

Reprinted from

# SPEECH COMMUNICATION

---

Speech Communication 16 (1995) 67-88

## Vocal tract aerodynamics in /aCa/ utterances: Simulations

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

*Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, USA*

Received 3 May 1994; revised 17 October 1994

## SPEECH COMMUNICATION

**Founding Editor**  
(1981-1993)  
Max Wajskop †

**Editor in Chief**  
Christel Sorin  
CNET LAA/TSS/RCP  
2 Route de Trégastel  
22301 Lannion, France  
Tel: (33) 96 05 33 06  
Fax: (33) 96 05 35 30  
E-mail: sorin@lannion.cnet.fr

**Assistant Editor**  
Laurent Miclet  
(same address)  
E-mail: miclet@lannion.cnet.fr

**Advisory Committee**  
S. Furui (Tokyo, Japan)  
B. Juang (Murray Hill, NJ, USA)

M. Kunt (Lausanne, Switzerland; EURASIP)  
M. Liberman (Philadelphia, PA, USA)  
J. Ohala (Berkeley, CA, USA)  
L. Pols (Amsterdam, The Netherlands; ESCA)

**Editorial Board**  
R. Billi (Torino, Italy)  
L.J. Boë (St Martin d'Hères, France)  
L. Boves (Nijmegen, The Netherlands)  
R. Carlson (Stockholm, Sweden)  
C.J. Darwin (Brighton, UK)  
B. De Boysson-Bardies (Paris, France)  
R. De Mori (Montréal, Canada)  
B. Delgutte (Boston, MA, USA)  
G. Fant (Stockholm, Sweden)  
J. Flanagan (Piscataway, NJ, USA)  
J.E. Flege (Birmingham, AL, USA)  
H. Fujisaki (Chiba, Japan)  
J.P. Haton (Vandoeuvre, France)  
J. Hirschberg (Murray Hill, NJ, USA)  
M. Hunt (Cheltenham, UK)  
F. Jelinek (Baltimore, MD, USA)

J.C. Junqua (Santa Barbara, CA, USA)  
J. Laver (Edinburgh, UK)  
M. Lennig (Montréal, Canada)  
B. Lindblom (Austin, TX, USA)  
J. Makhoul (Cambridge, MA, USA)  
J. Mariani (Orsay, France)  
J.B. Millar (Canberra, Australia)  
E. Moulines (Paris, France)  
H. Ney (Aachen, Germany)  
L.A. Petitto (Montreal, Canada)  
P.J. Price (Menlo Park, CA, USA)  
Y. Sagisaka (Kyoto, Japan)  
J. Schoentgen (Brussels, Belgium)  
S. Seneff (Cambridge, MA, USA)  
J. Spitz (White Plains, NY, USA)  
Y. Tohkura (Kyoto, Japan)  
I. Trancoso (Lisbon, Portugal)  
J.P. Tubach (Paris, France)  
D. Van Compernelle (Heverlee, Belgium)  
H. Wakita (Santa Barbara, CA, USA)  
C. Wellekens (Sophia Antipolis, France)  
A.C. Woodbury (Austin, TX, USA)

**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 calendar 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 date 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.

© 1995 Elsevier Science B.V. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Elsevier Science B.V., Copyright and Permissions Department, P.O. Box 521, 1000 AM Amsterdam, The Netherlands.

**Special regulations for authors:** Upon acceptance of an article by the journal, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.

**Special regulations for readers in the USA:** This journal has been registered with the Copyright Clearance Center, Inc. Consent is given for copying of articles for personal or internal use, or for the personal use of specific clients. This consent is given on the condition that the copier pays through the Center the per-copy fee stated in the code on the first page of each article for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. The appropriate fee should be forwarded with a copy of the first page of the article to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. If no code appears in an article, the author has not given broad consent to copy and permission to copy must be obtained directly from the author. The fee indicated on the first page of an article in this issue will apply retroactively to all articles published in the journal, regardless of the year of publication. This consent does not extend to other kinds of copying, such as for general distribution, resale, advertising and promotion purposes, or for creating new collective works. Special written permission must be obtained from the publisher for such copying.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Although all advertising material is expected to conform to ethical standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality of such product or of the claims made of it by its manufacturer.

© The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).



ELSEVIER

Speech Communication 16 (1995) 67-88

**SPEECH**  
COMMUNICATION

## Vocal tract aerodynamics in /aCa/ utterances: Simulations

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

*Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, USA*

Received 3 May 1994; revised 17 October 1994

---

### Abstract

Aerodynamic simulations of /aCa/ utterances were made using a low-frequency model for upper vocal tract airflow and a two-mass model for the voice source. These simulations helped increase insight into the results of an empirical study of flow during running speech. The various sources of flow, including wall compliance, were examined for their contributions to total flow from the mouth. The two-mass model was modified to allow for more natural glottal flow during abduction and adduction. Even with modifications the two-mass model was not sufficient to model source variations during running speech.

### Zusammenfassung

Es wurden aerodynamische Simulationen von /aCa/ Spracheinheiten durchgeführt, die ein Niedrigfrequenzmodell für den Luftstrom des oberen Vokaltrakts und ein 2-Masse-Modell für die Stimmerzeugungsquelle verwendet. Diese Simulationen helfen die Resultate einer empirischen Studie über den Luftstrom bei kontinuierlicher Sprache besser zu verstehen. Die verschiedenen Quellen des Luftstroms, einschließlich der Nachgiebigkeit des Mundraums, wurden auf ihre Beteiligung an dem gesamten Luftstrom aus dem Mund hin untersucht. Das 2-Masse-Modell wurde modifiziert, um einen natürlicheren Glottisfluß während der Abduktion und der Adduktion zu ermöglichen. Aber selbst mit diesen Modifikationen war das 2-Masse-Modell nicht ausreichend, um die Variationen der Stimmerzeugungsquelle bei kontinuierlicher Sprache zu modellieren.

### Résumé

On a réalisé des simulations aérodynamiques de séquences /aCa/ à l'aide d'un modèle basse-fréquence pour l'écoulement de l'air dans la partie supérieure du conduit vocal et d'un modèle à deux masses pour la source. Ces simulations nous ont permis de mieux comprendre les résultats d'une étude empirique sur l'écoulement lors de la parole continue. On a examiné la contribution des diverses sources d'écoulement, y compris la compliance des

---

\* Corresponding author. Tel.: (203) 865-6163; Fax: (203) 865-8963; E-mail: MCGOWAN@HASKINS.YALE.EDU.

parois, à l'écoulement buccal global. Le modèle à deux masses a été modifié pour permettre un écoulement glottique plus naturel pendant l'abduction et l'adduction. Malgré ces modifications, le modèle à deux masses s'est avéré insuffisant pour modéliser les variations de source en parole continue.

*Keywords:* Aerodynamics; Voice; Running speech

## 1. Introduction

### 1.1. Purpose and general method

To gain further insight into the results of Löfqvist, Koenig and McGowan (Löfqvist et al., 1995), a numerical simulation of the voice source and upper vocal tract aerodynamics was made. Löfqvist et al. (1995) used a pneumotacograph (the Rothenberg mask) to record oral-nasal flow. This device measures a bandlimited oral-nasal flow that has contributions from glottal flow as well as articulatory movement, the passive movements of the vocal tract walls and other flows. Therefore, even with the acoustic resonances of the vocal tract removed from the signal, the measured flow does not directly correspond to the glottal flow. One purpose of this work was to find the magnitudes of the various component flows that contribute to the flow measured by the mask during running speech. Another purpose of this work was to infer laryngeal and supralaryngeal activity during running speech. Also, the adequacy of the simulation model itself for use in synthesis of running speech was assessed. The adequacy of the simulations depends on which aspects of laryngeal activity, air flow, and the interaction between them are captured in the model.

A slightly modified version of the two-mass model (Ishizaka and Matsudaira, 1968; Ishizaka and Flanagan, 1972) was coupled with a low-frequency model of upper vocal tract aerodynamics to enable an analysis-by-synthesis of the experimental data obtained in Löfqvist et al. (1995). The term low frequency in reference to aerodynamic simulation means that lumped impedance elements for the upper vocal tract were used. Therefore, the validity of upper vocal tract flow analyses was restricted to frequencies of less than 100 Hz. (At 100 Hz, the wavelength of sound is

20 times that of a 17 cm vocal tract, so this may be considered a good upper bound on the frequency range of the flow model described here.) In a lumped-element analysis, individual pulses of the voice are not propagated in the supralaryngeal tract, but rather what is sometimes referred to as "D.C." flow or low-frequency aerodynamics is modeled. To simulate the low-frequency aerodynamics produced by a given utterance and its articulation, it is necessary to simulate the evolution of the vocal tract area function and supralaryngeal volume. The area function and the volume of the vocal tract between the glottis and any supralaryngeal constriction were simulated with the movement of the articulators in the vocal tract used in the Haskins Laboratories articulatory synthesizer, ASY (Rubin et al., 1981). (It should be noted that the flow simulations were not used to generate sound in ASY.)

Direct comparisons of the inverse-filtered measured oral-nasal flow and the simulated flow from the mouth cannot be made. The inverse-filtered measured oral-nasal flow, although bandlimited and lacking formant oscillations, does contain information on the shape of individual glottal pulses. Comparisons can generally be made between the smoothed measured flow and the simulated low-frequency mouth flow, with the further assumptions of no nasal leakage and no volume changes between the face and the mask during the experimental recordings. The smoothed measured flow traces are similar in type to those found in (Klatt et al., 1968). An exception in this comparability of the two flows is within the first two milliseconds of formation or release of a tight constriction, when flow with high frequency content is generated but its propagation is not modeled validly by a low-frequency aerodynamic model. Further, the inverse-filtered pulses and the simulated glottal pulses from the

two-mass model cannot, strictly, be compared. Aerodynamic effects, such as the inflation of the supralaryngeal tract when oral constrictions are formed, were not subtracted out in the processing of the measured flow signal. Even if the upper vocal tract aerodynamics were known and subtracted out, other factors prohibit a strict comparison between the inverse-filtered measured oral flow and the simulated glottal flow. One of the problems is that the measured flow contains effects of the upper vocal tract on the glottal pulse, even after inverse filtering. An example would be the pulse skewing caused by the inductive load of the vocal tract when the first formant frequency is above the fundamental (Rothenberg, 1983). Despite these limitations, there are useful inferences that can be made on the relative timing and magnitudes of articulatory events from comparisons of smoothed measured flow and the simulations. Further, the effects of aerodynamic quantities on the maintenance and initiation of voice during running speech can be studied using simulations.

### *1.2. Aerodynamic simulation*

There has been previous work on modeling low-frequency aerodynamics in the vocal tract using lumped elements (Rothenberg, 1968; Ohala, 1974; Müller and Brown, 1980; Westbury, 1983; Bickley and Stevens, 1986; Scully, 1990; Bickley, 1991). These models are of varying degrees of sophistication, but all take into account constriction resistance in the upper vocal tract, glottal resistance and vocal-tract wall compliance. More inclusive work would simulate low-frequency aerodynamic phenomena along with acoustic wave propagation – a high-frequency aerodynamic phenomenon that requires a more sophisticated model than one composed of lumped elements. Sondhi and Schroeter's (1987) is a model that might be used for such a purpose. However, with the current state of knowledge, much can be learned by just simulating low-frequency aerodynamics. Once this kind of model is understood, a high-frequency simulation can be incorporated.

One of the aspects of aerodynamic modeling with which we were concerned was that of

source-tract interaction. Recent modeling works on source-tract interaction can be grouped according to the aspects of interaction for which they account. Some studies use a parametric model of the glottal source, while others account for the mechanics of laryngeal vibration. The latter works directly simulate source-tract interaction from physical principles, including tract effects on the vibratory patterns of the folds, e.g. (Ishizaka and Flanagan, 1972; Bickley and Stevens, 1986; Stevens, 1991; Bickley, 1991), while the former account for the vocal tract's effect on the volume velocity pulse from the glottis given a glottal area function, e.g. (Ananthapadmanabha and Fant, 1982; Rothenberg, 1983; Lin, 1990). Both types attempt to model differences in the glottal volume velocity waveform in the case that the ratio of upper vocal tract impedance to glottal impedance is nonzero, compared to the idealized case where the ratio is zero. Another distinction that can be made among source-tract interaction models is the same distinction used above for flow modeling, between low-frequency aerodynamics and high-frequency aerodynamics. Examples of studies of source-tract interaction resulting from low-frequency aerodynamics are (Bickley and Stevens, 1986; Scully, 1990; Bickley, 1991; Stevens, 1991). Studies of high-frequency aerodynamic, or acoustic, effects on the voice source can be found in (Ishizaka and Flanagan, 1972; Flanagan et al., 1975; Ananthapadmanabha and Fant, 1982; Rothenberg, 1983; Koizumi et al., 1987).

In the present work, a low-frequency aerodynamic simulation and a mechanical model of the vocal folds, the two-mass model, were used to simulate flow in /aCa/ utterances. Thus, source-tract interaction was restricted to the effect that low-frequency intraoral pressure has on the voice source, and vice-versa. The work presented here is most closely related to that of Bickley and Stevens (1986) (see also (Bickley, 1991; Stevens, 1991)), but extends their studies by allowing the articulatory parameters, including those of the two-mass model, to vary with time. The only other simulation of low-frequency vocal tract aerodynamics incorporating a voice source that is known to the authors is that of Scully (1990). The major difference between Scully's

model and the one described in this paper is that Scully used a parametric model of the voice source volume velocity.

Because time-varying situations were studied here, some fundamental assumptions in using a lumped-element model for low-frequency aerodynamics need to be made, beyond those noted in the introduction. One is that the flow at the glottis is signaled at the mouth with a speed of about the speed of sound. Further, the time-of-travel from the glottis to the mouth, based on the speed of sound, is supposed to be negligible compared to other time scales, such as that of articulatory movement. (This assumption is unwarranted at the moment when tight constrictions are broken and there is a sudden change from an infinite to finite resistance.) Signals, therefore, were presumed to be transmitted at an infinite speed. However, approximation of infinite travel speed can depend on the assumption that the flow is potential flow and irrotational. There are experiments that call this assumption into question (see (Davies et al., 1993) for references). In fact, flow sensed at the mouth that has traveled from the glottis in a rotational mode may have done so at a speed comparable to the convection speed (usually, at least two orders of magnitude less than the speed of sound). This could mean that the measured flow at the mouth represents the summation of flows originating at the glottis at very disparate times. It can only be noted that potential flow was assumed in the low-frequency model and that the convective effects on rotational air motion were neglected. It was hoped that this assumption was sufficient for the purpose of simulating the results of the running speech experiments. Accounting for rotational flow would have required a field-theoretic fluid dynamic simulation based on the Navier–Stokes equations.

Default values for the size of the vocal tract and the dimensions of the two-mass model were used (Mermelstein, 1973; Ishizaka and Flanagan, 1972). These values are more likely to correspond closely to adult males than to adult females. This is not critical for the present work however, because we were interested in comparing the simulated low-frequency flow from the mouth with the

smoothed, measured flow. (As will be explained in Section 2, smoothing removed the individual pulses.) As discussed above, comparisons between simulated glottal flow from the two-mass model and measured flow were not strictly valid. Thus, only the grosser trends of voice were mimicked, such as declining pulse amplitude upon abduction, and no attempt was made to mimic the finer features of pulse shape, such as open quotient or absolute pulse amplitude. Also, because there were many uncertainties in the estimation of such quantities as volume between a constriction and the glottis for the calculation of volume flow due to movement of the articulators, and the fact that glottal area inaccuracies could be compensated for with subglottal pressure, these anatomical factors were not adjusted when simulations were compared to female subjects. In fact, we used a female subject, CS, from Löfqvist et al. (1995) to guide us in setting the timing of articulators.

## 2. Procedure

The low-frequency aerodynamic model used in conjunction with ASY and the two-mass model, and its lumped-element electrical analogue, are shown in Fig. 1. The subglottal pressure is denoted  $P_S$ , and the intraoral pressure is denoted  $P_{IO}$ . In the electrical analogue these specify voltages. The volume flows are represented by the symbol  $U$ , with the subscript G denoting glottal flow, the subscript W denoting flow due to wall movement, the subscript A denoting flow due to the compliance of air, and the subscript M denoting flow from the mouth. The volume velocities become currents in the electrical analogue, and the relations amongst them will be discussed further in the results. The electrical resistance,  $R_W$ , represents the lumped viscous resistance of the walls, while the capacitance,  $C_W$ , and inductance,  $L_W$ , represent the lumped compliance and mass of the walls, respectively. The glottal resistance,  $R_G$ , was modeled using the formulas in (Ishizaka and Flanagan, 1972, pp. 1241–1242) for a two-mass model with a rectangular glottis and with the areas between the upper and lower masses

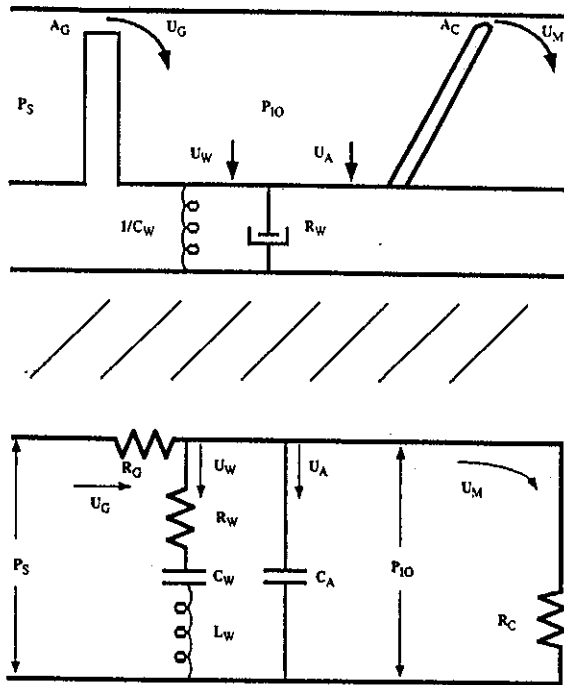


Fig. 1. Low-frequency flow model.

equal to  $A_G$ . The glottal slit was assumed to be 1.2 cm long in the anterior-posterior dimension, and the lower masses 0.25 cm long and the upper masses 0.05 cm long in the axial dimension. The vocal tract area just downstream of the glottis was derived from the section of ASY closest to the larynx. The supralaryngeal constriction resistance,  $R_C$ , was modeled assuming a circular opening of area  $A_C$ , so that the relation between the intraoral pressure,  $P_{IO}$ , and the flow through the mouth,  $U_M$ , was given by the formulas of Stevens in his model of fricative production with a circular constriction (Stevens, 1971, p. 1182).

Supraglottal area functions, necessary for the calculation of flow resistance, were derived from the ASY vocal tract. The coordinates of the various ASY articulators, such as those of the tongue body, were given new values every 10 ms, and these values were averaged using a 40 ms window. Using these averaged coordinate values, area functions were obtained every 10 ms. The averaging was done so that step changes would not be input at the 10 ms boundaries of the coordinate

values. These abrupt changes would have constituted high-frequency inputs to a model valid only at low frequencies. It turned out that this smoothing was not sufficient in the region very near a constriction release. Also, it was necessary to control the minimum constriction area,  $A_C$ , explicitly near fricatives to obtain precision in small constrictions. Thus,  $A_C$  was specified every 10 ms and averaged over 40 ms intervals during fricative articulations. The wall inductance, wall capacitance and wall resistance were all specified in 10 ms intervals, although these rarely changed.

Parameters specifically for control of the two-mass model and the low-frequency aerodynamics were also given every 10 ms. The subglottal pressure,  $P_S$ , and the Q-factor, QF, a generalized tension parameter, were used in controlling the two-mass model (Ishizaka and Flanagan, 1972). During simulated abduction, it was found that the volume velocity oscillation amplitudes would grow, which was contrary to observed flow measurements. To correct this, it was necessary to modify the two-mass model to simulate the third, anterior-posterior dimension, with increasing rest area from anterior to posterior (Fig. 2). This was accomplished by allowing for two different regions in the anterior-posterior dimension. The anterior region was modeled using the two mass model, with the masses reduced according to the proportion of the length of the folds in the posterior region. The posterior region simply provided a channel for flow, as there was no vibration allowed in the posterior. The anterior region was assigned a rest area (the same for both upper and lower masses), and the posterior region was assigned another area. The lateral rest distances between the anterior masses and the lateral distances between the walls were not altered. Therefore, there were three parameters to specify for glottal rest area: glottal rest area if the entire length of the folds was included in the anterior region, rest area if the entire length of the folds was included in the posterior region, and the actual proportion of the entire length of the folds contained in the posterior region. The folds were moved from a voicing posture to a completely abducted posture by changing the proportion of posterior-to-total length from zero to one (0%

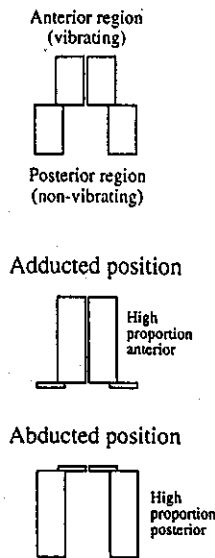


Fig. 2. Superior view of the modified two-mass model.

posterior was the default voicing condition, and 100% posterior was completely abducted). Generally, the equilibrium separation between the masses was left untouched, with exceptions to be noted later. Excluding these exceptions, if all the length of the folds was included in the anterior region, then the rest area between both the upper and lower masses was  $0.03 \text{ cm}^2$  (appropriate for voicing), and if all the length of the folds was included in the posterior region, the area of the channel was  $0.5 \text{ cm}^2$  (more appropriate for maximum abduction in voiceless consonants). The glottal rest area,  $TMA_G$ , was defined to be  $0.03 \text{ cm}^2$  times the proportion of the folds in the anterior region, plus  $0.5 \text{ cm}^2$  times the proportion of the folds in the posterior region. To ensure numerical stability, there was a constant 0.005 proportion of the length of folds in the posterior region during vowel portions when the folds were adducted for voicing. With this rudimentary modeling of the third dimension of the larynx, it was possible to abduct the folds without increasing the amplitude of the volume velocity oscillations. This was possible because the lateral rest distances between the masses in the anterior region

did not actually change during abduction. The idea of using variation in glottal rest area in the anterior-posterior dimension has been used previously in other contexts (Stevens, Personal communication; Cranen and Boyes, 1987).

For each utterance that was recorded experimentally in Löfqvist et al. (1995), except for /ama/, a simulation was performed. The utterances of a single subject (CS, female) were chosen for comparison between the data and the simulations. The simulated volume flow through the mouth and the volume flow measured by the Rothenberg mask, smoothed so that individual pulses propagated from the glottis were no longer visible, were compared. The smoothing of the measured flow was accomplished by applying a 101-point triangular window to the recorded flow signal. Such a window provides a low-pass capability that is 6 dB down at about 120 Hz for a sampling rate of 10 kHz.

Parameter settings were chosen to provide the best fit of simulation results to CS's smoothed flow data. Similar tongue body movements were used in all the utterances, with a slight change in vowel quality between the preceding and following vowel in keeping with the stress differences in the recorded utterances. In particular, the stressed second vowel /a/ was given a low jaw position and retracted tongue body; in the unstressed initial vowel /a/ the lowering and retraction were slightly less extreme for a more neutral articulation. Unless otherwise specified, the subglottal pressure during vowel production was set to  $7000 \text{ dy/cm}^2$  (approximately 7 cm  $\text{H}_2\text{O}$ ), and the Q-factor of the two-mass model was set to 1. The subglottal pressure was slightly lower than the standard 8 cm  $\text{H}_2\text{O}$  used in many modeling studies, but the magnitudes of the observed flows seemed to call for a reduced subglottal pressure. The wall impedance parameters were taken from (Westbury, 1983) for the tense wall condition, with wall capacitance,  $C_w$ , of  $2.5458 \times 10^{-4} \text{ cm}^5/\text{dy}$ , wall resistance,  $R_w$ , of  $18.56 \text{ dy-s/cm}^5$ , and wall inductance,  $L_w$ , of  $1.92 \times 10^{-2} \text{ dy-s}^2/\text{cm}^5$ . The tense wall condition was chosen because the laxer conditions provided too much capacitance, which resulted in too much mouth flow after stop releases for subject CS.



To provide an interface between the two-mass model and the low-frequency aerodynamics, an iterative procedure was designed so that the glottal flow into the low-frequency circuit matched that of the time-averaged glottal flow from the two-mass model. This procedure was necessary because intraoral pressure, which was determined by the low-frequency flow simulation, was a parameter used by the two-mass model. In the iterative procedure, the glottal orifice area of the low-frequency aerodynamic model,  $A_G$ , was adjusted so that the time-averaged glottal flows from the two models matched either within 40 cm<sup>3</sup>/s or 4%, whichever was least restrictive. After appropriate glottal areas were determined at 10 ms intervals, the entire simulation was run again, this time smoothing the glottal area function for the low-frequency aerodynamic model.

### 3. Results

The utterances can be divided into three groups for modeling purposes: those that involve little upper articulator movement, but substantial laryngeal adjustment (/aha/ and /a<sup>2</sup>a/), those that involve little laryngeal adjustment, but substantial upper articulator movement (/aba/ and /ava/), and those that require both substantial laryngeal adjustment and upper articulator movement (/asa/, /aspa/ and /apa/). The utterances will be considered in this order, noting highlights in the control of the upper articulatory, laryngeal and subglottal pressure parameters.

Beyond the mouth volume velocity ( $U_M$ ), the following simulation results are shown, when appropriate: glottal volume velocity ( $U_G$ ), volume velocity of the vocal tract walls ( $U_W$ ), volume velocity into the vocal tract caused by articulatory movement ( $U_{ARTIC}$ ), and intraoral pressure ( $P_{IO}$ ). (Whenever reference is made to a simulated quantity, it is assumed to be low frequency.) Note that positive values for  $U_M$ ,  $U_G$ ,  $U_{ARTIC}$  and  $U_W$  denote constriction flow out of the mouth, glottal flow into the oral cavity, flow filling the oral cavity, and expanding walls, respectively. In reference to the circuit in Fig. 1, the glottal volume velocity is equal to the sum of the wall volume

velocity, articulatory volume velocity, mouth volume velocity, and volume velocity due to air compression in the upper vocal-tract cavity (i.e.  $U_G = U_W + U_{ARTIC} + U_M + U_A$ ). ( $U_{ARTIC}$  is the volume velocity due to change in the physical dimensions of the volume between the glottis and the supralaryngeal constriction. The last component, volume velocity due to air compression,  $U_A$ , is not shown in the results, and it was ignored in the computation when the upper vocal tract was relatively unconstricted. This was considered to be the case when intraoral pressure was less than 40 dy/cm<sup>2</sup>.) Recall that, because the higher frequency aspects of the propagation have not been modeled, the inverse-filtered measured flow at the mask cannot strictly be compared to the simulated glottal flow,  $U_G$ , with the low-frequency wall flow,  $U_W$ , articulatory flow,  $U_{ARTIC}$ , and flow due to air compression,  $U_A$ , subtracted off. However, these simulations do illustrate how the low-frequency aspects of components like wall flow could contribute to the glottal flow to produce the measured flow out of the mouth.

Along with the volume velocities and intraoral pressure, various input quantities are also shown, as needed for the discussion. These include subglottal pressure,  $P_S$ , glottal rest area,  $TMA_G$ , and the Q-factor, QF, for the two-mass model.

#### 3.1. /aha /

The utterance /aha/ involved abduction of the glottis for devoicing and aspiration, and subsequent adduction and voicing as its main features. Simulating the details of /aha/ helped to determine the appropriate abductory-adductory glottal movements for other utterances studied here.

Variation in the input quantities was based on experimental results. It has been observed that subglottal pressure decreases upon abduction of the folds in /h/ (Löfqvist, 1975; Ohala, 1990). Thus, the subglottal pressure was allowed to drop from 7000 dy/cm<sup>2</sup> to 6000 dy/cm<sup>2</sup> at maximum glottal opening (Fig. 3a). The two-mass rest area trajectory,  $TMA_G$ , was given a slight asymmetry with a more rapid adduction than abduction (Fig. 3a), and this produced asymmetric mouth flow,

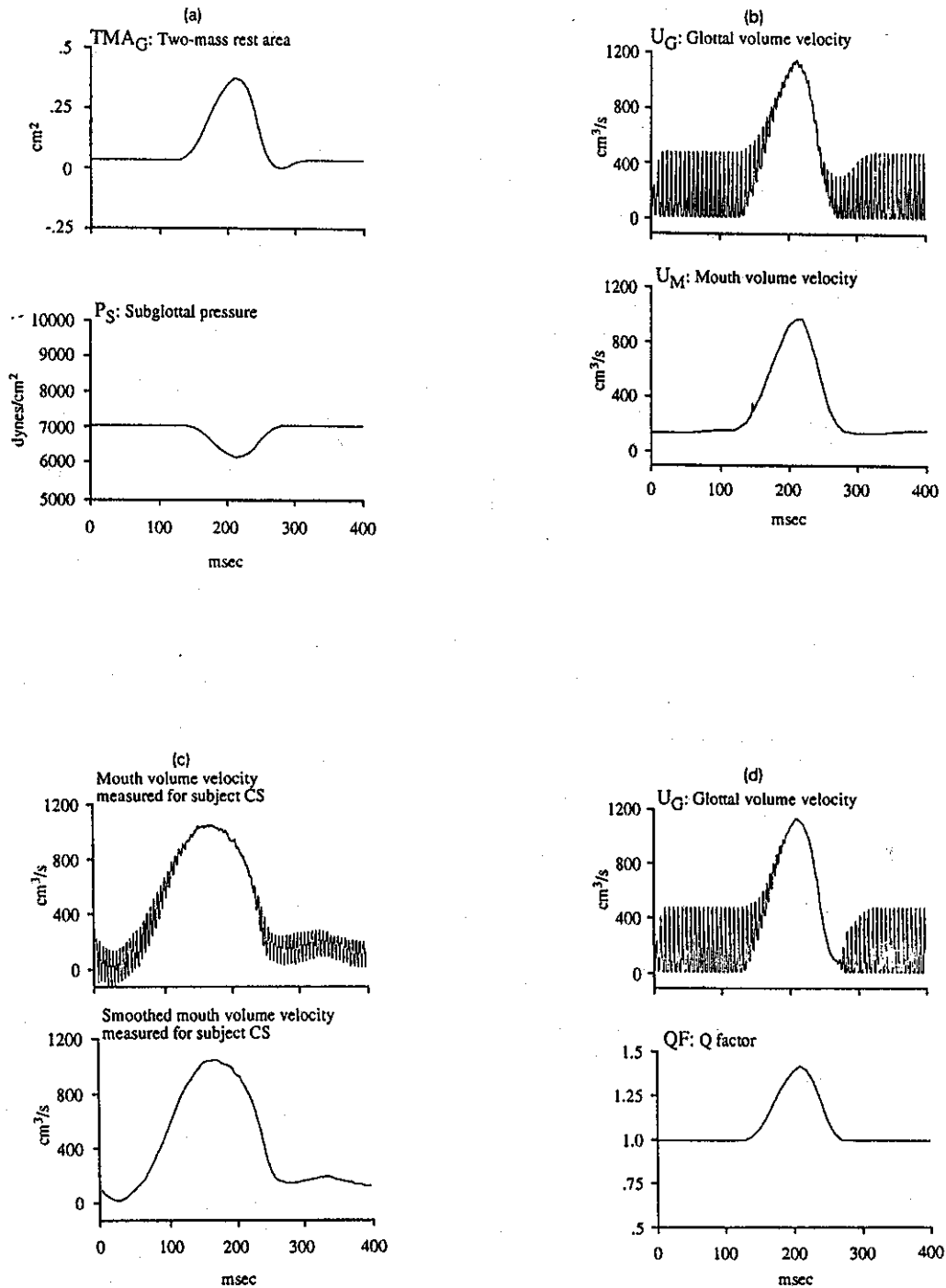


Fig. 3. (a) Two-mass rest area ( $TMA_G$ ) and subglottal pressure ( $P_S$ ) for /aha/ simulation. (b) Glottal volume velocity ( $U_G$ ) and mouth volume velocity ( $U_M$ ) for /aha/ simulation. (c) Measured mouth volume velocity and measured, smoothed mouth volume velocity for /aha/ of subject CS. (d) Glottal volume velocity ( $U_G$ ) and Q-factor (QF) for /aha/ simulation with increasing Q-factor.

$U_M$  (Fig. 3b). Further, a slight hyperadduction as the folds closed was assumed, so there was a decrease in effective glottal area slightly beyond the rest value at the moment of adduction. Such overshoot would explain phenomena observed during adduction in running speech (Löfqvist and McGowan, 1992; Löfqvist et al., 1995). Both the asymmetry and the overshoot were consistent with CS's mouth flow (Fig. 3c).

Overall, there was good agreement between the smoothed measured flow of CS and the simulated mouth flow,  $U_M$ , although CS's abduction-adduction phase was longer than the simulated version's (Figs. 3b and 3c). (Note that the absolute time scales of gestures throughout the simulations were somewhat coarse. Most attention was paid to the relative timing of articulators.) Because there was very little volume flow in the simulation from the movements of the articulators and the walls for this particular utterance, it is valid to compare the unsmoothed inverse-filtered measured flow and the simulated glottal flow (Figs. 3b and 3c). The amplitude of the voice pulsing was high in the simulated version, but the overall pattern of diminished amplitude of voicing on abduction and increased amplitude of voicing on adduction was mimicked in the simulation.

It was hypothesized that the Q-factor increases during abduction in voiceless consonants because the cricothyroid muscle may be invoked to suppress voicing (Löfqvist et al., 1989). In early simulations, it was seen that an increase in the Q-factor did have the effect of suppressing oscillation during the abduction. Also, the increase helped to alleviate the unnatural growth in volume velocity pulses seen when the rest area of the two-mass model was increased, as noted by (Bickley, 1991)<sup>1</sup>.

Although the increase in the Q-factor during

abduction appeared to be well-motivated based on previous experimental results, there were some facts that bounded the increase. The increase in fundamental frequency into the aspiration that would be expected with an increasing Q-factor did not appear consistently in the speech data. In the companion study, however, evidence for an elevated Q-factor appeared for some speakers during the adduction when a fundamental frequency higher than that of the subsequent vowel's was observed (Löfqvist et al., 1995). Also, for simulations of /aha/ run with the variation in Q-factor, it was found that too large an increase not only suppressed voice on abduction, but would not allow timely initiation of voice on adduction. Fig. 3d shows what happened when the Q-factor was increased beyond an upper limit: after the cessation of voicing, the initiation of voice took too long, even allowing rest area and Q-factor to change much more rapidly on adduction than on abduction. Thus, although an increase in the Q-factor was useful for suppressing the growth of volume velocity pulses during abduction, the difficulty in initiating voice restrained the use of too large an increase. This phenomenon is related to the hysteresis observed in the behavior of the two-mass model. Given symmetric trajectories in its parameters, voicing ceased for lower transglottal pressures and higher values of rest area and Q-factor than the values at which voicing started. Such a hysteresis may occur in real speech, e.g. (Hirose and Niimi, 1987). Here, the easiest way to get voicing during the aspiration was not to let it die in the first place, and for values of Q-factor less than about 1.4 this was found to be the case. The Q-factor was only allowed to increase to about 1.3 for simple abduction in the subsequent simulations. Also, as previously discussed, it was necessary to add control over the portion of the folds to be included in the abducted posterior region to avoid increasing the glottal volume velocity pulse amplitude while maintaining voicing during abduction.

### 3.2. /a<sup>?</sup>a/

It was a relatively simple matter to simulate the gross features of the glottal stop. All that had

<sup>1</sup> Bickley (1991) mentioned that another possible mechanism for suppressing vocal fold oscillation during abduction would be to abduct the upper and lower margins of the folds differentially. This differential abduction would be caused by setting the arytenoids so that the upper margins of the folds are more abducted than the lower margins, thus creating a divergent channel. This could be simulated in the two-mass model by setting a larger rest area for the upper masses than for the lower.

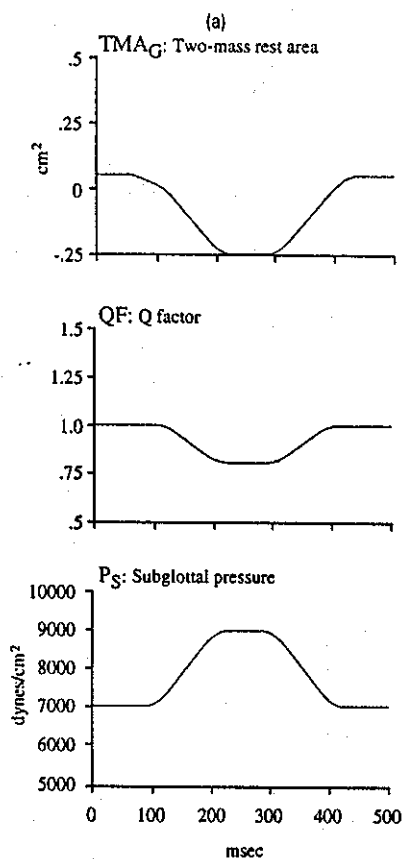
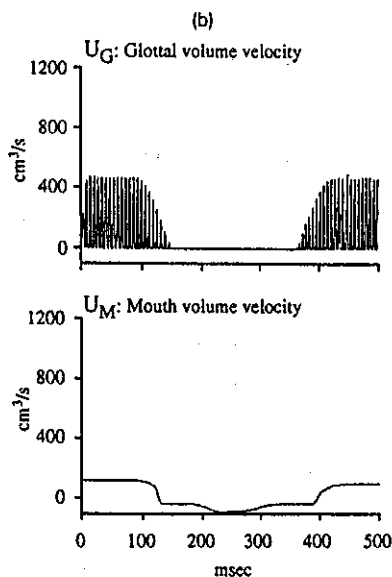


Fig. 4a. Two-mass rest area ( $TMA_G$ ), Q-factor (QF) and subglottal pressure ( $P_S$ ) for /a<sup>2</sup>a/ simulation.



to be done was to decrease the rest area between the masses to a negative value, simulating compression (Fig. 4a). Because the vocal folds were tightly closed, it was supposed both that average subglottal pressure would increase during /<sup>2</sup>/ and that the Q-factor would decrease (Fig. 4a).

Creaky voice was apparent in some of the measured tokens of glottal stops. It is the opinion of the authors that while a large decrease in the Q-factor does lead to erratic behavior in two-mass oscillation, a friction force between the folds of the two-mass model would be needed to model creaky voice. It is known that the viscosity of the tissue internal to the folds affects the voicing threshold of phonation, e.g. (Finkelhor et al., 1987), but the effect of the viscosity of the outer surface of the epithelium is unknown. The model we have in mind is that of a stick-slip phenomenon, similar to that of earthquakes. Earthquakes can be modeled as a self-organized critical system, where small and large energy events result from the same underlying cause, and a power law relates the frequency of large and small events (Bak et al., 1987).

In the preceding two utterances, the measured, unsmoothed mouth volume velocity was nearly equal to glottal volume velocity produced in the simulations, except for small contributions from articulatory movement. For the rest of the

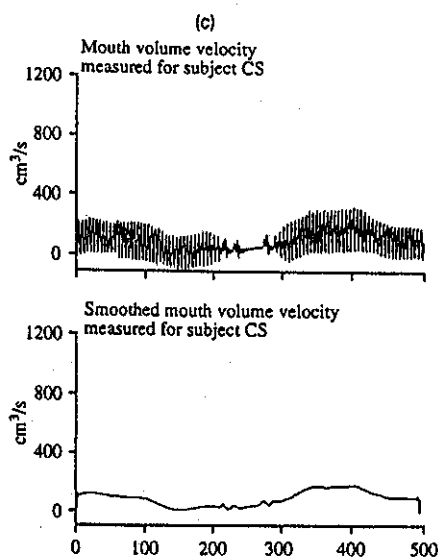


Fig. 4. (b) Glottal volume velocity ( $U_G$ ) and mouth volume velocity ( $U_M$ ) for /a<sup>2</sup>a/ simulation. (c) Measured mouth volume velocity and measured, smoothed mouth volume velocity for /a<sup>2</sup>a/ of subject CS.

utterances this was not the case. This is where simulation becomes important in interpreting the smoothed, measured data on flow through the mouth. That is, this analysis-by-synthesis helps to estimate the contributions of smoothed glottal flow, wall flow and air flow due to air compliance to the smoothed, measured flow from the mouth. For reasons given in the introduction, only a crude estimate of glottal flow (unsmoothed and with voice pulses) can be obtained from the unsmoothed mouth flow by subtracting off the low-frequency supralaryngeal components of flow due to wall and air compliance. Therefore, a direct comparison of simulated flow from the two-mass model and the measured mouth flow minus simulated low-frequency supralaryngeal components would be tenuous. It should also be noted that because a subglottal system was not a part of the model, the subglottal pressure trajectories were imposed, and not derived. The estimates of subglottal increase during supralaryngeal obstruction could be underestimated, thus making wall volume velocity upon release in the simulations lower than what it should have been. To keep mouth flow at voiceless releases from increasing too much in the simulations if the subglottal pressure were increased, it would be necessary to reduce glottal aperture area at release from the values in the simulations presented below.

### 3.3. /ava/

This utterance was simulated with a close approximation of the teeth to the lower lip and a minimum constriction area during the fricative of either 0.032 cm<sup>2</sup> or 0.017 cm<sup>2</sup>. Neither one of these values was held, but rather they were the low points of a constriction area trajectory. The subglottal pressure and laryngeal parameters were left untouched, except for a small increase in the effective glottal area entering the fricative (Fig. 5a).

Changes in minimum constriction area had a large effect on the results of the simulation. With a minimum constriction area of 0.032 cm<sup>2</sup>, voicing was diminished, but it continued throughout (Fig. 5b), because the increase in intraoral pressure was not sufficient to suppress voicing com-

pletely (Fig. 5a). On the other hand, with a minimum constriction area of 0.017 cm<sup>2</sup>, the rapid rise in intraoral pressure was sufficient to suppress voice soon after the constriction was formed (Fig. 5d). Because of the higher intraoral pressure in the more tightly constricted case, there was a greater mouth flow upon release (Figs. 5b and 5d). In both cases there was a contribution to outward flow from articulatory movement and vocal tract wall capacitance at release. Fig. 5b shows the full set of results for the larger constriction, which were similar to those of the smaller constriction shown in Fig. 5d. The slight rise in mouth flow just before closure was caused by the small abduction of the folds. The rise in mouth flow in this simulation just before closure was observed in the data (Fig. 5c), although it is uncertain whether this was due to articulatory movement or abduction of the folds. A slight abduction could make it easier to sustain voicing in an environment of reduced transglottal pressure.

The measured token of CS's /ava/ seemed more closely to match the situation with the tighter constriction because voice pulsing stopped sometime during the constriction (Figs. 5c and 5d). However, the magnitude of the mouth volume velocity pulse more closely matched the simulation with the looser constriction (Figs. 5b and 5c). The reason for this discrepancy may have been an overestimation of flow due to articulatory movement (Fig. 5b).

### 3.4. /aba/

This utterance involved complete closure of the lips. Other than this, the configurations of the glottal parameters and the supralaryngeal tract were much the same as in the utterance /ava/, except that the glottal rest area was not allowed to increase. With all parameters at default values, voicing was not maintained for longer than two or three pitch pulses into the closure. Modification of the glottal rest area and the Q-factor did not help maintain voicing. The only way found to sustain voicing through most of the closure was to increase the wall capacitance by a factor of nearly 6 from the default value. With this increase in

wall capacitance, there was a more gradual rise in intraoral pressure (Fig. 6a). CS's unsmoothed flow data showed sustained vibration throughout the closure (Fig. 6c). While she may have increased the capacitance of her vocal tract walls almost 6-fold from her apparent default capacitance to sustain voicing, it was also probable that her oral volume was actively changed using the larynx, pharyngeal walls or velum to maintain voicing (Bell-Berti, 1975; Westbury, 1983). The ASY vocal tract did not afford a means of increasing supralaryngeal volume using these articulators.

The peak wall volume velocity was much higher at the release of the /b/ than it was for the /v/, because the closure was complete for /b/, and the wall capacitance during /b/ was larger (Figs. 6b and 5b). The large volume velocity from the mouth after release of the labial closure was mostly due to the wall volume velocity (Fig. 6b). Comparison with the smoothed version of CS's mouth flow shows that the simulated mouth volume velocity was too large. This corroborates the theory that sustained voicing is due to active upper vocal-tract volume control rather than to a large increase in wall capacitance.

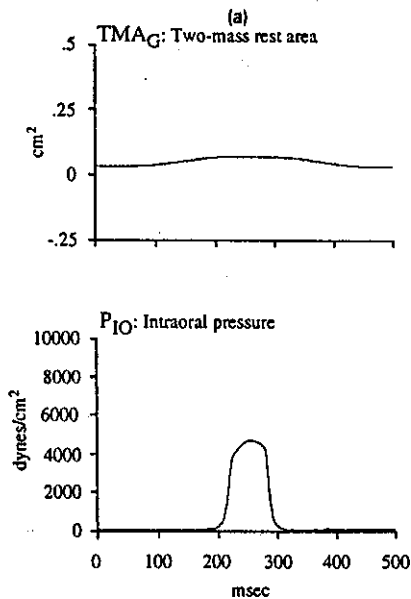


Fig. 5a. Two-mass rest area (TMA<sub>G</sub>) and intraoral pressure (P<sub>IO</sub>) for /ava/ simulation.

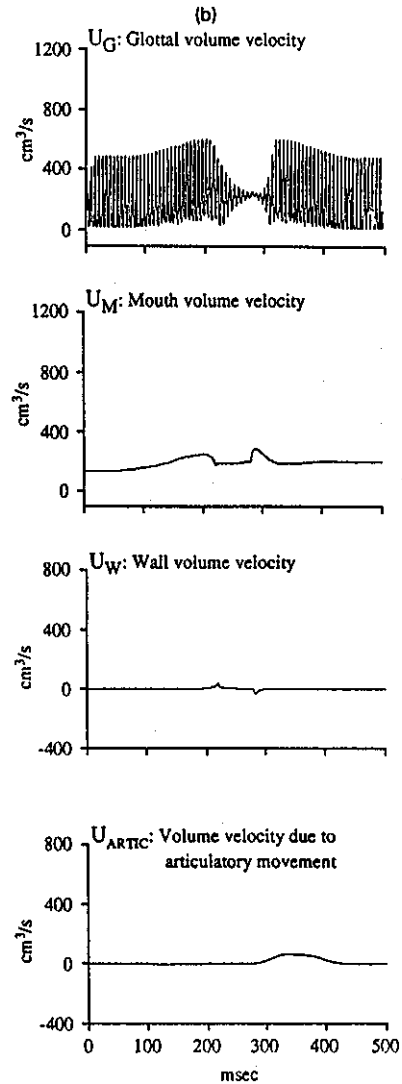


Fig. 5b. Glottal volume velocity ( $U_G$ ), mouth volume velocity ( $U_M$ ), wall volume velocity ( $U_W$ ) and volume velocity due to articulatory movement ( $U_{ARTIC}$ ) for /ava/ simulation.

The coordinated activity of the larynx and the upper articulators is important for the following utterances.

### 3.5. /apa/

The sequence of upper articulators movements here was similar to that for /aba/. The lips were closed for about 75 ms, with the glottal abduction

gesture begun about 60 ms before closure (Fig. 7a). There was a rise in the value of the Q-factor with the abduction from the default value of 1 to 1.25 (Fig. 7a). This was to simulate the tensing of the folds with abduction, as was discussed for the aspirate /h/ (Löfqvist et al., 1989). The subglottal pressure was allowed to decrease at the start of abduction, to increase during closure to the same level it had during voicing, to decrease after the release of the lips, and then to rise again for phonation (Fig. 7a). This pattern was based on data reported in Löfqvist (1975) with one exception: he observed a subglottal pressure during closure greater than that during voicing. This great a rise in subglottal pressure was not simulated because it led to too large a mouth volume velocity upon release.

The volume velocity due to the vocal tract walls was very important in effecting the rise in intraoral pressure during closure, and the constriction volume velocity peak at release (Fig. 7b). The sharp rise in intraoral pressure and narrow, tall positive spike in wall velocity for /apa/ (Figs. 7a and 7b) can be contrasted with the gradual

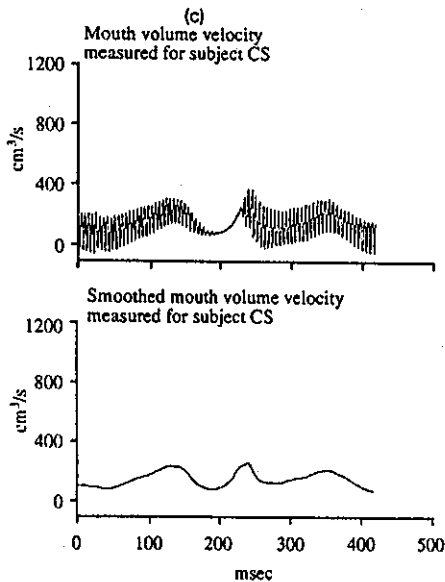


Fig. 5c. Measured mouth volume velocity and measured, smoothed mouth volume velocity for /ava/ of subject CS.

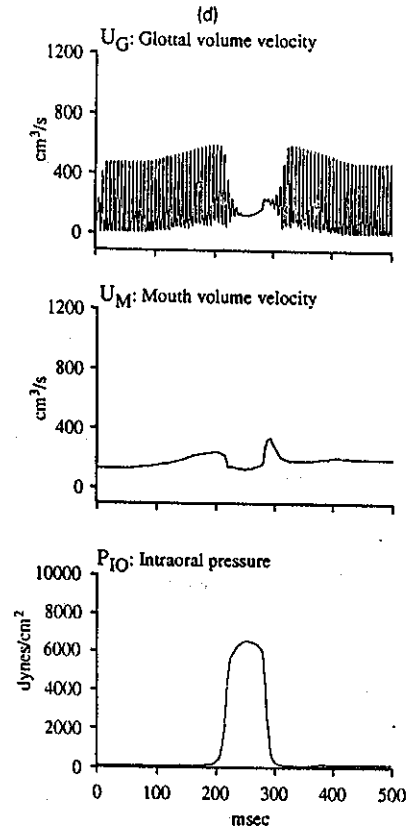


Fig. 5d. Glottal volume velocity ( $U_G$ ), mouth volume velocity ( $U_M$ ) and intraoral pressure ( $P_{IO}$ ) for /ava/ simulation for small constriction condition.

rise in intraoral pressure and broad, low positive wall velocity for /aba/ (Figs. 6a and 6b). These differences can be attributed to the differences in wall capacitance between the two simulations. At release of the /p/, the large negative spike in the wall volume velocity provides the initial surge in the mouth volume velocity (Fig. 7b).

The relative timing of glottal adduction and release of the lips was very important for initiating voice early in the aspiration phase. In order that some weak voicing be initiated quickly after release, it was necessary to begin adduction about 40 ms before release, so that the effective rest area of the glottis at release was about half of what it was at maximum opening. Adduction

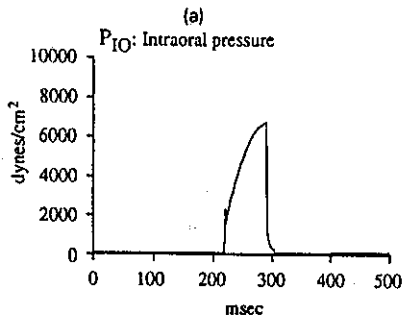


Fig. 6a. Intraoral pressure ( $P_{IO}$ ) for /aba/ simulation.

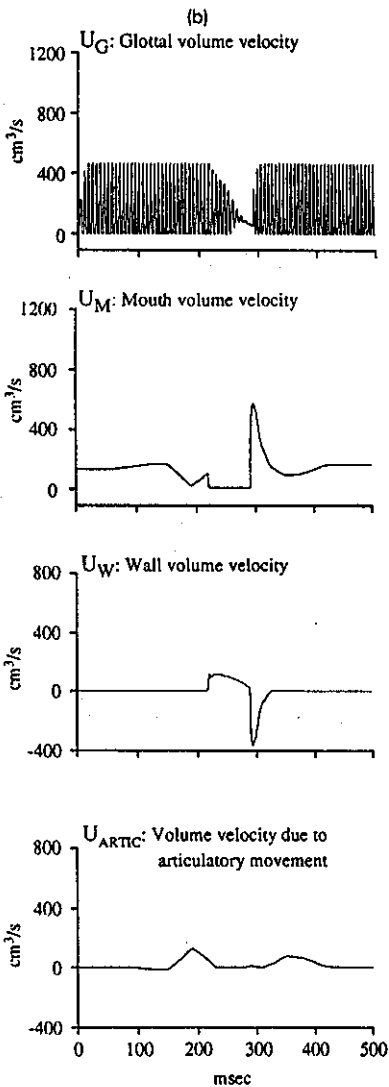


Fig. 6b. Glottal volume velocity ( $U_G$ ), mouth volume velocity ( $U_M$ ), wall volume velocity ( $U_W$ ) and volume velocity due to articulatory movement ( $U_{ARTIC}$ ) for /aba/ simulation.

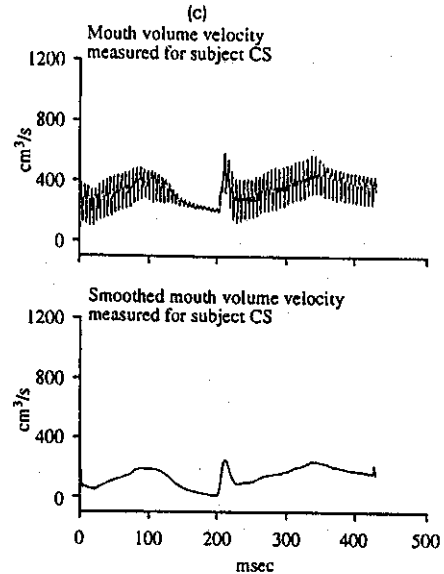


Fig. 6c. Measured mouth volume velocity and measured, smoothed mouth volume velocity for /aba/ of subject CS.

starting at release of the lips would result in voice being delayed too long<sup>2</sup>. It was probably unrealistic to suppose that glottal adduction actually was initiated so early (Löfqvist and Yoshioka, 1984). This represents a weakness in the model presented here, which might be alleviated by adding more physical detail. These simulations did not account for changes in glottal rest area caused by changes in intraoral pressure (Stevens, 1991). With an abrupt reduction of intraoral pressure there may be a quick reduction in glottal rest area, particularly for the upper margins of the folds. This acceleration toward a convergent configuration could result in sustained oscillation, that is, voicing.

<sup>2</sup> An attempt was made to help initiate voice by creating a convergent glottal channel (Bickley, 1991). However, little effect was seen using a two-to-one ratio of lower to upper rest area. Our results with regard to a converging/diverging channel are inconclusive because of the many possible combinations of parameters under control.



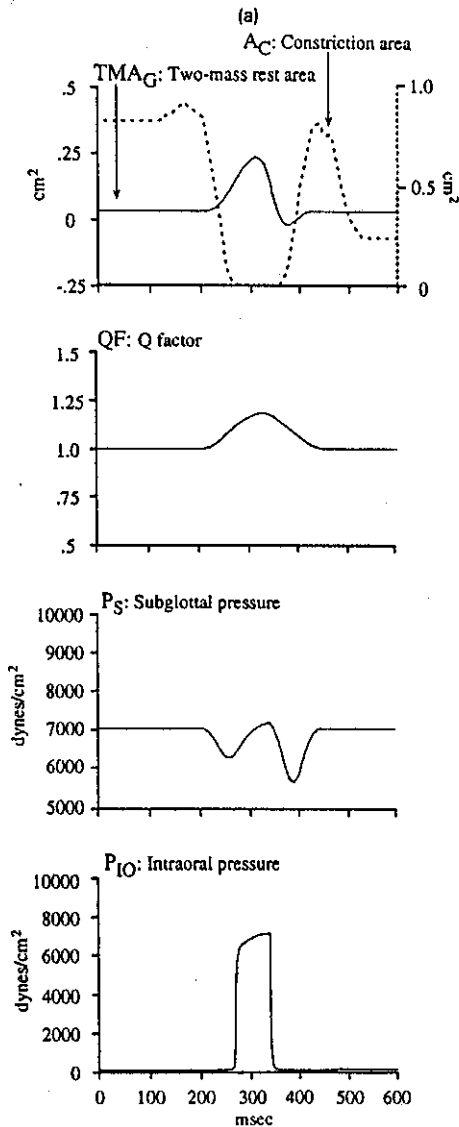


Fig. 7a. Two-mass rest area ( $TMA_G$ ), Q-factor (QF), subglottal pressure ( $P_S$ ) and intraoral pressure ( $P_{IO}$ ) for /apa/ simulation.

The simulated mouth flow,  $U_M$ , (Fig. 7b) and the smoothed measured mouth flow of CS (Fig. 7c) compare fairly well except for the rapid change in flow at closure for the simulation. This was the result of the abruptness with which the lips were closed. Without better temporal resolution in articulatory control, it was not possible to simulate a deceleration in the closure of the lips in the last few milliseconds.

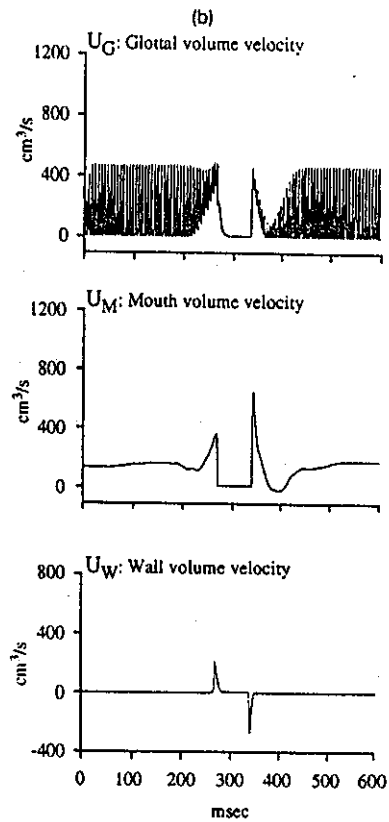


Fig. 7b. Glottal volume velocity ( $U_G$ ), mouth volume velocity ( $U_M$ ) and wall volume velocity ( $U_W$ ) for /apa/ simulation.

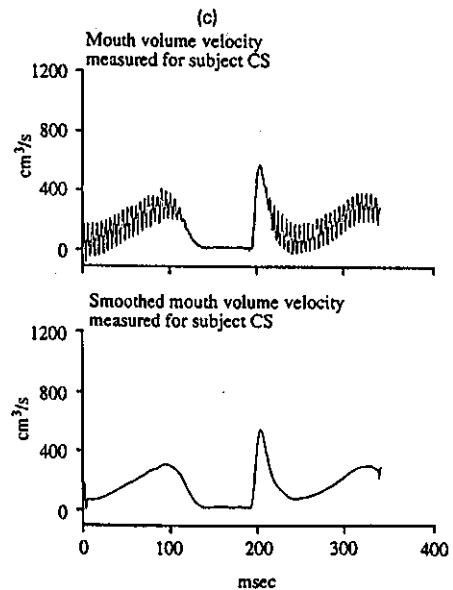


Fig. 7c. Measured mouth volume velocity and measured, smoothed mouth volume velocity for /apa/ of subject CS.

### 3.6. /asa/

The effective glottal area and the Q-factor trajectories for /asa/ were similar to those for /apa/ (Fig. 8a), except the magnitudes of the increases were larger in this case. The folds started to abduct about 50 ms before minimum tongue-tip constriction area, which was  $0.01 \text{ cm}^2$  for /s/. At the beginning of tight constriction, they were abducted to about one-half of the maximum glottal area. The subglottal pressure was allowed to drop to  $6000 \text{ dy/cm}^2$  before maximum constriction of the tongue tip was attained, and then it rose again during the fricative (Fig. 8a). Glottal adduction began about 50 ms before release of the tongue-tip constriction, and it was nearly adducted for voicing when the constriction was released (Fig. 8a). Again, a slight overshoot was allowed in the glottal rest area during the adduction phase.

The effects of changes in the glottal area and upper vocal tract constriction area can be seen in the intraoral pressure and constriction volume velocity (Figs. 8a and 8b). As the glottis was abducted, there was an increase in mouth flow volume velocity, and, until the time when minimum constriction was attained, there was a large average glottal flow (Figs. 8a and 8b). This glottal flow, with a small amount subtracted at closure due to the outward movement of the vocal-tract walls, provided the flow through the mouth (Fig. 8b). When the minimum area was attained, the glottal and mouth volume velocities were greatly reduced, and the vocal tract walls began to absorb some of the glottal flow. At release, with the glottis fairly well adducted, the mouth volume velocity was maintained momentarily only because the vocal tract walls and the capacitance of the volume of enclosed air began to push air out (Fig. 1). Normal voicing began soon after release, because the folds were well adducted at that time.

Overall, the simulated mouth flow,  $U_M$ , (Fig. 8b) compares well with the smoothed measured mouth flow of CS (Fig. 8c). The relatively sharp peaks in simulated mouth flow during abduction and constriction formation, and during adduction and constriction release, are due to the abrupt-

ness with which small constriction areas were attained and released. It appears that CS attains and releases small constriction areas more smoothly.

### 3.7. /aspa/

The utterance /aspa/ had a long abductory phase before the tongue-tip constriction, much like for /asa/ (Figs. 9a and 8a). In the simulation of /aspa/ the folds were allowed to start abducting about 80 ms before the constriction for /s/ was attained (Fig. 9a). As with utterance /asa/, the Q-factor covaried with the glottal area and subglottal pressure was decreased to  $6000 \text{ dy/cm}^2$  (Fig. 9a). Maximum glottal opening was attained during the /s/, and at about the time for /p/ closure the folds were fully adducted (Fig. 9a). Almost immediately, the intraoral pressure,  $P_{IO}$ , jumped from  $6000 \text{ dy/cm}^2$  to  $7000 \text{ dy/cm}^2$ , which was what the subglottal pressure,  $P_S$ , was at the time of p-closure (Fig. 9a). At the release of the bilabial closure, there was immediate onset of voice (Fig. 9b). Most of the flow from the mouth,  $U_M$ , was due to the inward movement of the vocal tract walls, with some contribution from articulatory movement,  $U_{ARTIC}$ , and the glottal flow,  $U_G$  (Fig. 9b).

The simulated  $U_M$  (Fig. 9b) compares well with the smoothed measured mouth flow (Fig. 9c). As for /apa/ and /asa/, the changes into and out of tight constrictions are a little too abrupt. In both the measured and simulated mouth flows, a slight increase in flow can be seen before the bilabial closure. At least part of this rise in the simulation was due to an increase in intraoral pressure,  $P_{IO}$ , which, in turn, was caused by an increase in subglottal pressure,  $P_S$ , from its minimum after abduction (Fig. 9a).

### 3.8. Voice pulses

In the simulations done here, the low-frequency flow patterns and voicing thresholds were of the greatest concern. However, it is instructive to compare the voice pulse characteristics from the two-mass model to the measured flows from Löfqvist et al. (1995), recalling the

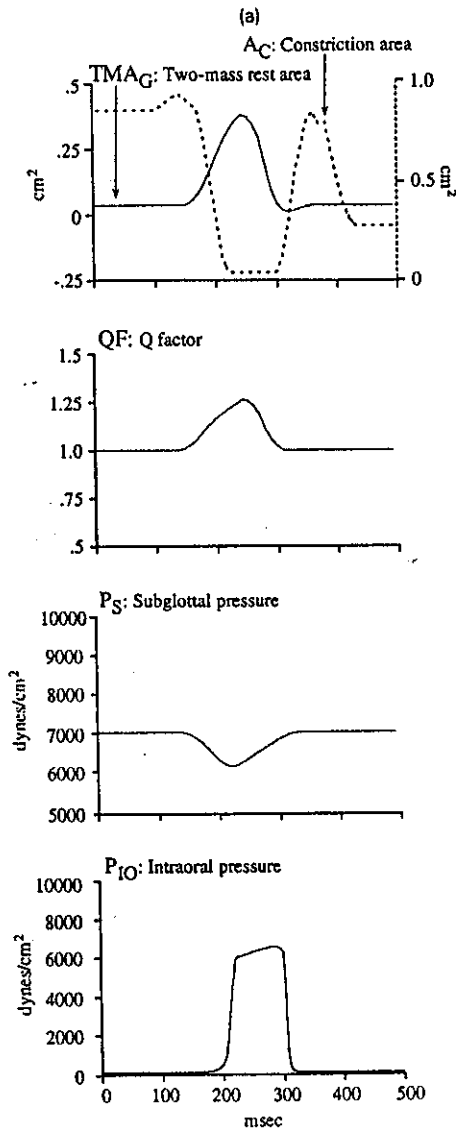


Fig. 8a. Two-mass rest area ( $TMA_G$ ), Q-factor (QF), subglottal pressure ( $P_S$ ) and intraoral pressure ( $P_{IO}$ ) for /asa/ simulation.

caveats of Section 1.1. Default values for the two-mass model were used for the voice during sustained phonation. In relation to CS's voicing in the middle of the vowels, the simulated glottal volume velocity pulse height was too large, the fundamental frequency too low, and the open quotient too small. This would seem to indicate that the default Q-factor used was too small for

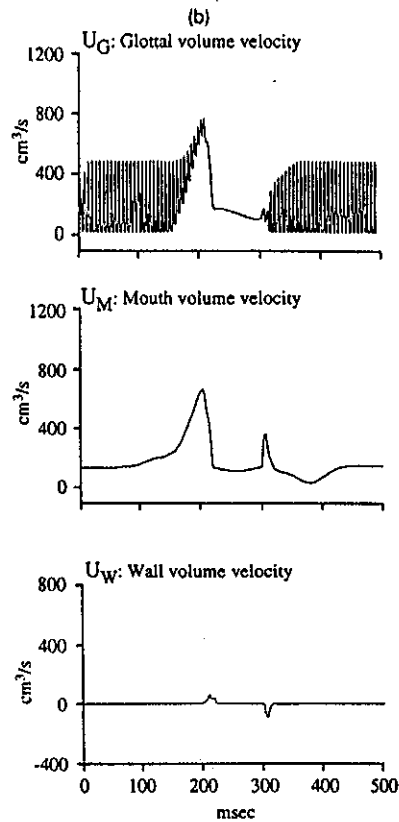


Fig. 8b. Glottal volume velocity ( $U_G$ ), mouth volume velocity ( $U_M$ ) and wall volume velocity ( $U_W$ ) for /asa/ simulation.

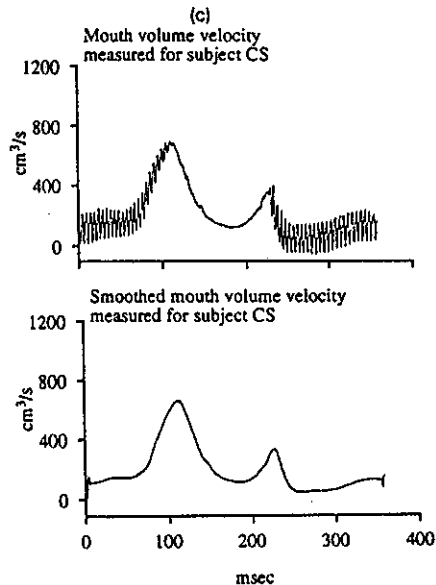


Fig. 8c. Measured mouth volume velocity and measured, smoothed mouth volume velocity for /asa/ of subject CS.

CS. A larger Q-factor was avoided because of the difficulty in initiating voice after constriction releases. Also, a two-mass model with shorter glottal length dimension, resulting in smaller rest areas, may have made the pulse amplitudes closer to those of CS.

In the simulations, the Q-factor was increased going into voiceless consonants /s/, /p/, /h/

and /sp/. These increases in the Q-factor meant heightened fundamental frequency before and after the consonants (Fig. 10). The simulations are consistent with the data of CS and others in showing a heightened fundamental frequency in the adductory portion of the /s/, /p/, /h/ and /sp/. However, the simulations are not consistent with the data in showing increased funda-

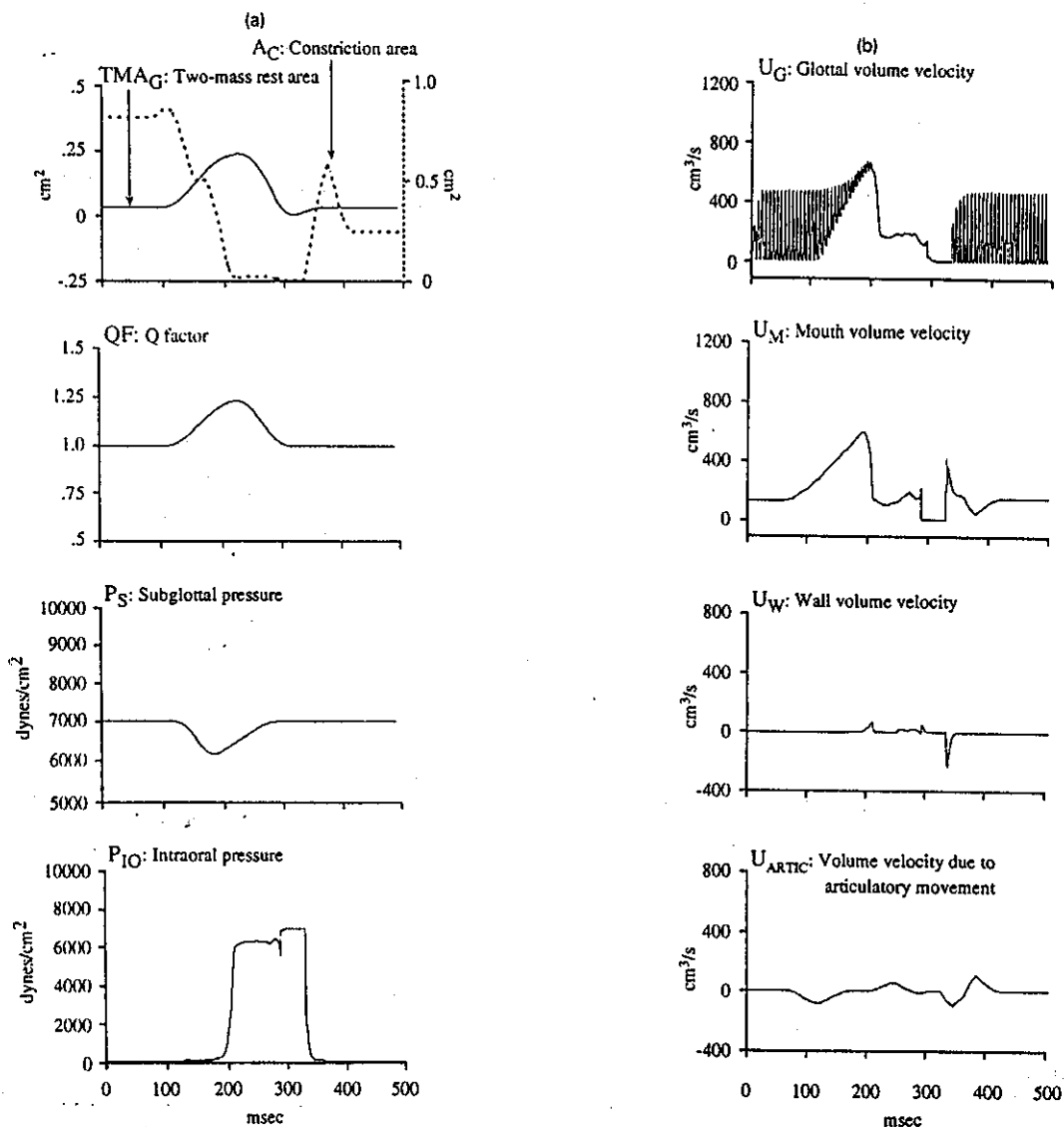


Fig. 9. (a) Two-mass rest area (TMA<sub>G</sub>), Q-factor (QF), subglottal pressure (P<sub>S</sub>) and intraoral pressure (P<sub>IO</sub>) for /aspa/ simulation. (b) Glottal volume velocity (U<sub>G</sub>), mouth volume velocity (U<sub>M</sub>), wall volume velocity (U<sub>W</sub>) and volume velocity due to articulatory movement (U<sub>ARTIC</sub>) for /aspa/ simulation.

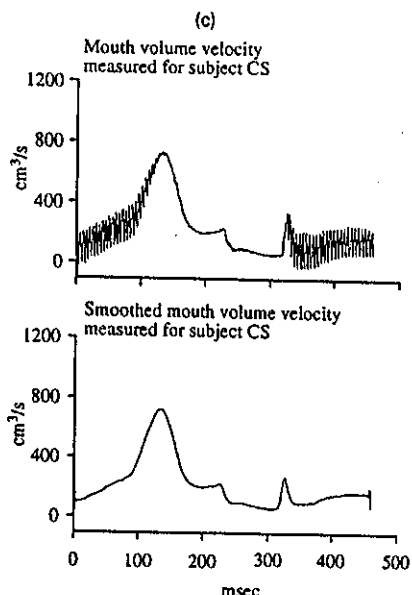


Fig. 9c. Measured mouth volume velocity and measured, smoothed mouth volume velocity for /aspa/ of subject CS.

mental frequency before voiceless consonants for CS and other subjects in Löfqvist et al. (1995). Either the increase in Q-factor was started too early in the simulations, or voicing was allowed to continue too long into the consonant. Also, there was no allowance made for the intonational contour in the simulations.

The simulated obstruents /b/, /v/, /s/ and /sp/ showed increasing open quotient before minimum constriction, and during constriction (if the consonant was voiced) due to decreasing transglottal pressure (Fig. 11). This was one of the most noticeable effects of the upper vocal tract's low-frequency aerodynamics on the voice source. The voiceless consonants, /s/, /sp/, /h/ and /p/ also show an increase in open quotient before maximum abduction because of the increasing rest area between the masses, as well as an increase in the Q-factor. The subject CS did not show these trends, probably because her open quotient was high to begin with. However, the

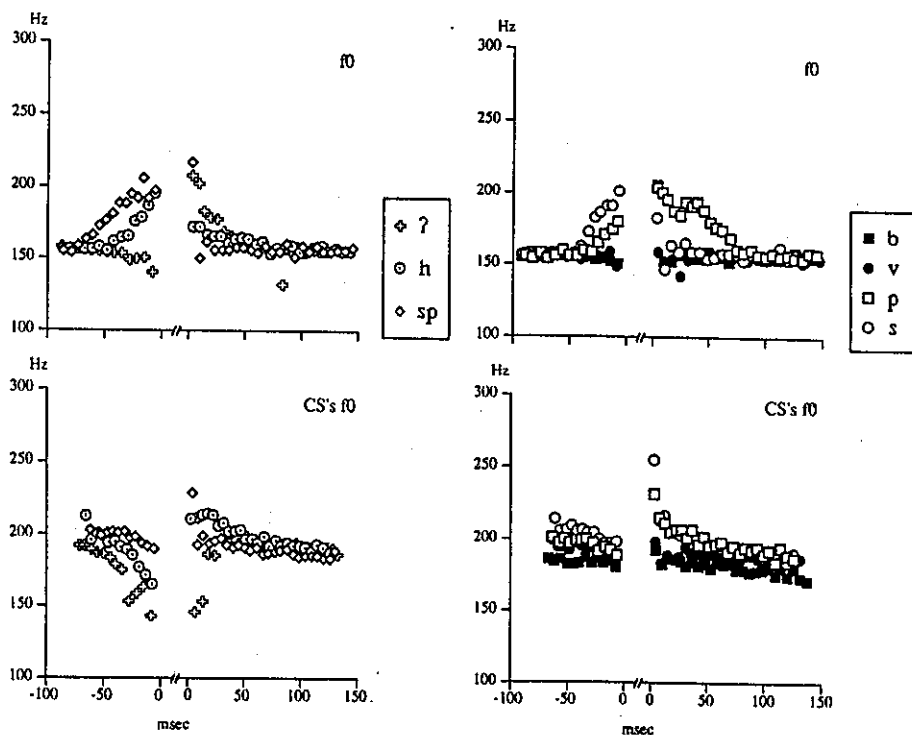


Fig. 10. Simulated and measured fundamental frequency (f<sub>0</sub>).

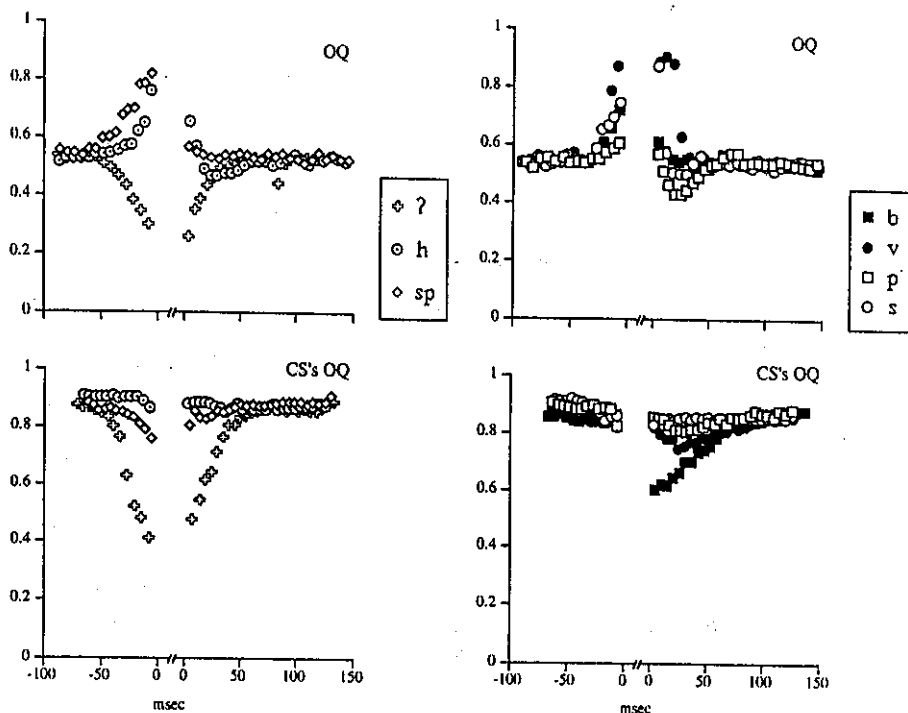


Fig. 11. Simulated and measured open quotient (OQ).

simulations' open quotients do seem to follow the majority of subjects in the (Löfqvist et al., 1995) experiments.

During the abduction and adduction phases, the minimum flow regions of the simulated pulses were too flat. Normally, in the data, when there is positive, nonzero minimum flow, the minimum flow phase is not static and straight, but rather changing and curvilinear. Thus, the simple fix of allowing two different regions, one posterior and the other anterior, did not produce a natural glottal pulse while the glottis was open. Although this change did allow for a diminished pulsing amplitude during abduction and adduction, Bickley and Stevens (1986) note that glottal pulses decrease in both area and abruptness of closure in the environment of a highly constricted vocal tract.

#### 4. Discussion

The results presented here show that a combination of the two-mass model with a low-

frequency aerodynamic model can be used to simulate the rudimentary aerodynamics involved in simple sonorant-obstruent sequences in English. A careful specification of the trajectories of and relative timing between the various control parameters is necessary to obtain simulations of mouth flow that fit the data. This provides some evidence that flow data can be used to infer articulatory movement, despite the crude estimates used here of such things as vocal-tract volume change and wall impedance. Also, these simulations, in agreement with previous simulations, e.g. (Müller and Brown, 1980), show that the flows due to vocal-tract wall movement and articulatory movement can add significantly to the glottal flow to produce the flow through the mouth.

The two-mass model appears to be too simple for good simulation of glottal flow pulses during transition, however, even with the addition of a simple model for the third dimension. A more sophisticated vibratory model is necessary to obtain the type of glottal pulsing associated with the breathy voice seen during many transitions, for

example into the /s/ constriction and after the /p/ release. The overall amplitude of vibration and the abruptness of closure could be reduced by using modifications to the two-mass model proposed by Koizumi et al. (1987). These include a vertical degree of freedom for the masses and variable thickness for the upper mass. A more sophisticated mechanical model for the anterior-posterior, or third, dimension could be used to round the minimum flow portions of the glottal pulses when the folds do not close completely during abduction and adduction. The development of a beam model for the third dimension should improve naturalness in abduction and adduction (Johnson, 1992).

The initiation of voice for the two-mass model may be more difficult after constriction release than for real larynges, even given the hysteresis noted for real larynges. Attempting a convergent channel with the upper masses closer together than the lower did not appear to help initiate voicing (cf. footnote 2). An improvement would be to let the equilibrium distance between the masses depend on the appropriate pressures, so that subglottal pressure acts to force the lower masses laterally, and intraoral pressure acts to force the upper masses as modeled by Stevens (1988). Allowing the rest distance between the masses to react to the rapid pressure changes during constriction release would allow one to see if moving toward a convergent configuration is more efficient in getting voice started than just allowing a convergent rest configuration without pressure coupling. The addition of a subglottal system might be tried for future simulations. A simple model, such as those used by Rothenberg (1968) and Ohala (1990), might be sufficient for a low-frequency simulation.

### Acknowledgements

This work was supported by NIH grants DC-00865, DC-01247 and DC-00121 to Haskins Laboratories, and in part by Esprit-BR Project 6975 – Speech Maps through grant P55 from the Swedish National Board for Industrial and Technical Development.

### References

- T.V. Ananthapadmanabha and G. Fant (1982), "Calculation of true glottal flow and its components", *Speech Communication*, Vol. 1, Nos. 3–4, pp. 167–184.
- P. Bak, C. Tang and K. Wiesenfeld (1987), "Self-organized criticality: An explanation of  $1/f$  noise", *Physical Rev. Letters*, Vol. 59, pp. 381–384.
- F. Bell-Berti (1975), "Control of pharyngeal cavity size for English voiced and voiceless stops", *J. Acoust. Soc. Amer.*, Vol. 57, pp. 456–461.
- C. Bickley (1991), "Vocal-fold vibration in a computer model of a larynx", in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, Inc., San Diego), pp. 37–46.
- C. Bickley and K.N. Stevens (1986), "Effects of a vocal-tract constriction on the glottal source: experimental and modeling studies", *J. Phonetics*, Vol. 14, pp. 373–382.
- B. Cranen and L. Boves (1987), "On subglottal formant analysis", *J. Acoust. Soc. Amer.*, Vol. 81, pp. 734–746.
- P.O.A.L. Davies, R.S. McGowan and C.H. Shadle (1993), "Practical flow duct acoustics applied to the vocal tract", in *Vocal Fold Physiology: Frontiers in Basic Science*, ed. by I.R. Titze (Singular Publishing Group Inc., San Diego), pp. 93–142.
- B.K. Finkelhor, I.R. Titze and P.L. Durham (1987), "The effect of viscosity changes in the vocal folds on the range of oscillation", *J. Voice*, Vol. 1, pp. 320–325.
- J.L. Flanagan, K. Ishizaka and K.L. Shipley (1975), "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *Bell System Tech. J.*, Vol. 54, pp. 485–506.
- H. Hirose and S. Niimi (1987), "The relationship between glottal opening and the translglottal pressure differences during consonant production", in *Laryngeal Function in Phonation and Respiration*, ed. by T. Baer, C. Sasaki and K. Harris (College-Hill Press, Boston), pp. 381–390.
- K. Ishizaka and J.L. Flanagan (1972), "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell System Tech. J.*, Vol. 51, pp. 1233–1268.
- K. Ishizaka and M. Matsudaira (1968), Fluid mechanical considerations for vocal-fold vibration, Monograph 8, Speech Commun. Res. Lab. Inc., Santa Barbara, CA.
- M.A. Johnson (1992), "Using beam elements to model the vocal fold length in breathy voice", *J. Acoust. Soc. Amer.*, Vol. 91, pp. 2420–2421.
- D.H. Klatt, K.N. Stevens and J. Mead (1968), "Studies of articulatory activity and airflow during speech", in *Annals of the New York Academy of Sciences*, Vol. 155, Art. 1, *Sound Production in Man*, ed. by A. Bouhuys (The New York Academy of Sciences, New York), pp. 42–55.
- T. Koizumi, S. Taniguchi and S. Hiromitsu (1987), "Two-mass model of the vocal cords for natural sounding voice synthesis", *J. Acoust. Soc. Amer.*, Vol. 82, pp. 1179–1192.
- Q. Lin (1990), Speech production theory and articulatory speech synthesis, Unpublished Ph.D. dissertation, KTH, Stockholm.

- A. Löfqvist (1975), "A study of subglottal pressure during the production of Swedish stops", *J. Phonetics*, Vol. 3, pp. 175-189.
- 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 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.
- A. Löfqvist, L.L. Koenig and R.S. McGowan (1995), "Vocal tract aerodynamics in /aCa/ utterances: Measurements", *Speech Communication*, Vol. 16, No. 1, pp. 49-66.
- P. Mermelstein (1973), "Articulatory model for the study of speech production", *J. Acoust. Soc. Amer.*, Vol. 53, pp. 1070-1082.
- E.M. Müller and W.S. Brown (1980), "Variations in the supraglottal air pressure waveform and their articulatory consequences", in *Speech and Language: Advances in Basic Research and Practice*, Vol. 4, ed. by N. Lass (Academic Press, New York), pp. 318-389.
- J.J. Ohala (1974), "A mathematical model of speech aerodynamics", in *Speech Communication: Proceedings of the Speech Communications Seminar, Stockholm, 1-3 August 1974 (Vol. 2): Speech Production and Synthesis by Rule*, ed. by G. Fant (Almqvist and Wiksell, Stockholm), pp. 65-72.
- J.J. Ohala (1990), "Respiratory activity in speech", in *Speech Production and Speech Modeling*, ed. by W.J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht), pp. 23-54.
- M. Rothenberg (1968), *The Breath-Stream Dynamics of Simple-Released-Plosive Production*. (S. Krager, Basel).
- M. Rothenberg (1983), "An interactive model of the voice source", in *Vocal Fold Physiology*, ed. by D.H. Bless and J.H. Abbs (College-Hill Press, San Diego), pp. 155-165.
- P. Rubin, T. Baer and P. Mermelstein (1981), "An articulatory synthesizer for perceptual research", *J. Acoust. Soc. Amer.*, Vol. 70, pp. 321-328.
- C. Scully (1990), "Articulatory synthesis", in *Speech Production and Speech Modeling*, ed. by W.J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht), pp. 151-186.
- M.M. Sondhi and J. Schroeter (1987), "A hybrid time-frequency domain articulatory speech synthesizer", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 35, pp. 955-967.
- K.N. Stevens (1971), "Airflow and turbulence noise for fricative and stop consonants: Static considerations", *J. Acoust. Soc. Amer.*, Vol. 50, pp. 1180-1192.
- K.N. Stevens (1988), "Modes of vocal fold vibration based on a two-section model", in *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*, ed. by O. Fujimura (Raven Press, Ltd., New York), pp. 357-371.
- K.N. Stevens (1991), "Vocal fold vibration for obstruent consonants", in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, Inc., San Diego), pp. 29-46.
- K.N. Stevens (Personal communication).
- J.R. Westbury (1983), "Enlargement of the supraglottal cavity and its relation to stop consonant voicing", *J. Acoust. Soc. Amer.*, Vol. 73, pp. 1322-1336.



## Instructions to authors

**General.** Prospective authors are encouraged to submit manuscripts within the scope and the subject matter coverage of the Journal. Papers published are expected to make a significant contribution to our knowledge of the speech communication process. To qualify for publication, papers must be previously unpublished and must not be under consideration for publication elsewhere. All material should be sent in quadruplicate (original plus three copies) to the Editor. Contributors are reminded that once their contribution has been accepted for publication, all further correspondence should not be sent to the Editor, but directly to the publishers (Editorial Department., Elsevier Science B.V. (North-Holland), P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands).

All submissions are transmitted to at least two anonymous referees and will only be accepted on the basis of stringent peer review. Upon acceptance of an article, author(s) resident in the USA will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information under the U.S. Copyright Law.

Articles of interest to a significant number of readers are considered for publication. Criteria for selection are originality, quality, conciseness and relevance to the aims of the Journal. Taking into account the number of different fields encompassed by *Speech Communication*, it is recommended that papers be written so as to be accessible to those outside the immediate specialization.

The accepted language is English. Each contribution should be accompanied by abstracts (see *Abstracts*). Page proofs will be sent to the author with an offprint order form. Fifty offprints of each article are supplied free of charge. Costs arising from alterations in proof, other than of printer's errors, will be charged to the authors. Typed material (including abstracts, list of references) should be double spaced on one side of A4 size (297x210 mm) paper with a wide margin. All pages should be numbered. The first page should include the article title and the author's name and affiliation, as well as a name and mailing address to be used for correspondence and transmission of proofs. The second page should include a list of unusual symbols used in the article and the number of pages, tables and figures. It should also contain the keywords in English.

**Abstracts.** The text of the paper should be preceded by abstracts of no more than 200 words in English, in French and in German. Abstracts should contain the substance of the methods and results of the paper.

**Figures.** All illustrations are to be considered as figures, and each should be numbered in sequence with Arabic numerals. Original finished art work of all illustrations should be supplied. Each figure should have a caption and these should be listed on a separate sheet. Figures should be planned to fit one column of the printed page (7½ cm). Care should be taken that lettering on the original is large enough to be legible after reduction. Each figure should be identified. The approximate place of a figure in the text should be indicated in the margin.

**Equations.** Equations to which reference is made in the text should be numbered; equations should be referred to by citing the equation number enclosed in parentheses. Special care should be taken of those symbols which might easily cause confusion.

**Tables.** Tables should be typed on separate sheets. They should preferably be supplied as camera ready copy. Each table should have a number and a title. The approximate places for their insertion in the text should be indicated in the margin.

**References.** References are to be listed alphabetically by author at the end of the articles. Within the text, contributors should cite authors and years of publication in parentheses. References must be in the style shown below.

- Book* G. Fant (1960), *Acoustic Theory of Speech Production* (Mouton, The Hague), pp. 212-214.
- Journal* M. Schroeder (1967), "Determination of the geometry of the human vocal tract by acoustic measurements", *J. Acoust. Soc. Amer.*, Vol. 41, pp. 1002-1010.
- Conference Proceedings* B. Lindblom (1972), "Phonetics and the description of language", *Proc. 7th Internat. Cong. Phon. Sc. Montreal, 22-28 August 1971*, ed. by A. Rigault and R. Charbonneau (Mouton, The Hague), pp. 63-97.
- Contributed Volume* M. Halle and K.N. Stevens (1979), "Some reflections on the theoretical bases of phonetics", in *Frontiers of Speech Communication Research*, ed. by B. Lindblom and S. Öhman (Academic Press, London), pp. 335-349.

**Footnotes in text.** Footnotes in the text should be identified by superscript numbers and listed consecutively on a separate page.

Detailed guidelines for authors are available either from the publishers or from the Editor-in-Chief.

### SUBMISSION OF ELECTRONIC TEXT

In order to publish the paper as quickly as possible after acceptance, authors are encouraged to submit the final text also on a 3.5" or 5.25" diskette. Both double density (DD) and high density (HD) diskettes are acceptable. The diskette may be formatted with either MS-DOS/PC-DOS or with Macintosh DR. See the Notes for Electronic Text Preparation at the back of this issue for further information. The final manuscript may contain parts (e.g. formulae, complex tables) or last minute corrections which are not included in the electronic text on the diskette; however, this should be clearly marked in an additional hardcopy of the manuscript. Authors are encouraged to ensure that apart from any such small last-minute corrections, the disk version and the hardcopy must be identical. Discrepancies can lead to proofs of the wrong version being made.

### European Association for Signal Processing (EURASIP)

Administrating Committee

*President:* U. Heute, LNS/Techn. Fakultät/CAU, Kaiserstraße 2, 24143 Kiel, Germany

*Secretary-treasurer:* P. Grant, Electrical Engineering, Univ. of Edinburgh, Edinburgh EH9 3JL, UK

*Workshops Coordinator:* W. Mecklenbräuker, Institut für Nachrichtentechnik, TU Wien, Gußhausstraße 25/389, A-1040 Wien, Austria

*Brussels Liaison:* J. Vandewalle, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium

*Past President:* M. Bellanger, CNAM, 292 rue Saint-Martin, 75141 Paris Cedex 03, France

### European Speech Communication Association (ESCA)

*President:* Louis C.W. Pols, University of Amsterdam, The Netherlands

*Vice-President:* Roger Moore, DRA Malvern, UK

*Secretary/Treasurer:* Christian Benoit, ICP Grenoble, France

*Conferences:* Jens Blauert, Bochum University, Germany

*Publications:* Paul Dalsgaard, University of Aalborg, Denmark

*Industrial relations:* Roberto Billi, CSELT Turin, Italy

*ESCA Tutorial and Research Workshops:* Bjorn Granstrom, KTH Stockholm, Sweden

*Educations:* Isabel Trancoso, INESC Lisbon, Portugal

*CEC relations and Eastern Europe:* Joseph Mariani, LIMSI Orsay, France

ESCA is Europe's leading professional organization for Speech Science and Technology, covering all aspects of spoken language processing. ESCA is most visible through its NESCA Newsletter, the journal *Speech Communication*, its biennial Eurospeech conferences, and its regular ETRW workshops on special topics. For all correspondence, please contact ESCA, ICP-Université Stendhal, BP 25X, 38400 Grenoble Cedex, France. For further information, preferably use e-mail: [esca@icp.grenet.fr](mailto:esca@icp.grenet.fr)