identical to the glottal flow. There are several factors contributing to oral flow in addition to the glottal flow, such as articulatory movement and vocal tract wall compliance. Hence, in an accompanying study (McGowan et al., 1995), we use an analysis-by-synthesis approach to assess the relative contributions of these factors, and to examine the effects of interarticulator timing and vocal fold myodynamics on the observed patterns of oral air flow using articulatory and aerodynamic modeling.

2. Procedure

2.1. Data recording

Air flow was recorded using a face mask and a differential pressure transducer (Glottal Enterprises) according to the method described by Rothenberg (1973). The system has a flat frequency response from DC to above 1 kHz, cf. (Badin et al., 1990). The flow signal was calibrated using a rotameter. During the experiment, the flow signal, together with the audio signal, was digitized at 10 kHz with 12 bit resolution. Since the mask covers the mouth and the nose, the flow signal represents both oral and nasal flow.

Six native speakers of American English, three females and three males without any history of voice disorders, produced twelve repetitions of the material listed in Table 1. The subjects produced these nonsense utterances as reiterant speech modeled after the sentence "It's raining in Oslo" at a self-selected speaking rate and intensity level. Reiterant speech was used to control for segmental effects and to obtain as uniform an intonation pattern as possible. An open vowel was used to ensure that the frequency of the first formant was considerably higher than

Table 1
The linguistic material (underlining marks the syllable carrying the sentential stress)

ma pa ma ma a ma
ma pa ma ma ba ma
ma pa ma ma ma ma
ma pa ma ma ha ma
ma pa ma ma va ma
ma pa ma ma sa ma
ma pa ma ma pa ma
ma pa ma ma spa ma

that of the fundamental; a nasal consonant was selected because it does not involve any specific laryngeal adjustments and thus minimizes changes in the setting of the larynx. The use of a single vowel carrying the sentential stress in the first utterance was intended to make the subjects produce a glottal stop (a hard attack) at the onset of this vowel.

2.2. Data processing

The processing performed in this study is slightly different from that performed in (Löfqvist and McGowan, 1992), but similar information is derived: the bandlimited (i.e., DC to less than 2 kHz) properties of the pulsed flow as a function of phonetic context. In both the previous and current study, the analysis is restricted by the frequency response of the mask used to record air flow. Here, instead of using a fixed hardware filter to smooth the flow pulse as was done in our earlier study, a software filter was used to infer adaptively the frequencies and bandwidths of ripples appearing in the flow signal during the glottal pulse and to filter them out. In both the earlier and the present study, a linear-phase low-pass filter was used for additional filtering of the inverse-filtered flow signal. In (Löfqvist and McGowan, 1992), a linear phase low-pass filter with a sharp cutoff was designed using the Remez exchange algorithm with the bandwidth designed to include the same number of harmonics regardless of the subject and the fundamental frequency. In the procedure used here, a simple five point triangular window, which is 6 dB down at about 2 kHz, was used for all subjects. In neither study is there a claim to complete inverse filtering, but, rather, filtering to remove the ripples found on the flow pulses.

After sampling, portions of the flow signal before and after the target consonant were extracted for software inverse filtering. Before each file was filtered, the open and closed phase of each pulse was marked by hand from the flow file. A completely closed phase analysis would have been preferable (Krishnamurthy and Childers, 1986; De Veth et al., 1990) but this was impossible due to the nature of the data set. In particular, the flow signal close to the boundary between a vowel and a voiceless consonant typically did not show any closed phase. When there was no closed phase, a relatively flat portion of the pulse near the minimum flow portion was used as the "closed phase". Sometimes this meant that only a few samples (< 10) were included in the "closed phase" portion of the flow pulse. While an electroglottographic signal could have been helpful in estimating the closed phase intervals in the more adducted conditions (Krishnamurthy and Childers, 1986), the transition regions between vowels and consonants were also of interest to us, so the cruder method of choosing closed intervals from the flow signal was used.

Linear Predictive Coding with covariance analysis and root finding was used to estimate the formant frequencies and bandwidths for removing ripple in the flow pulses. Default values for the analyses were used except where the filtering resulted in spurious high frequency formants, or no first or second formants. The default order of the analysis filter was 12 and the window size was 80, or 8 ms. A window of this size contains both maxima and minima of the glottal pulses. In those cases where the analysis did not work (e.g., the resulting pulses were not smooth or the program crashed), the order of analysis could be reset to 10, or the window size could be reset to anywhere between 60 to 120. Various parameters were set to exclude spurious formant frequencies at this stage. These included the minimum and maximum frequencies of 500 and 5000 Hz, respectively, as well as the minimum and maximum bandwidths of 50 and 500 Hz. While 50 Hz is a reasonable bandwidth for a first or second formant in a closed glottis condition (Fant, 1961), the analysis window was long enough so that the first formant was both damped and moved by subglottal coupling, both effects adding to the apparent bandwidth (Fant, 1972). Setting the minimum bandwidth lower would produce poles at harmonics of the voice.

Each analysis window used to compute the filter started at some time during the closed phase. A new filter was computed every 15 samples, as long as the earliest point in the analysis window was in the closed phase. By requiring the window to start in the closed phase, we tried to minimize effects due to changing relative phase between window and voice pulse. An improved method would have used a pitch synchronous analysis (Rabiner et al., 1977; Pierrehumbert and Talkin, 1992, p. 99). However, because the speech material included aspirated and glottalized sounds, a pitch synchronous analysis would have been difficult. The method applied here usually produced reasonable estimates of two frequencies and bandwidths. When reasonable estimates were not obtained, the files containing the frequencies and bandwidths used for filtering were edited to carry forward the latest good estimate or to bring back the earliest good estimate of the formants and bandwidths within the same period. The inverse filter was applied after editing the frequency and bandwidth values.

In addition to its bandlimited aspect, there are some reasons not to refer to the processed flow as glottal flow. As noted above, our inversion procedure could not recover either the true glottal flow or the short circuit glottal flow (Ananthapadmanabha and Fant, 1982; Fant, 1986). The signal recovered here could not be the former because it was not a strictly closed phase analysis, and it could not be the latter because no attempt was made to factor out all tract interactions with the glottal flow. In our processing, some of the effects of time-varying supra- and sub-glottal formants were removed, as was any superposed ripple from previous voice pulses or excitations within a pulse (Fant, 1972, 1986; Cranen and Boves, 1987; Gobl, 1988). Because there is a time-varying glottal area during the analysis windows, the effect of subglottal zeroes should be minimized. Also, the first subglottal resonance has a bandwidth broad enough that it should not have a great effect on the results. However, the effects of these phenomena on the spectral skewness, amplitude and average values of the glottal flow were not removed. Also, as mentioned above, the low-frequency aerodynamic effects of the supra-laryngeal tract were not removed.

In sum, our procedure can be considered as a bandlimited analysis of the oral-nasal flow with inverse filtering applied to smooth out the oscillations in the minimum flow regions of the pulses. As noted by Gobl and Ní Chasaide (1988), the Rothenberg mask does not have a frequency response such that a detailed model, such as the L–F model (Fant et al., 1985), can be fitted to the resulting waveform. However, from these measurements and with the help of flow simulations, we feel confident that inferences about glottal activity can be made.

2.3. Measurements

In preparing the filtered flow signal for measurements, relevant points in the waveform were labeled interactively on a computer. As illustrated in Fig. 1, these points were the onset (T1) and offset (T3) of the open phase during each period, and peak flow (T2). Points T1 and T3 were defined on the basis of a rapid change in the signal. This definition of the closed phase allows for flow occurring during the "closed" phase. Based on these labels, measurements were extracted for a predetermined number of periods. These measurements were selected to provide information on laryngeal adjustments, cf. (Gauffin and Sundberg, 1989), and also to make our results comparable to those presented in related studies. The value of peak flow was measured at point T2. Minimum flow during each cycle was taken as the average of the flow at points T3 and T4. The open quotient was calculated as the ratio between the open phase, T1–T3, and the period

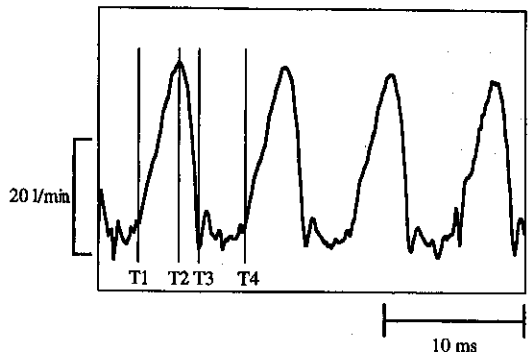

Fig. 1. A filtered flow signal with labels during a single pulse.

time, T1–T4. The area of the pulse was obtained in the following way. First, the flow signal was integrated between points T1 and T3. Second, a rectangle was defined with its length equal to the duration of the open phase (T1–T3) and its height equal to the mean of the flow at points T1 and T3. Third, the area of the rectangle was subtracted from the value obtained by the integration. Thus, the area of the pulse represents the area of the open phase with the base, or DC, flow removed.

Measurements were made before the offset of the vowel preceding and after the onset of the vowel following the target consonant. The identification of offsets and onsets was unproblematic in most cases (cf. Fig. 3 in (Löfqvist and McGowan, 1992) for examples of acoustic, aerodynamic and glottographic recordings of the different /aCa/ sequences). For the voiced consonants /b, v, m/, the vibrations continued through the consonant. Here, the first and last glottal pulses were defined as the last period before a significant reduction in flow amplitude and the first period with a significant increase in flow amplitude, respectively. Also for the laryngeal fricative /h/, glottal vibrations sometimes continued uninterruptedly through the consonant. Here, the offset was taken as the last period during the increase in air flow, and the onset as the first period during the decrease in air flow. Two subjects typically produced the /h/ with glottal vibrations throughout. In these two cases, no time interval separated the measurements for the two vowels, cf. below and Fig. 8. Twelve periods before vowel offset and 25 periods after vowel onset were measured. Fewer periods were measured before vowel offset because the first, unstressed vowel of the VCV sequence was short and often did not contain many glottal periods; in a few productions, some of the early periods measured before vowel offset actually occurred in the nasal preceding the vowel. The number of periods measured after vowel onset was selected to keep the measurement task manageable and still allow air flow to reach an asymptotic value during the vowel. This turned out not always to be true for female voices with a higher fundamental frequency, however.

3. Results

In the plots to be discussed below, the x-axis has been divided into two intervals showing the periods before vowel offset to the left, and the periods after vowel onset to the right. The time scale is given in milliseconds. To derive this time scale for the individual plots, the mean length of each period was calculated. The running total of period lengths determined the location of individual points along the x-axis. Note that the time scales differ between subjects, since the same number of periods will cover different temporal intervals depending on the fundamental frequency of the subject. In this way, the differences in fundamental frequency between male and female voices as well as differences in fundamental frequency following different consonants are visible in the plots.

The results to be reported here consist of averages over the 12 repetitions of each utterance. In order to keep the graphs clear and readable, no error bars have been included. Since there are obviously variations between repetitions of the same VCV sequence by a given subject, though, we shall briefly address this variability before proceeding with the main results. Fig. 2 plots peak flow during all productions of the sequences /aʔa/ and /apa/ by a male (VG) and a female (LZ) subject. While the values differ between productions, the pattern of change at the transition from the vowel to the consonant and from the consonant to the vowel is very consistent within the same sequence. For example, peak flow decreases during the vowel preceding the glottal stop /ʔ/, and increases after the onset of the vowel following it. This pattern is found for both subjects. One consequence of our measurement procedure, briefly discussed above, should also be pointed out here, since it is evident in Fig. 2. In this figure, the peak flow pattern in the vowel following the glottal stop is rising and falling for subject VG, but rising for subject LZ. The most likely explanation for this difference is that the measurements represent temporal windows of unequal duration due to the difference in fundamental frequency between the male and the female subject. Thus, for subject

LZ, the measured pulses would seem to correspond to those occurring during the initial increase in peak flow for subject VG (i.e., until approximately 100 ms into the following vowel).

In the following plots, the results for the male and female subjects are grouped together. The male subjects are ES, RM and VG, while the female subjects are CS, LK and LZ. In discussing the main results, we shall first deal with the pattern before vowel offset; second, the pattern after vowel onset; third, differences between female and male subjects; fourth, individual differences; and finally, possible effects of sentential stress.

Fig. 3 plots mean peak flow for all subjects. Before the offset of the vowel, peak flow shows an increase before the consonants /h, s/ and the cluster /sp/. Note that the pattern is virtually identical before the single fricative /s/ and the cluster /sp/. Peak flow generally decreases before the glottal stop /ʔ/ and the consonants /b, v, p, m/. After vowel onset, peak flow decreases following the laryngeal fricative /h/ and increases following the glottal stop /ʔ/. Peak flow also tends to decrease following the voiceless consonants /p, s/. Following the voiced stop and fricative /b, v/, peak flow generally shows a decreasing and an increasing-decreasing pattern,

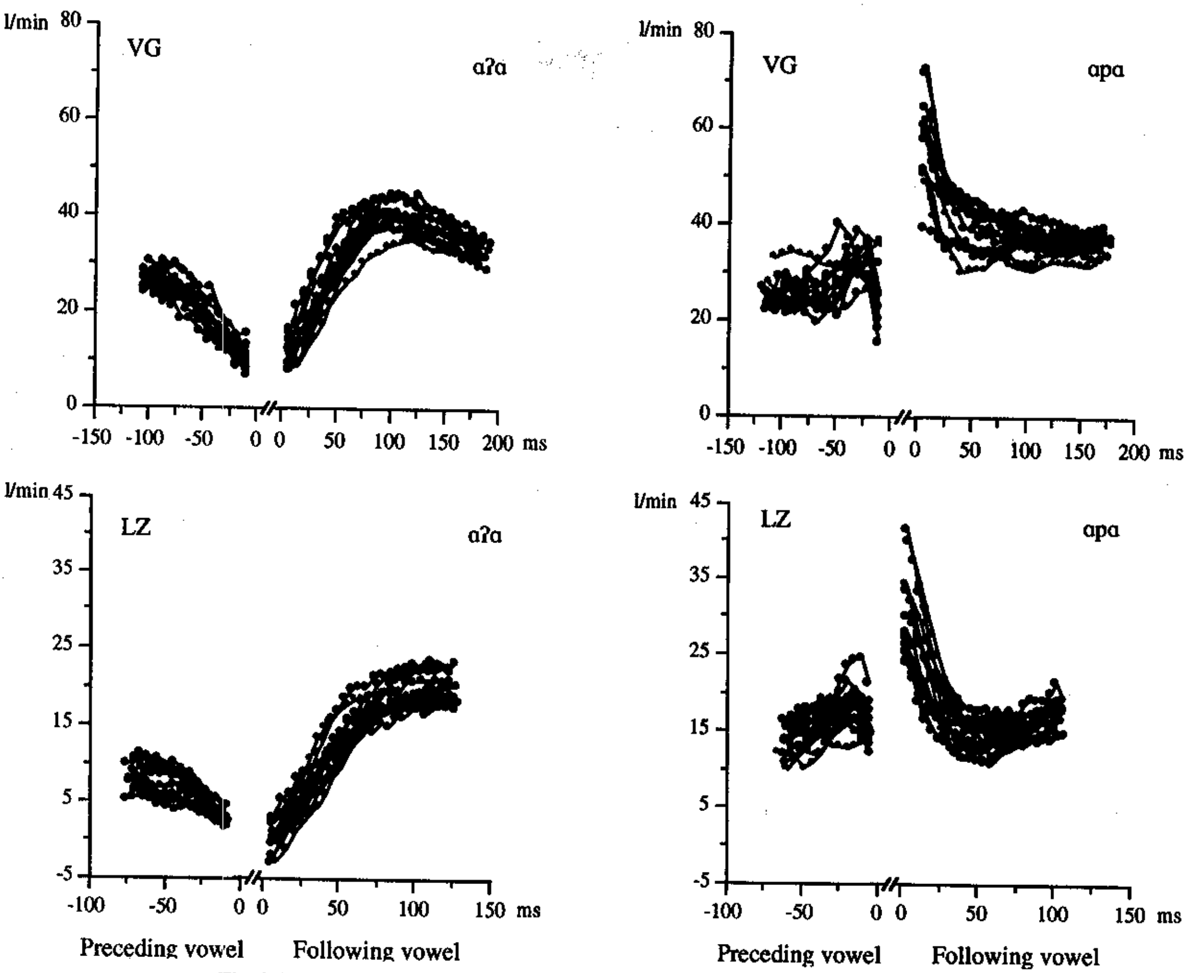

Fig. 2. Peak flow for all productions of /aʔa/ and /apa/ by subjects VG and LZ.

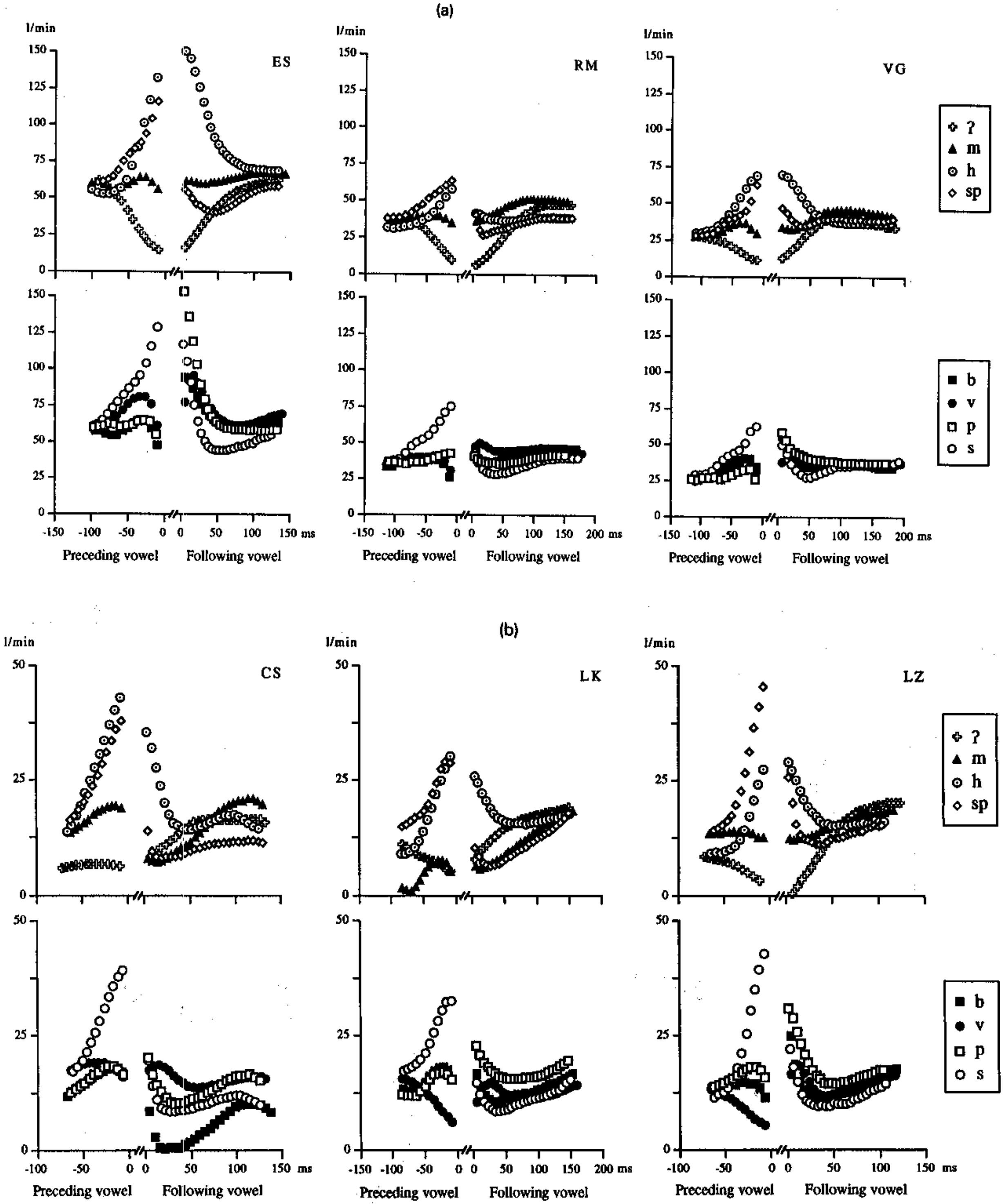

Fig. 3. Mean peak flow for all subjects.

respectively. If we compare the peak flow values following the aspirated single stop /p/ and the unaspirated stop in the cluster /sp/, we see that peak flow tends to be higher following the aspirated stop. Peak flow is consistently higher for the male subjects. This is the case not only for the high peak flow values in the vicinity of voiceless consonants, but also during the second vowel when peak flow has reached a steady level.

The influence of the upcoming consonant can be seen at least 50 ms before the offset of the vowel. At this point in time, the curves associated with the different consonants in Fig. 3 begin, or have already begun, to diverge from each other. In the following vowel, the influence of the consonant can be seen at least 50–75 ms after vowel onset. For some subjects, the curves have not converged at the rightmost data points in the plots. Again, remember the difference in fundamental frequency across subjects and across consonantal context within a subject.

The effect of sentential stress on peak flow can be examined by comparing the leftmost and rightmost data points in each plot. The leftmost data points occur early during the first vowel (or during the preceding nasal), where the influence of the upcoming consonant can be assumed to be at a minimum. Similarly, the rightmost data points

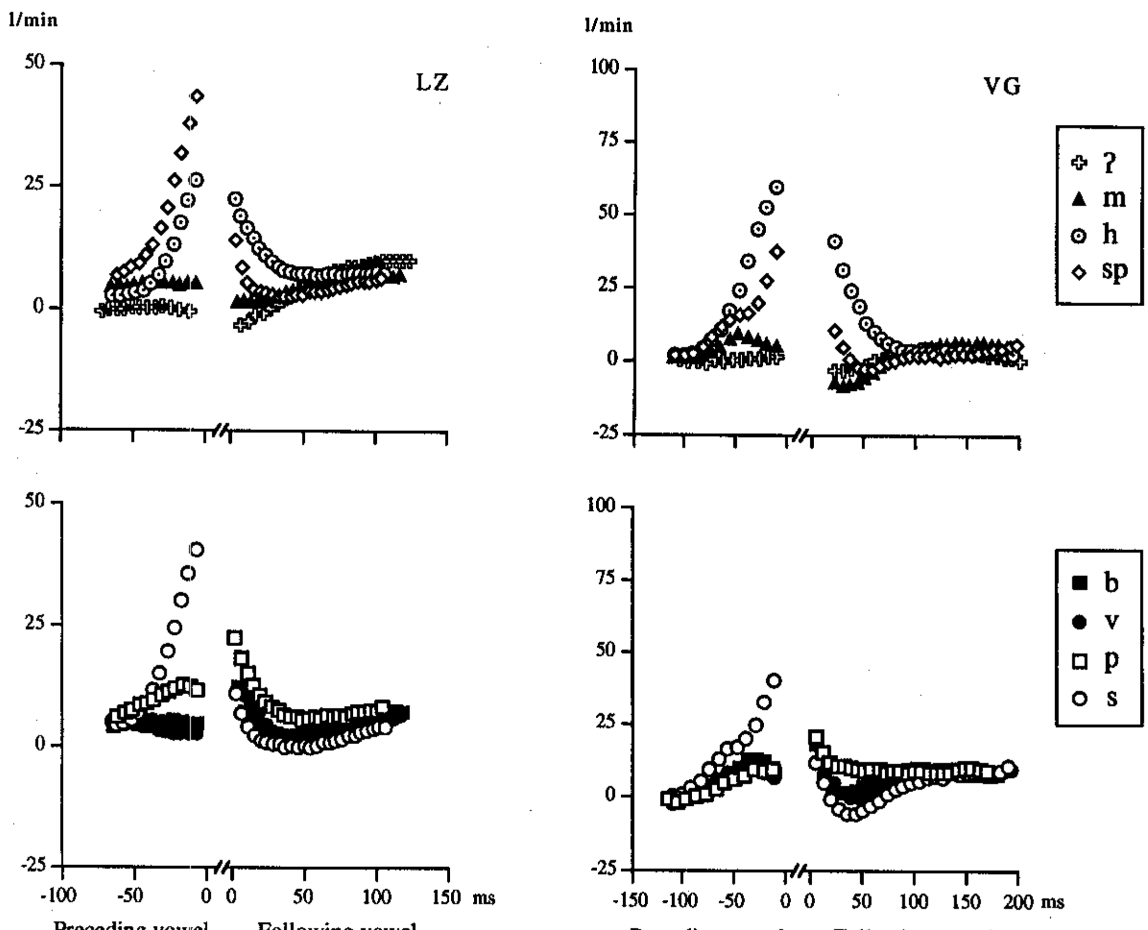

Fig. 4. Mean minimum flow for subjects LZ and VG.

occur late in the second vowel, when the influence of the preceding consonant has decreased. The overall tendency is for peak flow to be higher at the rightmost data points, i.e., during the second vowel where the sentential stress occurs.

Minimum flow shows variations that closely

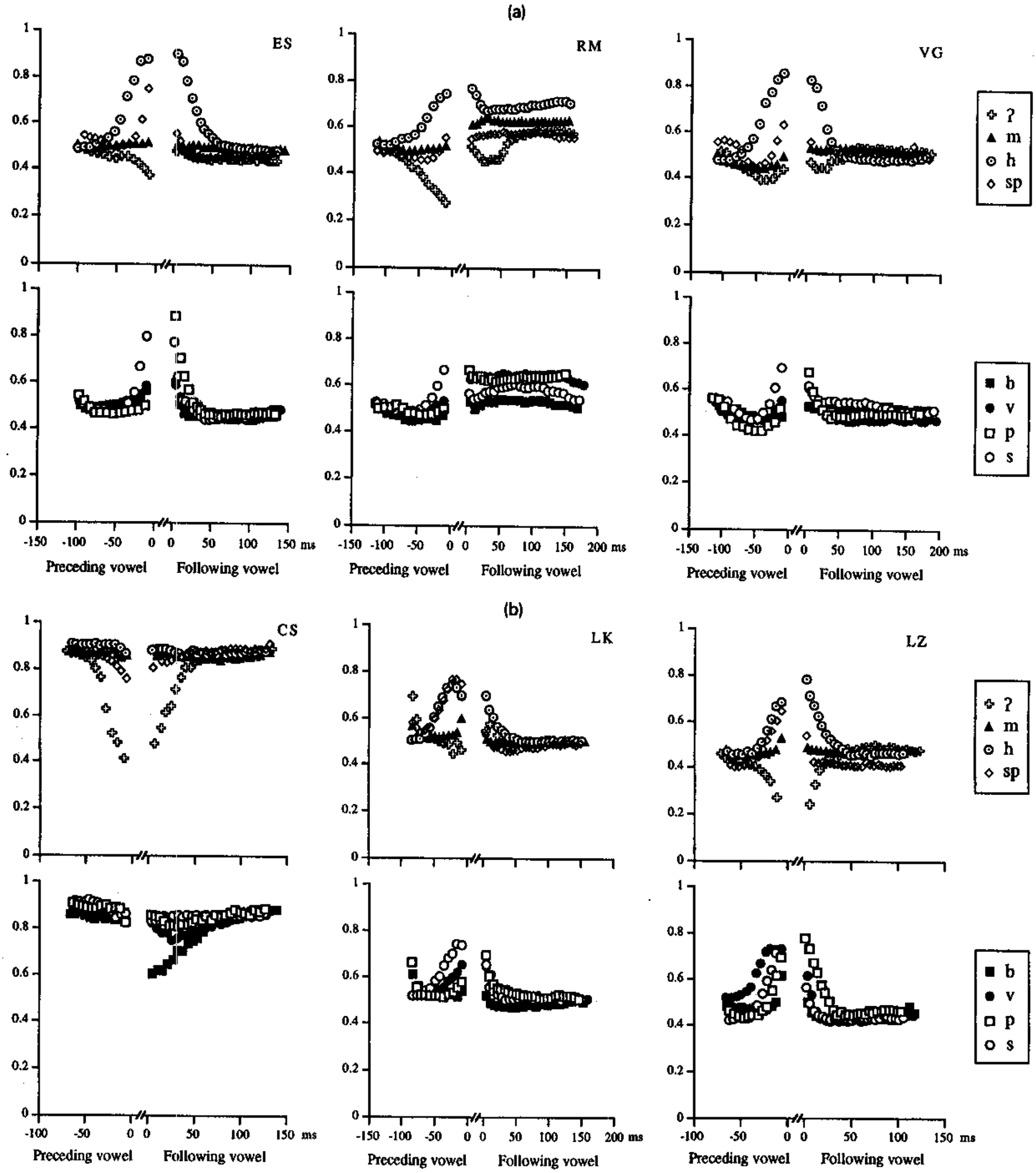

Fig. 5. Mean open quotient for all subjects.

parallel those of peak flow. This is illustrated in Fig. 4 which plots mean minimum flow for one male (VG) and one female (LZ) subject. Before vowel offset, minimum flow increases before the consonants /h, s, sp/. Again, the pattern is similar before the /s/ and the cluster /sp/. In contrast to peak flow, minimum flow does not show a marked decrease before the glottal stop.

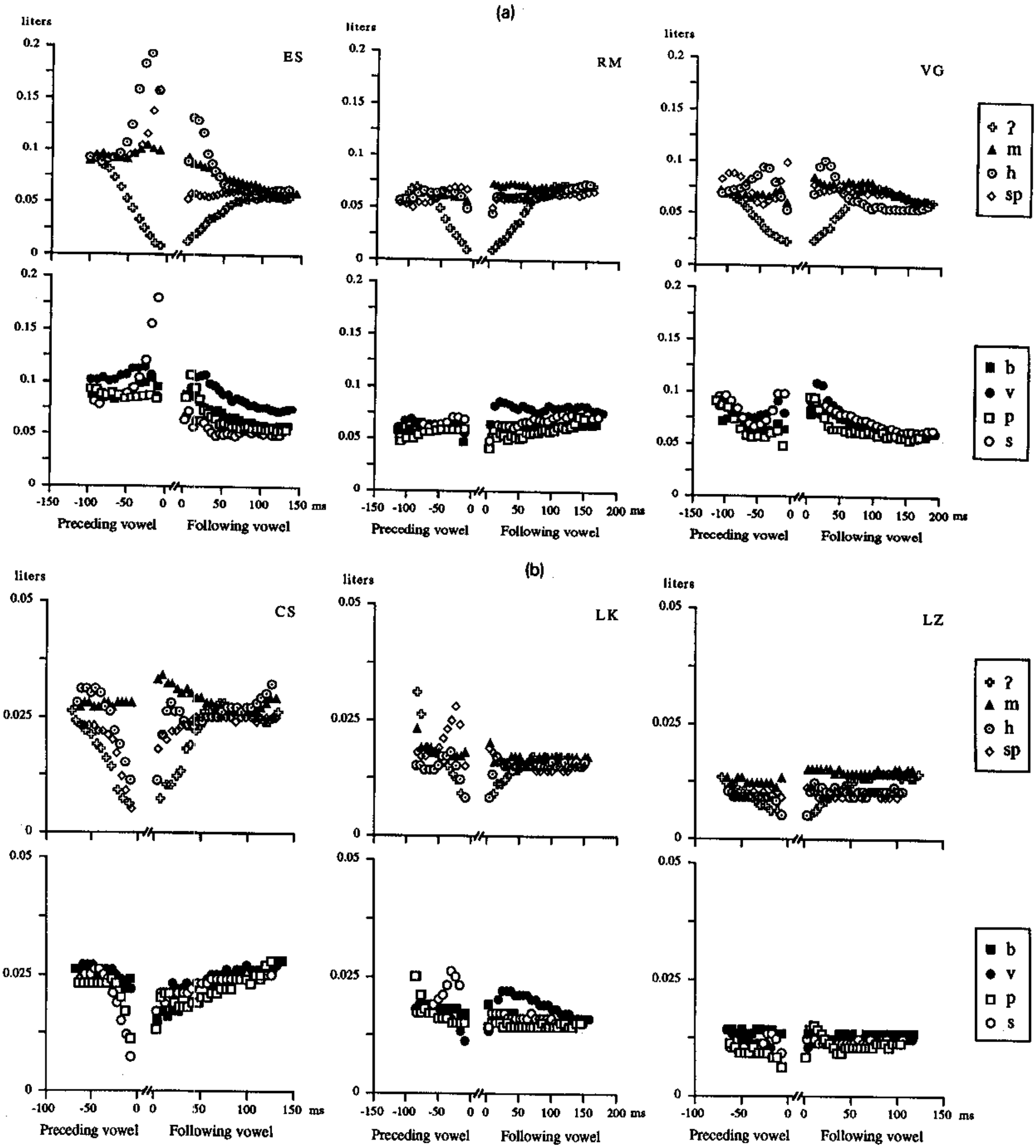

Fig. 6. Mean area of the pulse for all subjects.

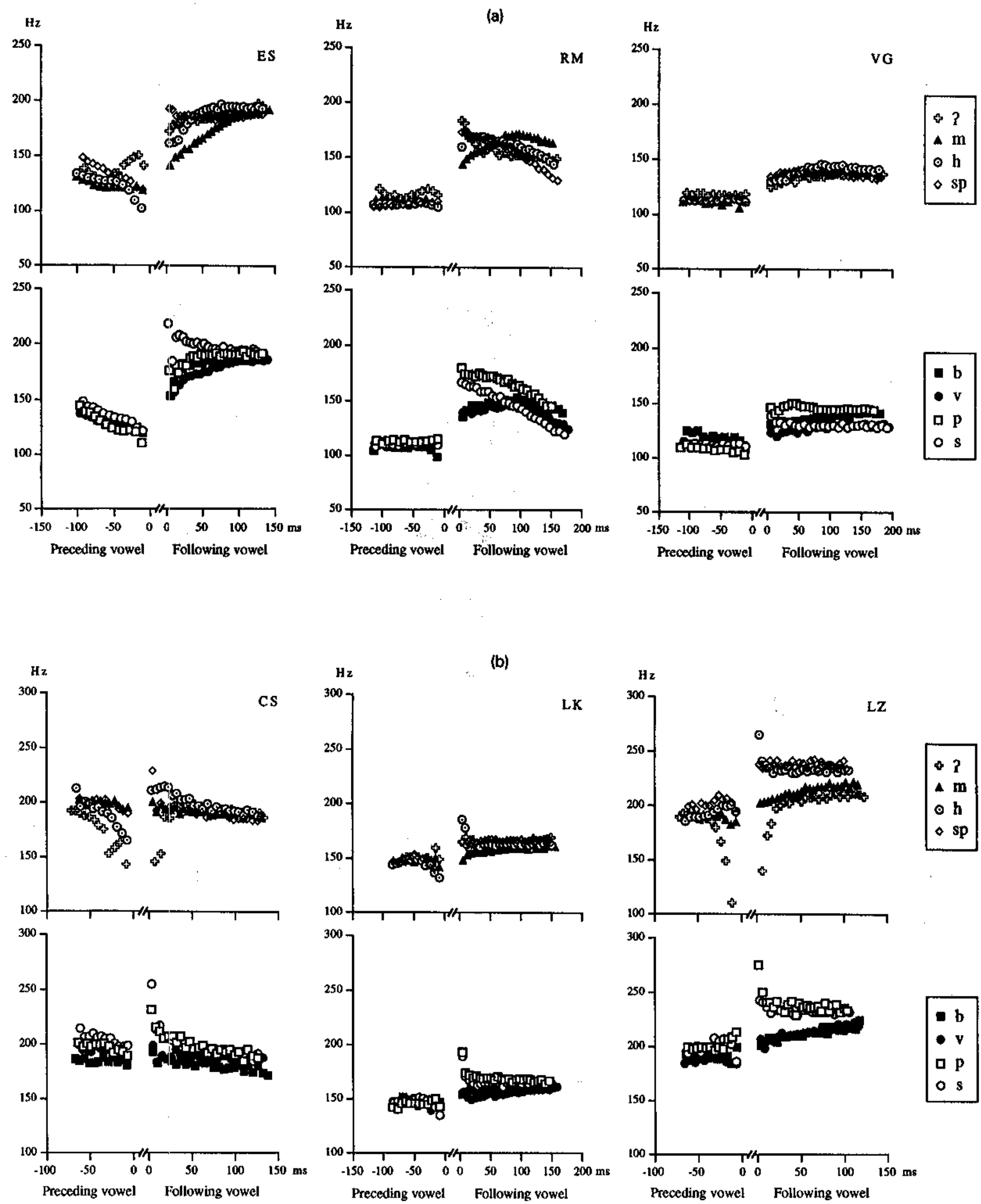

Fig. 7. Mean fundamental frequency for all subjects.

After vowel onset, minimum flow decreases after the consonants /h, sp, b, v, p, s/. Also, minimum flow is higher following the aspirated stop /p/ than after the unaspirated stop in the cluster /sp/. The minimum flow values do not appear to differ between the female subject and the male subject. A comparison between the leftmost and rightmost data points shows that the minimum flow is higher at the rightmost points. This would suggest that the minimum flow is higher during the vowel carrying the sentential stress.

The results for the mean open quotient are shown in Fig. 5. The open quotient generally increases before vowel offset when the following consonant is one of /h, s, sp/. A decrease in the open quotient tends to occur before the glottal stop /ʔ/. After vowel onset, the open quotient decreases following the consonants /h, p/. For all subjects except RM, the open quotient is larger following the aspirated stop /p/ than after the unaspirated stop in the /sp/ cluster. There do not appear to be any clear differences in the open quotient between male and female voices. Subject CS has a very large open quotient overall compared to all the other subjects; in her case, the open quotient also shows less variation before vowel offset and after vowel onset. The magnitude of the differences across consonantal contexts shows considerable variability between subjects. There does not appear to be any general effect of sentential stress on the open quotient. For some subjects, such as RM, the open quotient is larger during the second vowel, while for others, such as ES and LK, the opposite is true.

The mean area of the glottal pulse is plotted in Fig. 6. Before the offset of the vowel, the area decreases before the glottal stop. After vowel onset, the area increases after the glottal stop. In the other consonantal contexts, the patterns of change differ considerably between subjects. The area of the pulse is always larger for the male than for the female voices. No consistent effect of sentential stress is evident, most likely because this measure is affected by several factors, such as the peak flow and the fundamental frequency.

The fundamental frequency of a vowel has been found to vary systematically as a function of the voicing status of the preceding consonant, e.g. (Hombert et al., 1979; Ohde, 1984; Löfqvist et al., 1989). In particular, $F0$ has commonly been reported to be higher following a voiceless consonant. The results for mean $F0$ obtained in the present study are summarized in Fig. 7. Before the offset of the preceding vowel, $F0$ decreases in most consonantal contexts, though the change is quite small. In two subjects, CS and LZ, there is a marked decrease in F0 before the glottal stop, most likely reflecting a change to a creaky mode of phonation going into the stop. Note, however, that the opposite pattern occurs for subject ES. After vowel onset, $F0$ is generally higher following voiceless consonant and shows a falling pattern. After the voiced consonant, the fundamental frequency is lower and rising or almost constant. As might be expected, the female voices generally have a higher fundamental frequency than the male voices. This difference is particularly evident during the first, unstressed, vowel. Note however, that during the second, stressed, vowel, subject ES shows a higher $F0$ than subject LK. The influence of sentential stress on the second vowel is evident in higher values of $F0$.

4. Discussion

The results of this study generally agree with those reported by Löfqvist and McGowan (1992), as well as with those obtained in other studies referred to in the introduction. While thus corroborating the earlier results, the present study also extends them by showing both consistency among and variability across subjects. The largest variations in the measured parameters are found at the transitions between vowels and voiceless consonants, and between voiceless consonants and vowels. Also, the glottal stop often shows a substantial influence on the pulse properties. The variations are generally smaller in the context of voiced consonants, in particular /m/. As we have previously argued in more detail, analysis of simultaneous flow, pressure and transillumination records (Löfqvist and McGowan, 1992) suggests that the greater influence of voiceless consonants is due to the requisite glottal abduction gesture (for further physiological data, see also (Pétursson, 1976; Löfqvist and Yoshioka, 1980, 1981, 1984; Yoshioka et al., 1981, 1982)). In addition, different phasing patterns between the glottal gesture and the oral articulatory movements making the closure/constriction appear to be responsible for variations in pulse properties across voiceless consonantal contexts. We shall now examine these patterns in more detail.

For the aspirated stop /p/, the onset of the glottal abduction occurs near the onset of the labial closure; its exact location appears to differ even between speakers of the same language, cf. (Löfqvist and Yoshioka, 1984). Thus, the transition from the vowel to the consonant is characterized by a small increase in the open quotient and also in peak and minimum flow before the flow is shut off by the oral closure. At the release of the oral closure, the glottis is open and in the process of being adducted for the following vowel. The glottal vibrations begin while the glottal area is decreasing, and pulses at vowel onset have large values of peak flow, minimum flow, and the open quotient. During the first part of the vowel, the values of these three parameters decrease.

For the voiced stop /b/, there may be no specific glottal adjustments, cf. (Lisker and Baer, 1984). The glottal vibrations may continue during the oral closure or gradually cease, depending on whether there is sufficient transglottal pressure to sustain them. Since glottal vibrations during the stop closure are not a necessary feature of American English stops, e.g. (Lisker, 1986), we may suppose that speakers vary. Such variability may partly explain why subject LZ shows little difference in peak flow and open quotient between the voiced and voiceless stops /b, p/. In fact, a closer examination of all the individual productions revealed that subject LZ had more cases of voice breaks during her /b/ productions than the other subjects.

The voiceless stop in the cluster /sp/ is unaspirated in American English. That is, the glottal vibrations start very shortly after the release of the oral closure. In such a cluster, the glottis opens at the fricative onset and begins to adduct near its offset. By the release of the stop closure, the glottis is in a position suitable for voicing, and the vibrations start as soon as a sufficient transglottal pressure has developed. The glottal conditions at the release of the aspirated stop /p/ and the unaspirated stop in the cluster /sp/ are thus quite different. These differences are reflected in the generally higher values of peak flow and open quotient after vowel onset following the aspirated stop /p/.

The production of the voiceless fricative /s/ requires a glottal opening gesture and the formation of a narrow constriction in the vocal tract. The glottis begins to open during the transition into the oral constriction for the fricative. Hence, the peak flow and the open quotient show increasing values before vowel offset. At the transition from the fricative into the following vowel, the glottal vibrations start when the glottis is in a more adducted position than is the case following the voiceless aspirated stop /p/. Peak flow and the open quotient thus tend to be higher following the stop than the fricative.

The glottal condition during the voiced fricative /v/ would appear to be similar to that for the voiced stop /b/. That is, there does not seem to be any particular laryngeal adjustment for this sound. Hence, the pattern of change before the offset of the preceding vowel is similar for the stop and the fricative. Subjects LK and LZ do, however, show a marked decrease in peak flow before the voiced fricative /v/, as well as an increase in the open quotient. We noted above for the /b/ productions of subject LZ that she had more voice breaks than the other subjects. This also proved to be the case for her /v/ productions.

The glottal sounds /h/ and /ʔ/ are produced without any specific supralaryngeal articulatory movements. However, at the laryngeal level, these two sounds are made with the opposite adjustments. For the fricative /h/, the vocal folds are abducted, while for the glottal stop /ʔ/ they are adducted more or less forcefully. This difference in laryngeal conditions is evidenced by opposing patterns of change for peak flow, minimum flow and the open quotient. For the laryngeal fricative, all these parameters increase before vowel offset and decrease after vowel onset. (Note that subject RM does show a very small change in peak flow following the /h/.) In the environment of the glottal stop, however, the same parameters decrease before vowel offset and increase after vowel onset.

The linguistic status of the glottal stop is debatable in American English. In the present study, we tried to have the subjects produce "natural" glottal stops at the onset of the vowel in a stressed syllable preceded by another vowel. The subjects were not explicitly instructed to produce a glottal stop, however. Both the glottal stop and the laryngeal fricative showed some variability across subjects. This is illustrated in Fig. 8 which plots the unfiltered flow signal for a representative single production of the sequences /aʔa/ and /aha/ by each of the six subjects. For the glottal stop, all subjects with the exception of VG show a cessation of voicing. For the laryngeal fricative, subjects ES and VG have continuous glottal vibrations, whereas for the other subjects there is a period of voicelessness. While the subjects differed in whether they produced a distinct glottal stop and in their use of a creaky mode of phonation, the laryngeal conditions at vowel onset following the glottal stop and the laryngeal fricative are clearly different. In particular, voicing is generally more breathy following the fricative, as is illustrated in Fig. 9. This figure shows DFT spectra after voicing onset of the unfiltered flow signals for subject LK in Fig. 8. Note, in particular, that the difference between the amplitude of the two first harmonics is larger following the laryn-

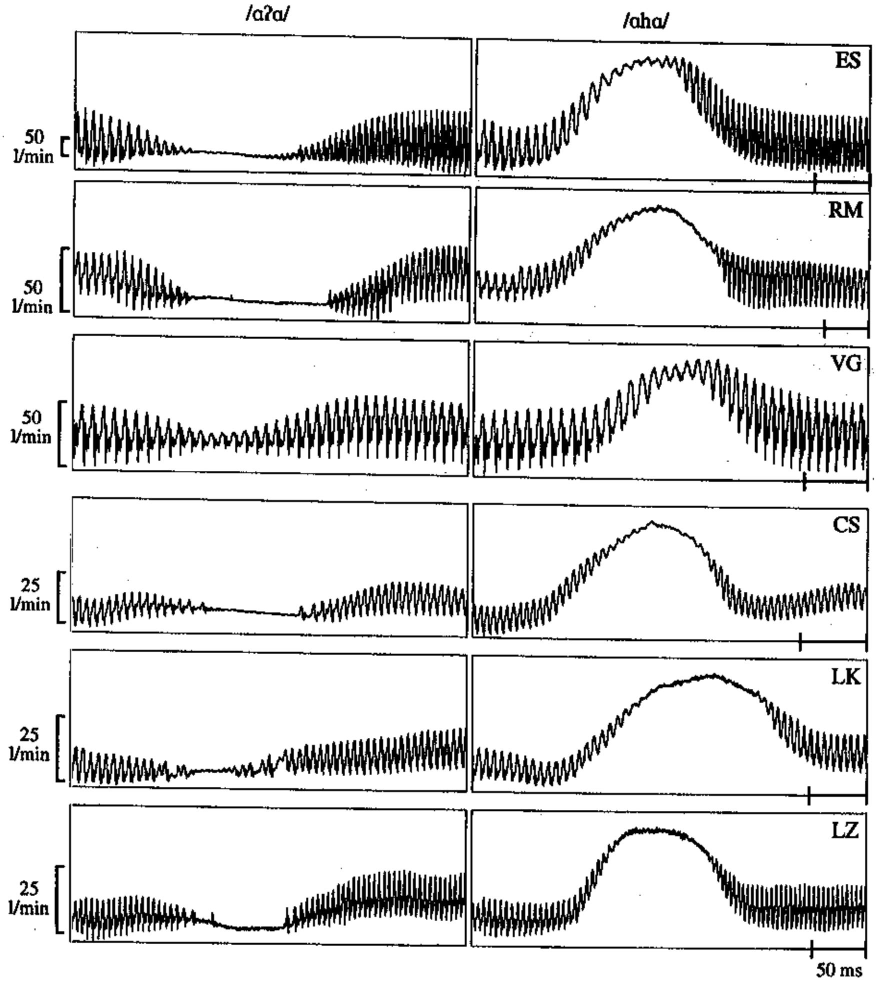

Fig. 8. Unfiltered flow signals for single productions of the sequences /aʔa/ and /aha/ by all subjects.

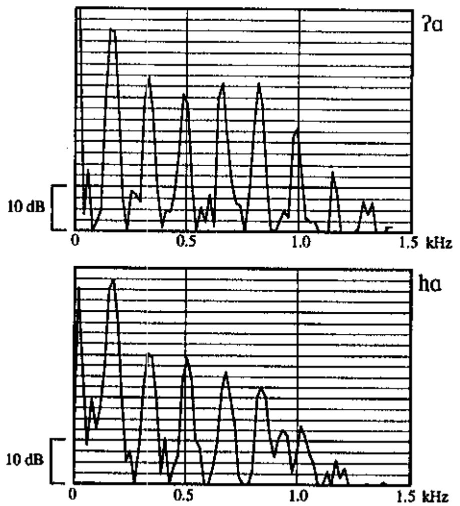

Fig. 9. Spectra at onset of voicing of the unfiltered flow signals following the glottal stop and the laryngeal fricative for the productions by subject LK shown in Fig. 8.

geal fricative /h/ than following the glottal stop. This difference can be used as an index of the tilt of the source spectrum. The greater tilt following the fricative is characteristic of a breathy type of phonation.

While there are individual differences in the magnitude of the changes of the different measures, many of the overall patterns are quite similar across speakers. The results for subject RM, however, show some exceptions that merit further discussion. Note that, in contrast to those of the other subjects, his data on peak flow in Fig. 3 show very little change following the laryngeal fricative /h/ and the voiceless consonants /p, s/. From the flow signals shown in Fig. 8, we can see that the glottal vibrations start later, at a lower flow value, for subject RM than for the other subjects in these consonantal contexts. Similarly, Fig. 5 shows that for subject RM the open quotient does not start high and then decrease dramatically following the same sounds /h, p, s/. We would hypothesize that this subject uses a considerable increase in the longitudinal tension of the vocal folds to suppress voicing, most likely using the cricothyroid muscle, cf. (Löfqvist et al., 1989). Further support in favor of this hypothesis can be found in the results for fundamental frequency plotted in Fig. 7. Note, in particular, that for subject RM the fundamental frequency after vowel onset is considerably higher than for the two other male subjects following the sounds /h, p, s/. Moreover, McGowan et al. (1995) show that when the Q-factor of the two-mass model is increased (see Fig. 3d in (McGowan et al., 1995)), the onset of glottal vibrations following the /h/ is delayed in a manner similar to that observed for subject RM in Fig. 8.

The variations in fundamental frequency observed in the present study generally agree with those found in other studies. They also indicate that subjects may vary both in the magnitude of the $F0$ difference following voiced and voiceless consonants and in the pattern of change. Subjects RM and LZ show large differences as a function of the voicing status of the preceding consonant, while for subjects VG and LK they are smaller. Following the voiced consonants, the fundamental frequency shows a rise for subjects ES, RM, LZ, but has a small fall for subject CS. The difference in fundamental frequency following voiced and voiceless consonants may remain at least 100 ms into the vowel. Even at this point in time, there is still a substantial difference between the two consonantal environments for subject LZ.

The measure of the open quotient that we have used here should more properly be seen as a characteristic of the waveform rather than as a reliable index to whether a complete closure of the glottis occurs or not. There is evidence from studies making direct observations of the glottis during phonation that a complete glottal closure often does not occur in normal voices, e.g. (Peppard et al., 1988; Biever and Bless, 1989; Södersten and Lindestad, 1990; Fex et al., 1991; Hertegård et al., 1992). In particular, an incomplete glottal closure appears to be more common among female subjects.

Studies comparing male and female voices have shown group differences in fundamental frequency, e.g. (Pegoraro Krook, 1988), and the open quotient (Holmberg et al., 1988). On the average, females exhibit a higher $F0$ and a larger open quotient than males. Acoustic studies also suggest that the female voice typically has more noise components and a higher tilt of the source spectrum than the male voice (Monsen and Engebretson, 1977; Nittrouer et al., 1990; Klatt and Klatt, 1990); see also (Price, 1989). The most reliable differences between the male and female voices in the present study were those of peak flow and the area of the pulse. Both were larger in the male voice. This is most likely due to the larger dimensions of the male larynx. We should add, however, that there are substantial variations within the male and female subjects for most of the measures used in the present study. Some of this variability may well be due to variations in the intensity level used by the individual subjects.

The present results show how the aerodynamic patterns at consonant-vowel and vowel-consonant transitions differ as a function of the consonant, in particular its voicing status. The extensive variations in air flow that occur before and after voiceless consonants reflect the coordination between the oral and laryngeal valves. In addition, data have suggested that subjects differ in their relative use of laryngeal abduction and longitudinal tension; some individual differences in flow patterns can be explained in terms of such differences in laryngeal setting. Modeling experiments reported by McGowan et al. (1995) yield variations comparable to those found in the present study. The modeling work also suggests that vocal tract wall impedance plays an important role in producing high air flows at the release of the closure/constriction for voiceless consonants. Hence, modeling allows us to separate the component factors and assists in interpreting the measurements. The experimental data presented here in combination with the modeling results presented by McGowan et al. (1995) illustrate the usefulness of an analysis-by-synthesis approach to the understanding of speech aerodynamics.

Acknowledgements

We are grateful to Louis Boves, University of Nijmegen, for providing us with a copy of the inverse filtering program developed there. This work was supported by Grants DC-00865, DC-01247 and DC-00121 from the National Institute on Deafness and Other Communication Disorders, and in part by Esprit-BR Project 6975 – Speech Maps through Grant P55 from the Swedish National Board for Industrial and Technical Development.

References

A.S. Abramson (1977), "Laryngeal timing in consonant distinctions", *Phonetica*, Vol. 34, pp. 295–303.

T. Ananthapadmanabha and G. Fant (1982), "Calculation of the true glottal flow and its components", *STL-QPSR*, 1/1982, pp. 1–30.

H.J. Arkebauer, T.J. Hixon and J.C. Hardy (1967), "Peak intraoral air pressures during speech", *J. Speech Hearing Res.*, Vol. 10, pp. 196–208.

P. Badin, S. Hertegård and I. Karlsson (1990), "Notes on the Rothenberg mask", *STL-QPSR*, 1/1990, pp. 1–7.

D. Biever and D. Bless (1989), "Vibratory characteristics of the vocal folds in young adult and geriatric women", *J. Voice*, Vol. 3, pp. 120–131.

B. Cranen and L. Boves (1987), "On subglottal formant analysis", *J. Acoust. Soc. Amer.*, Vol. 81, pp. 734–746.

J. De Veth, B. Cranen, H. Strik and L. Boves (1990), "Extraction of control parameters for the voice source in a text-to-speech system", *Internat. Conf. Acoust. Speech Signal Process. 1990*, pp. 301–304.

R.P. Dixit (1989), "Glottal gestures in Hindi plosives", *J. Phonetics*, Vol. 17, pp. 213–237.

G. Fant (1961), "The acoustics of speech", *Proc. Third Internat. Congress Acoustics, Stuttgart, 1959*, ed. by L. Cremer (Elsevier, Amsterdam); Reprinted in G. Fant (1973) *Speech Sounds and Features* (MIT Press, Cambridge), pp. 3–16.

G. Fant (1972), "Speech analysis and speech production: Subglottal formants", *STL-QPSR*, 1/1972, pp. 1–12.

G. Fant (1986), "Glottal flow: Models and interaction", *J. Phonetics*, Vol. 14, pp. 393–399.

G. Fant, J. Liljencrants and Q. Lin (1985), "A four-parameter model of glottal flow", *STL-QPSR*, 4/1985, pp. 1–13.

S. Fex, A. Löfqvist and L. Schalén (1991), "Videostroboscopic evaluation of glottal open quotient, related to some acoustic parameters", in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, San Diego), pp. 273–278.

J. Gauffin and J. Sundberg (1989), "Spectral correlates of glottal voice source waveform characteristics", *J. Speech Hearing Res.*, Vol. 32, pp. 556–565.

C. Gobl (1988), "Voice source dynamics in connected speech", *STL-QPSR*, 1/1988, pp. 123–159.

C. Gobl and A. Ní Chasaide (1988), "The effect of adjacent voiced/voiceless consonants on the vowel voice source: A cross-language study", *STL-QPSR*, 2–3/1988, pp. 23–59.

S. Hertegård, J. Gauffin and I. Karlsson (1992), "Physiological correlates of inverse filtered waveforms", *J. Voice*, Vol. 6, pp. 224–234.

E. Holmberg, R. Hillman and J. Perkell (1988), "Glottal airflow and transglottal air-pressure measurements for male and female speakers in soft, normal, and loud voice", *J. Acoust. Soc. Amer.*, Vol. 84, pp. 511–529.

J.-M. Hombert, J. Ohala and W. Ewan (1979), "Phonetic explanations for the development of tones", *Language*, Vol. 55, pp. 37–58.

N. Isshiki and R. Ringel (1964), "Airflow during the production of selected consonants", *J. Speech Hearing Res.*, Vol. 7, pp. 233–244.

D. Klatt and L. Klatt (1990), "Analysis, synthesis and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Amer.*, Vol. 87, pp. 820–857.

D. Klatt, K. Stevens and J. Mead (1968), "Studies of articulatory activity and airflow during speech", *Ann. New York Acad. Sci.*, Vol. 155, pp. 42–55.

J. Koreman, L. Boves and B. Cranen (1992), "The influence of linguistic variations on the voice source characteristics", *Proc. 1992 Internat. Conf. Spoken Language Processing*, ed. by J. Ohala, T. Nearey, B. Derwing, M. Hodge and G. Wiebe (Univ. of Alberta, Edmonton), Vol. 1, pp. 125–128.

A. Krishnamurthy and D. Childers (1986), "Two-channel speech analysis", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 34, pp. 730–743.

L. Lisker (1986), "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees", *Language and Speech*, Vol. 29, pp. 3–11.

L. Lisker and T. Baer (1984), "Laryngeal management at utterance-internal word boundary in American English", *Language and Speech*, Vol. 27, pp. 163–171.

A. Löfqvist (1980), "Interarticulator programming in stop production", *J. Phonetics*, Vol. 8, pp. 475–490.

A. Löfqvist (1992), "Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants", *Language and Speech*, Vol. 35, pp. 15–28.

A. Löfqvist and R.S. McGowan (1992), "Influence of consonantal environment on voice source aerodynamics", *J. Phonetics*, Vol. 20, pp. 93–110.

A. Löfqvist and H. Yoshioka (1980), "Laryngeal activity in Swedish obstruent clusters", *J. Acoust. Soc. Amer.*, Vol. 68, pp. 792–801.

A. Löfqvist and H. Yoshioka (1981), "Laryngeal activity in Icelandic obstruent production", *Nordic J. Linguistics*, Vol. 4, pp. 1–18.

A. Löfqvist and H. Yoshioka (1984), "Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production", *Speech Communication*, Vol. 3, No. 4, pp. 279–289.

A. Löfqvist, T. Baer, N.S. McGarr and R. Seider Story (1989), "The cricothyroid muscle in voicing control", *J. Acoust. Soc. Amer.*, Vol. 85, pp. 1314–1321.

R.S. McGowan, L.L. Koenig and A. Löfqvist (1995), "Vocal tract aerodynamics in /aCa/ utterances: Simulations", *Speech Communication*, Vol. 16, No. 1, pp. 67–88.

R. Monsen and A. Engebretson (1977), "A study of variations in the male and female glottal wave", *J. Acoust. Soc. Amer.*, Vol. 62, pp. 981–993.

S. Nittrouer, R.S. McGowan, P.H. Milenkovic and D. Beehler (1990), "Acoustic measurements of men's and women's voices: A study of context effects and covariation", *J. Speech Hearing Res.*, Vol. 33, pp. 761–775.

R. Ohde (1984), "Fundamental frequency as an acoustic correlate of stop consonant voicing", *J. Acoust. Soc. Amer.*, Vol. 75, pp. 224–230.

S. Palmer and J. House (1992), "Dynamic voice source changes in natural and synthetic speech", *Proc. 1992 Internat. Conf. Spoken Language Processing*, ed. by J. Ohala, T. Nearey, B. Derwing, M. Hodge and G. Wiebe (University of Alberta, Edmonton), Vol. 1, pp. 129–132.

M. Pegoraro Krook (1988), "Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis", *Folia Phoniat.*, Vol. 40, pp. 82–90.

R. Peppard, D. Bless and P. Milenkovic (1988), "Comparison of young adult singers and nonsingers with vocal nodules", *J. Voice*, Vol. 2, pp. 250–260.

M. Pétursson (1976), "Jointure au niveau glottal", *Phonetica*, Vol. 35, pp. 65–85.

J. Pierrehumbert and D. Talkin (1992), "Lenition of /h/ and glottal stop", *Papers in Laboratory Phonology II*, ed. by G. Docherty and D. Ladd (Cambridge Univ. Press, Cambridge), pp. 90–117.

P. Price (1989), "Male and female source characteristics: Inverse filtering results", *Speech Communication*, Vol. 8, No. 3, pp. 261–277.

L. Rabiner, B. Atal and M. Sambur (1977), "LPC prediction error-analysis of variation with the position of the analysis frame", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 25, pp. 434–442.

M. Rothenberg (1973), "A new inverse filtering technique for deriving the glottal airflow waveform during speech", *J. Acoust. Soc. Amer.*, Vol. 53, pp. 1632–1645.

C. Scully, E. Castelli, E. Brearly and M. Shirt (1992a), "Articulatory paths and aerodynamic patterns for some fricatives", *J. Phonetics*, Vol. 20, pp. 39–51.

C. Scully, E. Grabe-Georges and E. Castelli (1992b), "Articulatory paths for some fricatives in connected speech", *Speech Communication*, Vol. 11, Nos. 4–5, pp. 411–416.

M. Södersten and P.-Å. Lindestad (1990), "Glottal closure and perceived breathiness during phonation in normally speaking subjects", *J. Speech Hearing Res.*, Vol. 33, pp. 601–611.

J. Subtelny, J. Worth and M. Sakuda (1966), "Intraoral pressure and rate of flow during speech", *J. Speech Hearing Res.*, Vol. 9, pp. 498–518.

H. Yoshioka, A. Löfqvist and H. Hirose (1981), "Laryngeal adjustments in the production of consonant clusters and geminates in American English", *J. Acoust. Soc. Amer.*, Vol. 70, pp. 1615–1623.

H. Yoshioka, A. Löfqvist and R. Collier (1982), "Laryngeal adjustments in Dutch voiceless obstruent production", *Ann. Bull.* (Res. Inst. Logopedics and Phoniatrics, Univ. of Tokyo), Vol. 16, pp. 27–35.