



ELSEVIER

Speech Communication 19 (1996) 1–19

---

**SPEECH**  
COMMUNICATION

---

# Physiologically motivated modelling of the voice source in articulatory analysis/synthesis

Bert Cranen<sup>a,\*</sup>, Juergen Schroeter<sup>b</sup>

<sup>a</sup> Department of Language and Speech, Nijmegen University, P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands

<sup>b</sup> Acoustics Research Department, Rm 2C-576, AT&T Bell Laboratories, 600 Mountain Ave., Murray Hill, NJ 07974-063, USA

Received 25 February 1994; revised 29 December 1995

---

## Abstract

This paper describes the implementation of a new parametric model of the glottal geometry aimed at improving male and female speech synthesis in the framework of articulatory analysis synthesis. The model represents glottal geometry in terms of inlet and outlet area waveforms and is controlled by parameters that are tightly coupled to physiology, such as vocal fold abduction. It is embedded in an articulatory analysis synthesis system (articulatory speech mimic). To introduce naturally occurring details in our synthetic glottal flow waveforms, we modelled two different kinds of leakage: a ‘‘linked leak’’ and a ‘‘parallel chink’’. While the first is basically an incomplete glottal closure, the latter models a second glottal duct that is independent of the membranous (vibrating) part of the glottis. Characteristic for both types of leaks is that they increase dc-flow and source/tract interaction. A linked leak, however, gives rise to a steeper roll-off of the entire glottal flow spectrum, whereas a parallel chink decreases the energy of the lower frequencies more than the higher frequencies. In fact, for a parallel chink, the slope at the higher frequencies is more or less the same as in the no-leakage case.

## Zusammenfassung

Dieser Aufsatz beschreibt die Implementierung eines neuen parametrischen Modells der glottalen Geometrie. Unsere Arbeit zielt auf eine bessere Synthese männlicher und weiblicher Sprache im Rahmen von Systemen zur artikulatorischen Analyse/Synthese. Das Modell repräsentiert die glottale Geometrie als abhängig von den Zeitfunktionen der Querschnittsflächen am Ein- und Ausgang der Glottis. Die Steuerparameter des Modells sind stark an die Physiologie angelehnt, wie zum Beispiel glottale Abduktion. Unser Modell ist Teil eines artikulatorischen Analyse/Synthese-Systems. Um die im natürlichen Vorbild vorhandenen Details in der synthetischen Zeitfunktion des glottalen Strömungs zu reproduzieren, haben wir zwei verschiedene Arten von akustisch-wirksamen glottalen Lecks (Undichtigkeiten) implementiert: ein ‘‘verbundenes Leck’’ und eine ‘‘parallele Spalte’’. Während es sich im ersten Fall im wesentlichen um einen unvollständigen glottalen Verschluss handelt, stellt der zweite Fall einen zweiten glottalen Kanal dar, der unabhängig von dem knorpeligen (vibrierenden) Teil der Glottis ist. Charakteristisch für beide Lecktypen ist, dass sie die DC-Strömung und die Interaktion von Quellsignal und Ansatzrohr erhöhen. Ein verbundenes Leck bewirkt jedoch einen steileren Abfall des gesamten glottalen Strömungsspektrums; eine parallele Spalte hingegen erniedrigt die Energie der tieferen Frequenzen stärker als die der höheren Frequenzen. Tatsächlich ist es so, dass für eine parallele Spalte der Abfall bei höheren Frequenzen mehr oder weniger derselbe ist wie im Falle nicht vorhandener Lecks.

---

\* Corresponding author.

## Résumé

Cet article décrit l'implémentation d'un nouveau modèle paramétrique de la géométrie de la glotte qui a pour but d'amélioration, la synthèse de voix masculines et féminines dans le cadre d'une analyse/synthèse articulatoire. Le modèle représente la géométrie de la glotte en termes d'ondes glottiques d'entrée et de sortie. Il est contrôlé par des paramètres qui sont étroitement liés à la réalité physiologique, comme l'abduction vocale, par exemple. Ce modèle est inclu dans un système d'analyse/synthèse articulatoire visant une simulation articulatoire de la parole. Pour introduire dans les ondes synthétiques du flux glottique des détails que l'on observe au naturel, deux types de fuite différents ont été modélisés: la fuite couplée et la fuite parallèle. Alors que la première correspond, en principe, à une fermeture incomplète de la glotte, la deuxième correspond à la modélisation d'un conduit glottique supplémentaire, indépendant de la partie membranique (vibratoire) de la glotte. Ces deux types de fuite ont pour trait caractéristique commun d'augmenter le flux DC et l'interaction source/conduit vocal. Toutefois, une fuite couplée produit une pente plus forte du spectre du flux glottique, alors qu'une fuite parallèle réduit l'énergie plus fortement dans les basses fréquences que dans les hautes fréquences. En fait, pour une fuite parallèle, la pente dans les hautes fréquences est à peu près la même que celle qu'on observe dans les cas sans fuite.

**Keywords:** Glottal modelling; Glottal leakage; Articulatory synthesis of male and female speech

## 1. Introduction

Automatic analysis-by-synthesis using an articulatory speech synthesizer (speech mimicking) is an essential step towards using physiologically meaningful parameters in speech coding and synthesis (Schroeter and Sondhi, 1991; Gupta and Schroeter, 1993). One of the many questions that must be answered in this field is how to model the voice source in a parsimonious, but physiologically realistic fashion. In this paper we develop a parametric voice source which is intended to cover a wide range of voice qualities so that it is applicable to modelling the voice of many different, male and female speakers. The parameters comprise both physical dimensions in order to allow simulation of gender differences as well as non-dimensional parameters to describe source characteristics which are common for both sexes.

Ideally, a physiologically realistic model of the voice source should describe a self-oscillating system driven by appropriate aerodynamic forces. A well-known model that attempts to achieve this ideal is the one by Ishizaka and Flanagan (1972). However, at present we feel that there is not enough data to develop an accurate model of that type that is suited to be incorporated in an articulatory synthesizer. Among the factors that have been shown to affect the acoustic voice source signal are the mechanical properties of the vocal folds and surrounding tissues as well as the geometrical details of the glottal slit and the cavities and obstructions in its direct neighbourhood. Under what circumstances these factors play a role and to what extent they may interact is not yet clear.

At this point, we therefore take the standpoint that for the description of the acoustics of the voice source it is irrelevant how the glottal geometry has come about and that it is more fruitful to parameterize the time-varying geometry of the glottis directly. In other words, we prefer to keep away from the mechanical oscillator system and assume that the resulting glottis geometry can be described in some stylized form analogous to the one developed by Titze (1984). In doing so, our main assumption is that *all relevant acoustic features of the glottal source signal can be modelled adequately if geometric waveforms of the glottal inlet and outlet openings are parameterized in sufficient detail*. Note that this assumption also underlies the acoustic portion of the two-mass model approach (Ishizaka and Flanagan, 1972), and that by making this assumption, the problem of parameterizing the acoustic voice source is translated mainly into a problem of parameterizing glottal geometry.

The rest of this paper is organized as follows. In Section 2 we start out by summarizing the most important features of the parameterization proposed by Titze (1984). Next, on the basis of comparisons of real measured data with simulation results, we propose some extensions to this model in order to enable simulation of two

different types of glottal leakage (Cranen and Schroeter, 1995). Furthermore, we try to improve Titze's parameterization by providing extra controls for the glottal area derivatives at opening and closing in order to allow the elimination of some undesirable effects in the acoustic domain. This finally leads us to a model that describes glottal geometry in terms of inlet and outlet area *waveforms*. The waveforms are described in terms of power series of the control parameters proposed by Titze plus a few additional parameters. Subsequently, in Section 3, we set up the equations that constitute our synthesizer, which we have used to generate a few illustrative examples of how the most important control parameters, in particular the different forms of leakage, affect the acoustic excitation in our synthesizer when synthesizing stationary vowel sounds. Finally, in Section 4, we present our conclusions.

## 2. Glottal geometry parameterization

For any speech synthesizer the quality of the acoustic end product is the ultimate criterion by which the quality of the entire system is judged. From this perspective, it is important to realize that, since geometry and acoustics are related in a non-linear way, perceptually disturbing sounds may easily be created even when the glottal geometry seems to be modelled quite adequately. Therefore, since the acoustic excitation of the vocal tract takes place at instants when the glottal flow changes abruptly, we consider it particularly important to have explicit control of first and second order derivatives of the glottal inlet and outlet areas at the moments of opening and closing (also cf. Cranen and Boves, 1987).

The parameterization proposed by Titze (1984) was primarily designed to describe the most coarse characteristics of vocal fold motion. For instance, no attempt has been made to incorporate details like the changes in vocal fold geometry due to changing tissue compression at glottal closure and opening. As a consequence, the glottal opening and closing are handled in a similar fashion, and the resulting acoustic excitations at opening and closing of this model are equally strong. We believe that the control offered by Titze's model of the glottal area (derivative) at opening or at closing is too coarse for acoustic purposes and that it is not suited for use in an articulatory synthesizer without modifications. Nevertheless, Titze's parameterization does have some very nice properties that are worth preserving. Therefore, we will first give a brief summary of his model.

### 2.1. The Titze parameterization

Titze assumes that vibrating vocal folds can be described as a simple static structure upon which some time-varying modulations (representing the first normal vibrational mode of the tissue) are superimposed. With the  $\xi$ -axis in the lateral direction, the  $y$ -axis in the anterior/posterior direction and the  $z$ -axis along the direction of the flow (cf. Fig. 1), the static and dynamic displacements ( $\xi_0$  and  $\xi_t$ , respectively) are described by

$$\xi_0(y, z) = \left\{ \xi_{01} - \left[ (\xi_{01} - \xi_{02}) / d_g \right] z \right\} (1 - y/l_g), \quad (1)$$

$$\xi_t(y, z, t) = -\xi_m \sin(\pi y/l_g) \cos(2\pi F_0 t - \phi z/d_g), \quad (2)$$

with

- $\xi_{01}$  distance between posterior ends of folds at glottal inlet ( $z = 0$ ),
- $\xi_{02}$  distance between posterior ends of folds at glottal outlet ( $z = d_g$ ),
- $\xi_m$  amplitude of vibration,
- $\phi$  phase angle between lower and upper margins of the vocal folds,
- $F_0$  fundamental frequency =  $1/T$ ,  $T$ : fundamental period,
- $l_g$  length of glottis,
- $d_g$  depth of glottis.

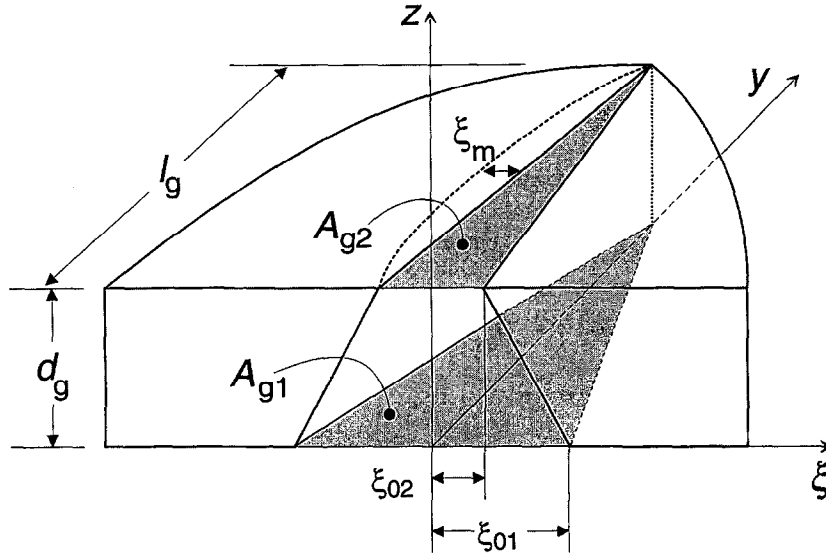


Fig. 1. Parameterization of glottal geometry after Titze (1984).  $A_{g1}$  and  $A_{g2}$  are the glottal inlet and outlet areas, respectively. Other notation see text.

Note that Eq. (2) differs from Titze's original in that we use “ $-\cos$ ” instead of “ $\sin$ ”. This only constitutes a different choice of time origin and has mainly historical reasons reflecting our preference to use the moment of minimum area (closure) as a reference.

By normalizing the static settings with respect to vibratory amplitude, three dimensionless parameters can be introduced:

$$\begin{aligned} \text{abduction quotient:} \quad Q_a &= \xi_{02}/\xi_m, \\ \text{shape quotient:} \quad Q_s &= (\xi_{01} - \xi_{02})\xi_m, \\ \text{vertical phase quotient:} \quad Q_p &= \phi/2\pi. \end{aligned}$$

These three parameters can be used to describe the displacement at each point of the fold from the midline as the sum of the quasi-static  $[\xi_0(y, z)]$  and the dynamic components  $[\xi_t(y, z, t)]$  of the glottal geometry:

$$\begin{aligned} \xi(y, z, t) &= \max\{0, \xi_0(y, z) + \xi_t(y, z, t)\} \\ &= \max\left\{0, \xi_m \left[ \left(1 - \frac{y}{l_g}\right) \left(Q_a + Q_s - Q_s \frac{z}{d_g}\right) - \sin\left(\pi \frac{y}{l_g}\right) \cos\left[2\pi \left(F_0 t - Q_p \frac{z}{d_g}\right)\right] \right] \right\}. \end{aligned} \quad (3)$$

The collision process is modelled by clipping all negative values of the resulting displacement at zero, as indicated by the max operation.

## 2.2. Glottal measurements

In order to claim that an articulatory synthesizer reflects physiological reality, it is mandatory that not only the speech signal is modelled correctly, but that also some “internal” signals assume realistic values and show characteristics that are observed in reality. As far as the voice source is concerned two such internal signals are important: transglottal pressure and glottal flow. In the following we describe some signal characteristics that, in our view, should be simulated correctly in order for the synthesizer to deserve the predicate “articulatory”.

From subglottal and supraglottal pressure measurements (e.g., Cranen and Boves, 1985) we infer that for male voices and normal speaking [at an average subglottal pressure of 8–10 cm H<sub>2</sub>O (784–980 Pa)], the

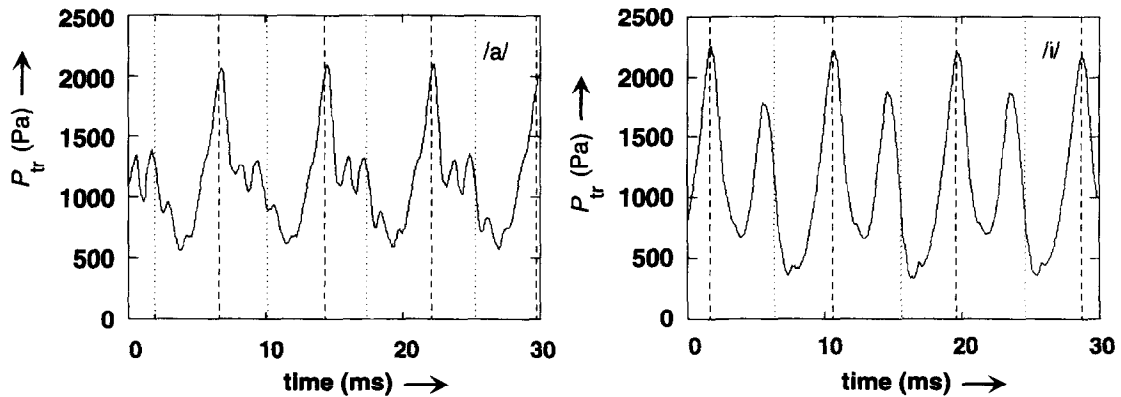


Fig. 2. Typical measured transglottal pressure waveforms during the vowels /a/ (left) and /i/ (right). The dashed lines denote instants of glottal closure, the dotted lines glottal opening. These instants were derived by marking the extrema in the first derivative of the EGG.

transglottal pressure has peak values occurring at glottal closure which are in the range of 15–30 cm H<sub>2</sub>O (1470–2940 Pa) while minima occur somewhere in the open glottis interval and range from 2–6 cm H<sub>2</sub>O (196–588 Pa) (cf. Fig. 2).

Similarly, from glottal flow estimates that may be obtained by inverse filtering oral flow recordings (cf. Fig. 3), we infer that dc-offset values of 100 cm<sup>3</sup>/s in combination with a relatively abrupt cessation of the glottal pulse are quite typical [also cf. (Cranen, 1990) and (Holmberg et al., 1988)]. Dependent on the loudness, normal peak-to-peak values for the glottal flow ( $U_g$ ) are about 300–600 cm<sup>3</sup>/s. In addition, we often note some unexplained ripples on the flow (derivative) waveforms in the (supposedly) closed glottis interval, which, in general, cannot be eliminated by fine-tuning the inverse filter.

### 2.3. Glottal leakage

One might be tempted to explain dc-offset flow solely by abduction of the vocal folds. However, in the model, abduction gives rise to an increased open quotient and a less abrupt opening and closing. In many measurements, of which Fig. 3 is an example, we observe a normal open quotient of approximately 50% as well

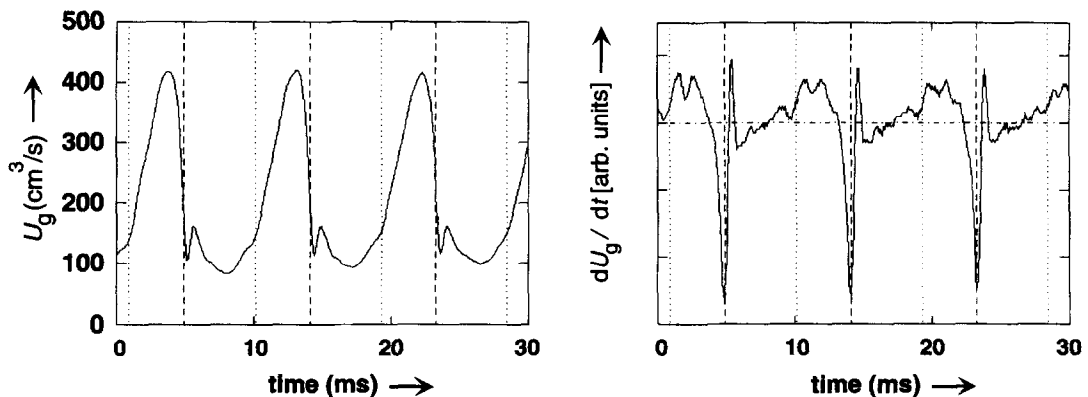


Fig. 3. A typical glottal flow estimate and its derivative during the vowel /ae/. The dashed lines denote instants of glottal closure, dotted lines glottal opening. These instants were derived by marking the extrema in the first derivative of the EGG.

as an abrupt closure in combination with an appreciable dc-offset in  $U_g$ . From simulations (Cranen, 1990; Cranen and Schroeter, 1995) we concluded that it is very unlikely that a leak flow of  $100 \text{ cm}^3/\text{s}$  at a normal open quotient is due to abduction. At the same time, we found that in these simulations realistic peak pressures for the transglottal pressure (cf. Fig. 2) could only be obtained if we assumed the vocal folds to close abruptly.

Evidently, a glottal leak flow may be caused by at least three different mechanisms (also cf. Södersten, 1994): (1) by abduction, (2) by an opening in the cartilaginous portion of the glottis which is usually denoted by ‘‘glottal chink’’, and (3) vertical tissue motions. In Cranen and de Jong (1995) it was argued that a leak flow of  $100 \text{ cm}^3/\text{s}$  is too large to be caused by the third mechanism alone. For that reason we will investigate mechanisms (1) and (2) more closely. Since abduction changes the waveform of the time-varying part of the glottal area whereas a glottal chink does not, the two types of leakage must be expected to have quite different acoustic consequences. Therefore, these two mechanisms must be carefully distinguished in a glottal geometry parameterization and an additional parameter is needed for controlling which mechanism to use.

To avoid confusion we introduce two terms denoting the different types of leakage. The terms are adopted from earlier unpublished work of Ishizaka (1989) by which our study was inspired. We will call a leak opening created by means of abduction (an opening that is necessarily connected to the membranous glottis) a **linked leak**. An opening that is virtually separated from the membranous glottis will be called a **parallel chink**. Note that a parallel chink can exist only when the folds are adducted. As soon as the folds are abducted a parallel chink inevitably has to become a linked leak.

#### 2.4. From geometry to waveform parameterization

If we require that our synthesizer be able to generate realistic glottal flow waveforms, i.e., waveforms with both a proper dc-offset, a reasonable open quotient, and a sufficiently abrupt change at glottal closure, it is necessary to *introduce a parallel chink*. We will do this by assuming a shunt tube of the same depth as the model of the membranous glottis with a cross-sectional area  $A_{c1}$  at the inlet and a cross-sectional area  $A_{c2}$  at the outlet. Note that this extension does not affect the parameterization of the membranous glottis in any way, although a way must be found to incorporate the change from a glottis with parallel chink to a configuration with linked leak if abduction increases.

The second extension we think is necessary, is to provide a *more flexible control of the area (and its derivative) at opening and closing*. Ideally, such an extension should reflect insights into the mechanisms that determine the changes of the shape of the vocal folds when they collide and separate again. As a consequence, it may seem most logical to modify Eq. (3) in such a way that instead of modelling the collision by clamping the  $\xi(y, z, t)$  at zero, the conservation of vocal fold tissue volume (i.e., its incompressibility) is properly accounted for. However, we feel that for our purposes such an approach would make the control over the acoustic behaviour of the model too indirect. Therefore, in order to be able to study the acoustic consequences of small area changes at opening and closing separately without being hampered too much by the peculiarities of some arbitrary stylization of geometry (as well as for computational efficiency reasons), we adopted an approach which is essentially a parametric description of glottal area inlet and outlet waveforms. We specify glottal geometry by describing both the membranous ( $A_{m1}$ ,  $A_{m2}$ ) and the chink components ( $A_{c1}$ ,  $A_{c2}$ ) of the cross-sectional area waveforms of the total glottal inlet and outlet and assume a linear variation of the areas in between.

In describing our area waveforms we have opted for a strategy in which we decompose the membranous glottal area waveforms into four components: the first component represents the dc-offset area, the second describes the time-varying part of the glottal area when the folds are freely moving, while the third and fourth component describe the area changes due to the collision process at opening and closing, respectively. All components can be described in terms of the dimensionless control parameters  $Q_a$ ,  $Q_s$  and  $Q_p$  thus keeping the link to Titze’s original model (and allowing refinement when additional knowledge becomes available).

As long as the folds are freely moving, the time-varying component of the area can be obtained by integrating  $\xi_i$  [Eq. (2)] with respect to  $y$ . Realizing that there are two folds, this leads to

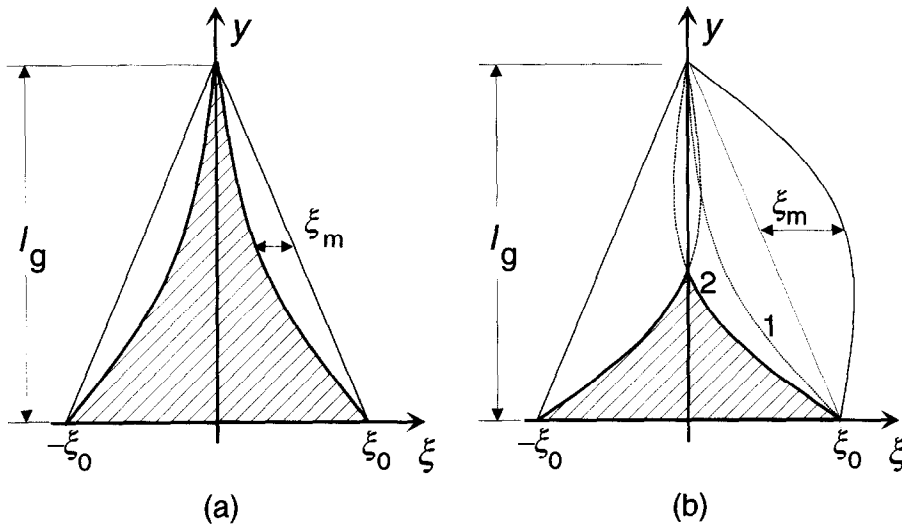


Fig. 4. (a) Minimum glottal area (cross-hatched) without collision, (b) computing minimum glottal area with collision. (1) starts when  $d\xi/dy|_{y=l_g} = 0$ , (2) increasing the excursion beyond a threshold given by (1) will diminish the minimum area.

$$A_t(z, t) = 2 \int_0^{l_g} \xi_t(y, z, t) dy = -\hat{A} \cos \left[ 2\pi \left( F_0 t - Q_p \frac{z}{l_g} \right) \right], \quad (4)$$

with  $\hat{A} = 4\xi_m l_g / \pi$ .

Of course, the glottal area is not a sinusoid throughout the entire glottal cycle. When the folds collide, the sinusoid will be truncated. To derive at what level the truncation happens, we make the following three observations (see Fig. 4(b)). First, for positive  $Q_a$ , collision can only occur during negative excursions of the vocal folds. Second, collision always starts at  $y = l_g$  and then proceeds zipper-like towards  $y = 0$ . Finally, the tangent of the folds at  $y = l_g$  must be in the direction of the  $y$ -axis. Hence, collision starts when

$$\left. \frac{d\xi}{dy} \right|_{y=l_g} = 0. \quad (5)$$

Assuming  $\xi \geq 0$  and, for example, letting  $z = d_g$  (i.e., considering the glottal outlet, denoted by the index ‘‘2’’) we obtain

$$\left. \frac{d\xi_2}{dy} \right|_{y=l_g} = -\frac{\xi_m}{l_g} \{ Q_a - \pi \cos[2\pi(F_0 t - Q_p)] \} = 0. \quad (6)$$

From the three observations stated above and Eq. (6) we conclude that collision (at the outlet) starts when  $\cos[2\pi(F_0 t - Q_p)] = Q_a / \pi$ . (For the glottal inlet,  $Q_a$  has to be replaced by  $Q_a + Q_s$ .) Hence, a first-order approximation of the areas of the membranous glottis is

$$A_{m1}(t) \approx A_{\min 1} + \max\{0.0, \hat{A}[(Q_a + Q_s)/\pi - \cos(2\pi F_0 t)]\}, \quad (7)$$

$$A_{m2}(t) \approx A_{\min 2} + \max\{0.0, \hat{A}[Q_a/\pi - \cos[2\pi(F_0 t - Q_p)]]\}, \quad (8)$$

where  $A_{m1}(t)$  is the area at the inlet of the membranous glottis ( $z = 0$ ), and  $A_{m2}(t)$  is the area at the outlet of the membranous glottis ( $z = d_g$ ).

As long as abduction is large enough for the folds not to collide (i.e., if  $Q_a + Q_s \geq \pi$  at the inlet, and  $Q_a \geq \pi$  at the outlet), the static dc-offset areas ( $A_{\min 1}$ ,  $A_{\min 2}$ ) that must be added to the time-varying

components in order to obtain the total glottal areas, may be found by integration of the static part of the fold-displacements [Eq. (1)] after which  $\hat{A}$  is subtracted. In this case, we find for the dc-offset area at the inlet ( $z = 0$ )

$$A_{\min 1} = (Q_a + Q_s) \xi_m l_g - \hat{A} = \hat{A} \left( \frac{Q_a + Q_s}{\pi} - 1 \right), \quad (9)$$

and at the outlet ( $z = d_g$ )

$$A_{\min 2} = Q_a \xi_m l_g - \hat{A} = \hat{A} \left( \frac{Q_a}{\pi} - 1 \right). \quad (10)$$

However, if the folds collide during the glottal cycle these minimum areas are intricate functions for which we were not able to find analytic expressions. Therefore, we synthesized a series of area waveforms with Titze's model and measured  $A_{\min 1}$  and  $A_{\min 2}$  manually. From these measurements we determined an approximation in terms of a power series. We found

$$A_{\min 1}(t) \approx \begin{cases} \xi_m l_g [-0.051 + 0.219(Q_a + Q_s) + 0.126(Q_a + Q_s)^2] & (Q_a + Q_s < \pi), \\ \xi_m l_g (Q_a + Q_s - 4/\pi) & (Q_a + Q_s \geq \pi); \end{cases} \quad (11)$$

$$A_{\min 2}(t) \approx \begin{cases} \xi_m l_g [-0.051 + 0.219Q_a + 0.126Q_a^2] & (Q_a < \pi), \\ \xi_m l_g (Q_a - 4/\pi) & (Q_a \geq \pi). \end{cases} \quad (12)$$

To reflect the fact that the dc-offset components of  $A_{m1}$  and  $A_{m2}$  can be considered as a linked leak and to make the notation more consistent with the notation used for the chink area ( $A_{c1}$ ,  $A_{c2}$ ), we will from now on denote  $A_{\min 1}$  by  $A_{l1}$  and  $A_{\min 2}$  by  $A_{l2}$ .

Our approach thus far describes the glottal area waveform as the sum of two components: a dc-offset value and a truncated sine wave. Using a truncated sine wave for modelling the time-varying portion of the glottal area is equivalent to modelling the collision process as an instantaneous event. This does not do justice to the fact that it will generally happen in a zipper like fashion (and thus it would constitute an even poorer approximation than the original Titze parameterization), nor does it take into account that tissue deformation may affect the glottal area as well. For obtaining smooth derivatives at opening and closure (in which we try to account for all collision effects without explicitly modelling them) we will now pursue our approach of superposing area components and add two additional (small) area components which have their centre of gravity exactly at the opening ( $t_o$ ) and closing ( $t_c$ ) instants. Thus, we obtain four components (cf. Fig. 5).

Instead of specifying the area directly, we prefer to specify the area derivative:

$$\frac{dA_g^{(m)}}{dt} = \Sigma \begin{cases} g_0(t) = C_0 & (0 \leq t < T), \\ g_2(t) = \begin{cases} \frac{d}{dt} \{ \hat{A} [1 - \cos(2\pi t/T)] \} & (t_o \leq t \leq t_c), \\ 0 & (0 \leq t < t_o \text{ and } t_c < t < T), \end{cases} \\ g_1(t) = \begin{cases} C_1(t - t_o + T_1)^3 & (t_o - T_1 \leq t \leq t_o), \\ C_1(t - t_o - T_1)^3 & (t_o < t \leq t_o + T_1), \\ 0 & (0 \leq t < t_o - T_1 \text{ and } t_o + T_1 < t < T), \end{cases} \\ g_3(t) = \begin{cases} C_3(t - t_c + T_3)^3 & (t_c - T_3 \leq t \leq t_c), \\ C_3(t - t_c - T_3)^3 & (t_c < t \leq t_c + T_3), \\ 0 & (0 \leq t < t_c - T_3 \text{ and } t_c + T_3 < t < T). \end{cases} \end{cases} \quad (13)$$



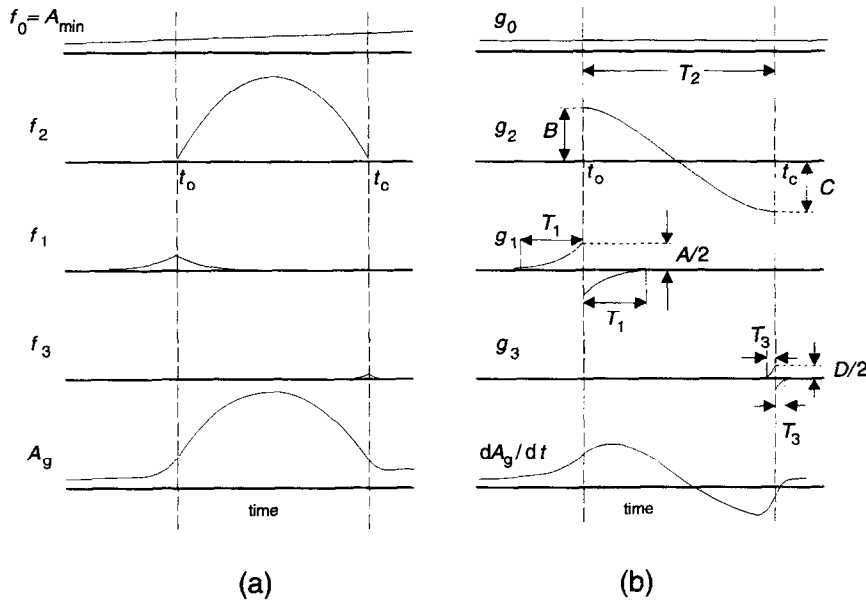


Fig. 5. The prototypes of the cross-sectional areas of the membranous glottis are built by superposition of four components (see text).

The first component ( $f_0$  in Fig. 5) represents the linked leak component ( $A_{l1}$  c.q.  $A_{l2}$ ). We allow this component to vary linearly over time within one glottal cycle, since in a situation where glottal parameters like abduction and vibratory amplitude may vary dynamically, this approach constitutes a simple way to construct  $A_g^{(m)}$ -waveforms that are continuous in time while the parameters are specified at discrete instants in time: we calculate the linked leak area at the reference points (moments of closure) and make a linear interpolation in between. As a consequence, the derivative ( $g_0$ ) is constant [ $C_0$  in Eq. (13)] over an entire period. Note that, if abduction changes very fast, our approach may give rise to discontinuities in the derivative of the glottal area at the boundaries of consecutive cycles, but since these will co-occur with the main excitation point they are expected to be relatively harmless from an acoustic point of view.

The second component ( $f_2$ ) is the truncated cosine wave of Eq. (8) and forms the core portion of the area waveform. It starts at  $t_o$  and ends at  $t_c$ ; its duration is denoted by  $T_2$ . Added on behalf of reviewer A (final review) Note that  $t_o$  and  $t_c$  are the instances where truncation takes place and thus  $T_2$  is implicitly controlled by  $Q_u$  in Eq. (7). If abduction is very large  $T_2$  will become equal to  $T$ ; if abduction is zero  $T_2$  will be  $0.5T$ . This component has a derivative [ $g_2$  in Eq. (13)] with discontinuities at  $t_o$  and  $t_c$ . In Fig. 5 the magnitudes of these discontinuities are denoted by  $B$  and  $C$ , respectively.

The third component ( $f_1$ ) is a symmetrical function centered around  $t_o$ . It extends over an interval  $T_1$  to the left and  $T_1$  to the right. Its derivative is denoted by  $g_1$  and has a maximum  $A/2$  (cf. Fig. 5). By allowing  $A$  to be different from  $B$  it becomes possible to construct area derivative waveforms that have a discontinuity at opening. Analogously, the fourth component ( $f_3$ ) is a symmetrical function centered around  $t_c$ . It extends over an interval  $T_3$  to the left and  $T_3$  to the right. Its derivative is denoted by  $g_3$ . By allowing  $D$  to be different from  $C$  one can introduce a discontinuity in the  $A_g^{(m)}$ -derivative at closure. In our simulations (cf. Section 3), however, we did not use this option, i.e. we always used  $A = B$  and  $C = D$ .

The constants  $C_1$  and  $C_3$  in Eq. (13) bear a direct relation to the constants  $A$  and  $D$  in Fig. 5: they determine the maximum amplitude of the area corrections at the instants of glottal opening ( $t_o$ ) and closing ( $t_c$ ), respectively. As can be easily verified from Eq. (13)  $A = C_1 T_1^3$  and  $D = C_3 T_3^3$ .

Note that the components  $f_1$  and  $f_3$  can be considered to model the area modulation due to the combined effects of zipper-like opening and closing as well as vocal fold tissue deformation. Adjusting the parameters  $C_1$ ,

$T_1$ ,  $C_3$  and  $T_3$  one can try to model the net effect of all these phenomena. Although we realize that it would be more natural to model the collision effects in the geometry domain, we feel that we might be hampered too much by the peculiarities of the geometry parameterization we have chosen. For this reason (as well as for computational efficiency reasons), we prefer to manipulate the area derivative waveforms directly.

To make our model comparable with the Titze model, we had to determine nominal values of  $T_1$  and  $T_3$  for the glottal outlet areas  $A_{g1}$  and  $A_{g2}$  as a function of abduction. To this end we took the output waveforms of the Titze model and measured the duration of the interval in which the area deviates significantly from a sinusoid. We did this for 21 values of  $Q_a$  in the range  $-1$  to  $4$  (i.e. from a pressed voice to a situation where the folds are abducted beyond the point that they do not collide anymore which is at  $Q_a = \pi$ ). Next we fitted a polynomial through these measured values. We found that for the outlet area the lowest order polynomial that gave an almost perfect fit to our measurement points was

$$\frac{T_1}{T} = \frac{T_3}{T} = \begin{cases} 10^{-2} (-0.16Q_a^5 + 0.44Q_a^4 - 0.76Q_a^3 + 2.16Q_a^2 + 3.45Q_a + 3.41) & (Q_a \leq \pi), \\ 0 & (Q_a > \pi). \end{cases} \quad (14)$$

A similar equation holds for the glottal inlet area  $A_{g1}$  with  $Q_a$  replaced by  $Q_a + Q_s$ .

Note that  $T_1$  and  $T_3$  in Eq. (14) are equal because the Titze model treats opening and closing similarly. If one would like to account for possible tissue deformation at opening and closing and avoid the behavior of the Titze model that the acoustic excitation is approximately equally strong at opening and closing (and which is in contradiction with normal practical findings), one may adjust  $T_1$  and  $T_3$  around the nominal values. By increasing  $T_1$  and decreasing  $T_3$  with respect to the nominal values (guided by the observations of our glottal flow simulations we arrived at typical values of  $T_1$  which are at least 20 times as large as  $T_3$ ), we created a workaround by means of which the acoustic behaviour of the model can be made more realistic while maintaining the link to the original Titze model. The symmetry of the final area waveform can be affected very strongly by choosing appropriate values for  $T_1$  and  $T_3$ . However, as long as one chooses  $f_2$  to be a truncated raised cosine wave,  $B$  and  $C$  (cf. Fig. 5) will be equal. This symmetry of the core portion might in some cases be considered too strong a restriction. We would like to point out that by applying some non-linear operator to the cosine function (or even better, by choosing a different function which reflects the fact that the oscillatory motion of the folds may contain higher modes as well) one can easily create a situation where  $B$  and  $C$  are not equal without affecting the basic experimentation scheme. Although we have experimented with such options and we found that it does add some extra flexibility in mimicking different voices we will not present any data about this in this paper, i.e., in the rest of the paper we will simply use the truncated cosine.

### 3. Synthesis

For synthesis purposes, it is necessary to describe how pressure changes in the various parts of the system due to viscosity, inertia, or due to a conversion of potential energy into kinetic energy (caused by the narrowing or widening of ducts). A blue-print of how to set up a set of equations which expresses transglottal pressure in terms of glottal flow and geometry can be found in (Ishizaka and Flanagan, 1972).

Obviously, for the calculation of the pressure drops in a system with a parallel chink, three different flows must be distinguished (viz., flow through the membranous glottis  $U_g$ , flow through the parallel chink  $U_c$ , and the flow  $U_{tot}$  that results when  $U_g$  and  $U_c$  have joined). Note that  $U_{tot} = U_g + U_c$  and that the flows  $U_g$  and  $U_c$  may occupy a region which has not necessarily the same dimensions as the glottis itself. In fact, as depicted schematically in Fig. 6, it seems reasonable to expect that in the case of a parallel chink the air flow splits some distance upstream of the actual glottal inlets and re-unites again at some distance downstream of the actual outlets. In deriving the equations, we discern three different regions.

*Region I.* The common entrance region in which there is only one flow component  $U_{tot}$ . We assume this flow to be uniform and laminar and the pressure drop, like in the original two-mass model, to be described by an

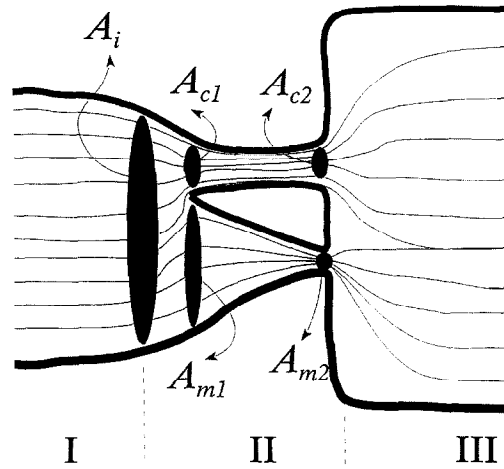


Fig. 6. Schematic cross-section of a glottis with parallel chink. The glottal flow is assumed to split some distance upstream. Downstream the glottal outlet the two flows will mix.

essentially lossless conversion of kinetic into potential energy. If it is considered necessary to account for contraction losses, this can be done by a simple correction factor [cf.  $\eta$  in Eq. (15)].

*Region II.* The glottal region which starts at the point where the flow splits into two separate components ( $U_g$  and  $U_c$ ) and ends at the glottal outlets. In this region the pressure distribution within the two flow channels are independent. Both components are assumed to be laminar and uniform as well. In the region, just upstream of the glottal entrance openings, the pressure drop is calculated in a similar way as in region I, but, of course, for each component individually. Within the membranous duct and the glottal chink viscous resistance and inertive effects are taken into account analogously to (Ishizaka and Flanagan, 1972).

*Region III.* The expansion region which starts directly at the glottal exits. The two flow components will exit their respective openings and mix until, finally, one uniform laminar flow is formed again which is attached to the vocal tract wall. It is assumed that the pressure in the pharynx is uniform and that the mixing process causes energy losses. The pressure difference across this region is approximated by assuming that momentum is preserved [analogously to (Ishizaka and Flanagan, 1972)].

The basic equation we use for calculating the (kinetic) pressure change  $\Delta P$  that is due to a flow  $U$  passing through a converging or diverging channel with inlet area  $A_1$  and outlet area  $A_2$  is given by

$$\Delta P = 0.5 \rho U^2 \left\{ \frac{1}{A_2^2} - \frac{1}{A_1^2} + \eta \left( \frac{1}{A_2} - \frac{1}{A_1} \right)^2 \right\}. \quad (15)$$

Note that the term with the  $\eta$ -coefficient accounts for energy losses [ $\eta = 0$  would correspond to a lossless conversion of potential into kinetic energy (Bernoulli flow);  $\eta = 1$  describes the situation where energy is lost but where momentum is preserved like in a strongly diverging duct].

It should be obvious that the transglottal pressure ( $P_{tr}$ , i.e. the pressure across region II in Fig. 6) can be expressed by three different equations, one for each flow path: (1) the membranous part of the glottis, (2) the parallel chink, and (3) the flow path through the subglottal and supraglottal system. Since we have three unknowns (viz.  $U_g$ ,  $U_c$  and  $P_{tr}$ ) these three equations are sufficient to describe our synthesizer.

To derive the first equation for  $P_{tr}$  we have assumed that for the inlet region Eq. (15) can be applied, and that for the expansion region, analogously to the approach in the original two-mass model, conservation of momentum must be assumed. Thus we find: *Pressure across regions I and III* ( $U_{tot} = U_g + U_c$ ):

$$P_{tr}(t) = P_{lung}(t) - [h_{sub}(t) + h_{supra}(t)] \otimes U_{tot}(t) \quad (\text{pressure change across acoustic load})$$

$$\begin{aligned}
& - (1 + \eta_i) \frac{\rho}{2} \frac{U_{\text{tot}}^2(t)}{A_i^2(t)} \quad (\text{pressure change across common inlet region}) \\
& - \frac{\rho U_{\text{tot}}^2(t)}{A_i^2(t)} + \frac{\rho U_g^2(t)}{A_{g2}(t) A_i(t)} + \frac{\rho U_c^2(t)}{A_{c2}(t) A_i(t)} \\
& \quad (\text{pressure change across common outlet region}).
\end{aligned} \tag{16}$$

For deriving the second and third equation it is important to realize that we assume a uniform velocity profile at the point where the total flow  $U_{\text{tot}}$  splits. This means that at that point the area occupied by  $U_g$  ( $A_{ig}$ ) and the area occupied by  $U_c$  ( $A_{ic}$ ) are related to the area that is occupied by the total flow ( $A_i$ ) in the following way:

$$\frac{U_{\text{tot}}}{A_i} = \frac{U_g}{A_{ig}} = \frac{U_c}{A_{ic}}. \tag{17}$$

Using Eq. (15) for describing the pressure loss at the entrance while using Eq. (17) to rewrite the inlet areas  $A_{ig}$  and  $A_{ic}$  in terms of  $A_i$ ,  $U_g$ ,  $U_c$  and  $U_{\text{tot}}$  we find: *Pressure across the membranous glottis (including the linked leak):*

$$\begin{aligned}
P_{\text{tr}}(t) &= \frac{\rho(\eta_i - 1)U_{\text{tot}}^2(t)}{2A_i^2(t)} + \frac{\rho(\eta_i + 1)U_g^2(t)}{2A_{g1}^2(t)} - \frac{\rho\eta_i U_g(t)U_{\text{tot}}(t)}{A_{g1}(t)A_i(t)} \\
& \quad (\text{pressure change across membranous inlet region}) \\
& + R_{gv}(t)U_g(t) + \frac{d[L_g(t)U_g(t)]}{dt} \\
& \quad (\text{pressure change due to viscous loss and inertance of the air}) \\
& + \frac{\rho}{2}U_g^2(t) \left\{ \left[ \frac{1}{[A_{g2}(t) + A_{l2}(t)]^2} - \frac{1}{[A_{g1}(t) + A_{l1}(t)]^2} \right] \right. \\
& \quad \left. + \eta_{l2}(t) \left[ \frac{1}{[A_{g2}(t) + A_{l2}(t)]} - \frac{1}{[A_{g1}(t) + A_{l2}(t)]} \right]^2 \right\} \\
& \quad (\text{kinetic pressure change due to widening/narrowing of duct}).
\end{aligned} \tag{18}$$

*Pressure across the parallel chink:*

$$\begin{aligned}
P_{\text{tr}}(t) &= \frac{\rho(\eta_i - 1)U_{\text{tot}}^2(t)}{2A_i^2(t)} + \frac{\rho(\eta_i + 1)U_c^2(t)}{2A_{c1}^2(t)} - \frac{\rho\eta_i U_c(t)U_{\text{tot}}(t)}{A_{c1}(t)A_i(t)} \\
& \quad (\text{pressure change across chink inlet region}) \\
& + R_{cv}(t)U_c(t) + \frac{d[L_c(t)U_c(t)]}{dt} \\
& \quad (\text{pressure change due to viscous loss and inertance of the air}) \\
& + \frac{\rho}{2}U_c^2(t) \left\{ \left[ \frac{1}{A_{c2}^2(t)} - \frac{1}{A_{c1}^2(t)} \right] + \eta_{l2}(t) \left[ \frac{1}{A_{c2}(t)} - \frac{1}{A_{c1}(t)} \right]^2 \right\} \\
& \quad (\text{kinetic pressure change due to widening/narrowing of duct}),
\end{aligned} \tag{19}$$

with  $A_{g1}(t)$  the time-varying part of the inlet area of the membranous glottis,  $A_{g2}(t)$  the time-varying part of the outlet area of the membranous glottis,  $A_{l1}(t)$  area of the linked leak at the glottal inlet,  $A_{l2}(t)$  the area of the linked leak at the glottal outlet,  $A_{c1}(t)$  the area of the parallel chink at the glottal inlet,  $A_{c2}(t)$  the area of the parallel chink at the glottal outlet,  $A_i(t)$  the area at which the entrance flow splits,  $h_{\text{sub}}(t)$  the subglottal impedance impulse response,  $h_{\text{supra}}(t)$  the vocal tract impedance impulse response,  $\otimes$  the convolution, and  $0 < \eta_i < 0.37$  the entrance loss factor,

$$\eta_{12}(t) = \begin{cases} 0.4 & \text{if } [A_{g1}(t) + A_{l1}(t)] \geq [A_{g2}(t) + A_{l2}(t)], \\ 1.0 & \text{if } [A_{g1}(t) + A_{l1}(t)] < [A_{g2}(t) + A_{l2}(t)], \end{cases}$$

loss coefficient to account for tapering of membranous glottis (cf. Ishizaka, 1983),

$$R_{v,g}(t) = 12\mu l_g^2 \int_0^{d_g} \frac{1}{[A_g(y,t) + A_l(y,t)]^3} dy,$$

$$R_{v,c}(t) = 12\mu l_c^2 \int_0^{d_g} \frac{1}{A_c^3(y,t)} dy,$$

viscous resistance of membranous glottis and parallel chink, respectively,

$$L_g(t) = \rho \left[ \frac{\Delta d_1}{(A_{g1} + A_{l1})} + \int_0^{d_g} \frac{1}{A_g(y,t) + A_l(y,t)} dy + \frac{\Delta d_2}{A_{g2} + A_{l2}} \right],$$

$$L_c(t) = \rho \left[ \frac{\Delta d_{1,c}}{A_{c1}} + \int_0^{d_g} \frac{1}{A_c(y,t)} dy + \frac{\Delta d_{2,c}}{A_{c2}} \right],$$

inductance of membranous glottis and parallel chink, respectively.

*Note.*  $\Delta d_{1,g} = 0.61\sqrt{(A_{g1} + A_{l1})/\pi}$ ,  $\Delta d_{2,g} = 0.61\sqrt{(A_{g2} + A_{l2})/\pi}$ ,  $\Delta d_{1,c} = 0.61\sqrt{A_{c1}/\pi}$  and  $\Delta d_{2,c} = 0.61\sqrt{A_{c2}/\pi}$  are end correction terms for a small tube ending in a large volume without baffle (cf. Beranek, 1986, p. 133.)

Using the relation  $U_{\text{tot}}(t) = U_g(t) + U_c(t)$ , Eqs. (16)–(19) constitute a set of three equations with three unknowns ( $P_v$ ,  $U_g$  and  $U_c$ ). To solve this set of equations, we used a discrete time representation where differential quotients were replaced by finite backward differences and where quadratic terms of the form  $U^2(n)$  ( $n$  denoting the time index) were approximated by  $2U(n-1)U(n) - U^2(n-1)$ . Furthermore, by using the same scheme adopted in (Sondhi and Schroeter, 1987) for calculating the convolution in terms of reflectances instead of impedances, Eqs. (16)–(19) were transformed into a set of three new equations which describe  $P_v(n)$  in terms of linear combinations of  $U_g(n)$  and  $U_c(n)$ . Thus, the synthesis problem is transformed into solving a set of linear equation (cf. Appendix A).

### 3.1. Some synthesis examples

Fig. 7 shows synthetic glottal waveforms and spectra for an /a/-tract load and an /i/-tract load, respectively. The bottom row applies to a linked leak, the upper row to a parallel chink. The thick solid lines apply to a glottis with leakage, while the dashed lines represent the no-leakage case as a reference. The thin solid lines around the zero dB line in the spectrum represent the difference between the (thick solid lines and dashed lines) spectra after those spectra have been normalized with respect to their maximum.

It can be seen that a leak has two effects. First, of course, a leak causes an extra dc flow. The leaks were chosen such that the dc-offset flow amounts approximately 100 cm<sup>3</sup>/s. Obviously, a dc-offset flow gives rise to an extra pressure drop in the lungs [we have modeled the lung resistance with 10 cgs acoustical Ohms ( $\approx 1$  MN/m<sup>5</sup>)] and as a consequence the overall intensity level drops as well. Second, and more interestingly, there

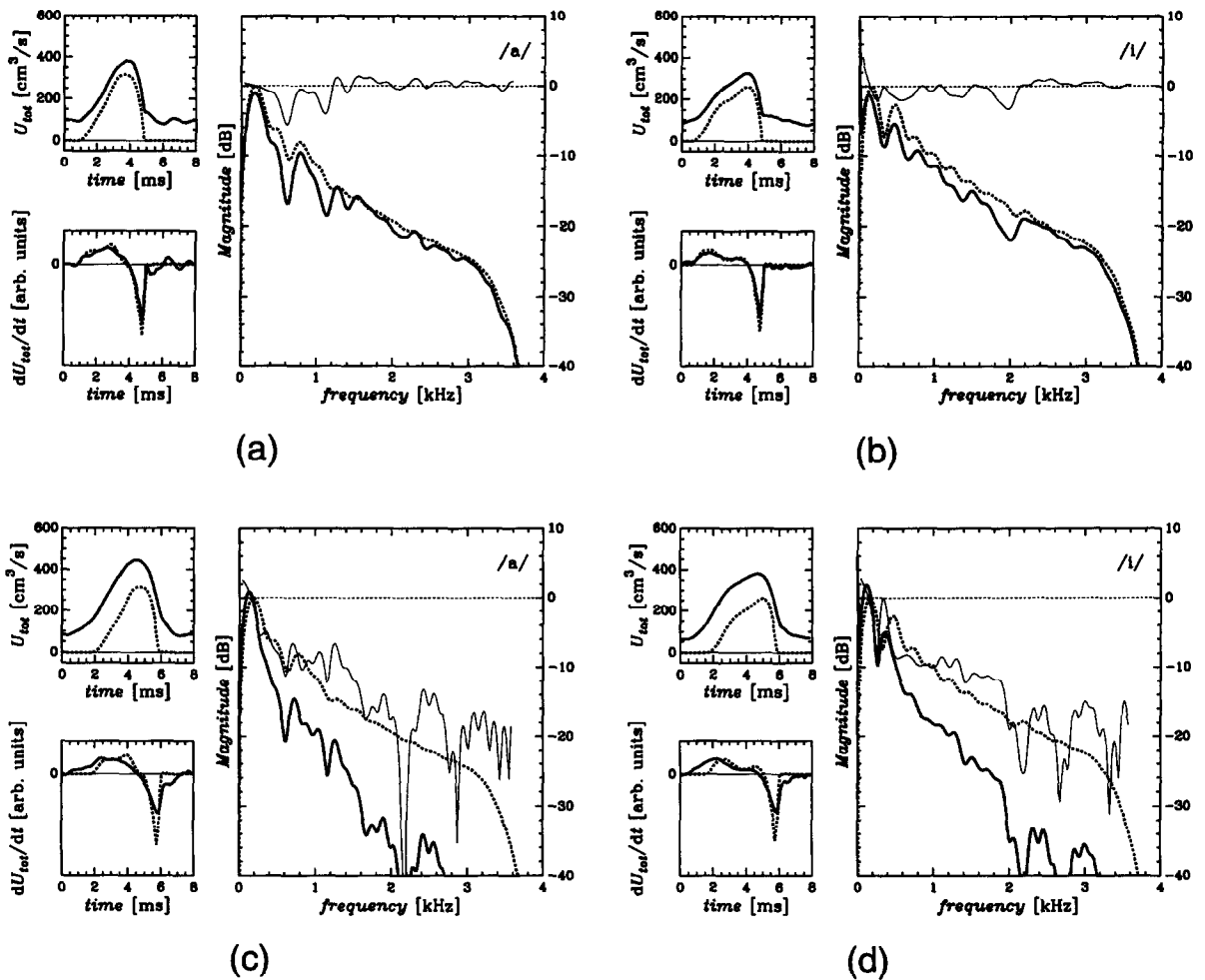


Fig. 7. Four figures, each showing waveforms of glottal flow ( $U_{tot}(t)$ ), flow derivatives ( $dU_{tot}/dt$ ), and glottal flow magnitude spectra. Figures (a) and (c) pertain to the vowel /a/, figures (b) and (d) to the vowel /i/. The dashed time-domain signals represent a no-leakage reference case; the solid lines pertain to a glottis with a leak (solid) causing a dc-offset flow of approximately  $100 \text{ cm}^3/\text{s}$ . In the top row ((a) and (b)) the leak has been modeled as a parallel chink of ( $\approx 0.03 \text{ cm}^2$ ); in the bottom row ((c) and (d)) the leak was created by abduction ( $Q_a = 0.78$ ). In this example  $A_l$  was chosen 1.05 times the total glottal inlet area, but this choice appears not to be critical. Note that the steep spectral roll-off above 3.3 kHz is due to low-pass filtering in the articulatory synthesizer.

are also changes to the spectral shape. Clearly visible is an increase of the interaction between voice source and acoustic load, reflected by the deepening of the zero's at the formant frequencies. Above remarks hold both for the linked leak as for the parallel chink.

Due to the fact that abduction affects the abruptness of glottal closure, the main difference between a linked leak and a parallel chink is the extent to which overall spectral slope is changed. A linked leak affects the spectral slope as a whole, attenuating the high frequencies the most. As may be inferred from Fig. 7, a parallel chink mainly affects the lower frequencies. Apparently, a parallel chink tends to act as a short circuit for the lower frequencies, while at the higher frequencies the spectral slope is hardly affected. Thus, if one compares the spectra neglecting absolute intensity levels and normalizing to some arbitrary reference level (as has been done creating the difference spectra in Fig. 7), one may in some occasions even get the impression that certain higher frequency regions get slightly boosted.

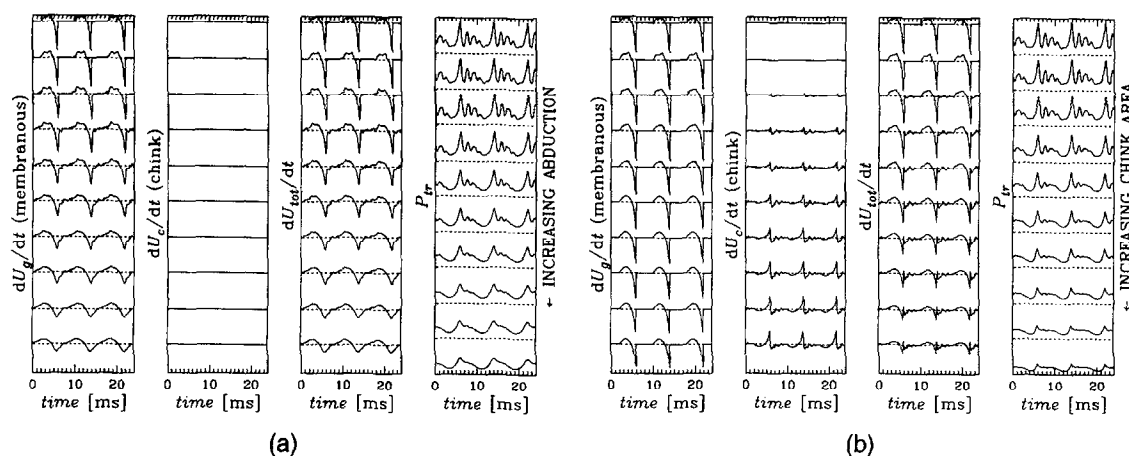


Fig. 8. The effect of an increasing leak (in each panel going from top to bottom). (a) Panels at the top show the effect of increasing abduction ( $-0.37 < Q_a < 1.5$ ) with no parallel chink present. (b) Panels at the bottom illustrate the effect of a parallel chink that increases gradually in size from 0–0.2 cm<sup>2</sup> while the folds are fully adducted. From left to right, the panels show the first derivatives of the membranous flow  $dU_g/dt$ , of the kchink flow  $dU_c/dt$ , and of their sum  $dU_{tot}/dt$ , and the transglottal pressure ( $P_{tr}$ ), respectively. The vocal tract load is that for an /a/.

Note that no glottal noise was modelled for deriving the result of Fig. 7. This means that if one would account for noise generation (and assuming that a dc flow would generate basically high frequency noise), the effect of high frequency boost due to the presence of a parallel chink must be expected to be more pronounced.

Whereas in the no-leakage case the interaction ripple can exist only in the open glottis interval, the acoustic interaction ripple for the leaky glottis exists throughout the entire glottal cycle. Since the acoustic excitation of the tract is generally strongest at the moment of “closure”, the interaction ripple in the “closed glottis” interval is also most clearly visible just after the moment of glottal “closure”. Compared to the no leakage case, the interaction ripple in the open glottis interval seems to decrease both for a linked leak (cf. Fig. 8(a)) as for a parallel chink (cf. Fig. 8(b)).

Fig. 8(b) can also be used to get a better understanding of the effects of a parallel chink. Because we have modelled the parallel chink impedance as a resistive component in series with an inductive component [cf. Eq. (19)] it is clear that the waveform of the chink flow is a filtered version of the transglottal pressure. Since the peak in transglottal pressure is caused by the main excitation of the subglottal and supraglottal cavities, it roughly coincides with glottal closure. Thus, it can be inferred that after the flow in the membranous portion of the glottis has completely stopped, the chink flow will cause a ripple in the total flow with a peak at or slightly after closure. Of course, the details of the ripple in the flow waveform and the corresponding spectrum will depend on how the value of resistive and inductive components of the chink impedance compares to the other impedances. However, it is worth noting that we did not observe a qualitatively different behavior of our model when different assumptions were made about the area where the flow splits [i.e., the specific value of  $A_i$  in Eq. (17)], the lung resistance, the loss coefficient in the entrance region, the end corrections applied, or the depth of the parallel chink, suggesting that the effects which are shown in Fig. 7 and Fig. 8 are quite robust with respect to geometrical details of the glottal region.

#### 4. Summary and conclusions

In this paper we have presented a parametric glottal geometry model that was designed to be capable of generating various voice qualities, to have a reasonable degree of physiological credibility, and to have enough

flexibility to be incorporated in an articulatory synthesizer. Although computational efficiency was not the primary goal, the implementation has been made so that in a later stage of the research, the parameters of the model can be estimated automatically and pitch synchronously in a speech mimic (cf. Schroeter and Sondhi, 1991; Gupta and Schroeter, 1993). Our desire to make control over the acoustic end product of our synthesizer as direct as possible explains why our glottal geometry parameterization is essentially a parameterization of glottal inlet and outlet area waveforms. By describing the major characteristics of our glottal area waveforms in terms of parameters as proposed by Titze (1984), we try to ensure maximal physiological interpretability.

Our model is an extension of Titze's glottal geometry parameterization in the sense that we (1) added a model for a parallel chink and (2) we allowed an explicit control of the area derivative, both at glottal closure and at glottal opening. The first extension was introduced in order to be able to explain characteristics observed in real glottal flow and transglottal pressure recordings. The second extension was thought necessary to allow independent control of acoustic excitations at glottal opening and closure (and in a very indirect way accounts for a more detailed modelling of the fold collision process).

When using our model to synthesize stationary vowels, the introduction of a leak appears to have two effects which are similar, irrespective how the leak was caused. First the extra opening gives rise to a dc-offset flow which, in combination with a non-zero lung resistance (which we think is the major resistive component of the acoustic load), causes the average subglottal pressure to decrease in comparison with the no-leakage case. Since a decreased subglottal pressure also decreases the average transglottal pressure and thus the flow pulse amplitudes, a leak decreases the overall intensity of the speech signal. The second effect of a leak is that the acoustic coupling between subglottal and supraglottal cavities becomes stronger. In the frequency domain the increased interaction is characterized by zeros in the glottal flow spectrum at the subglottal and supraglottal formants.

The main differences between a linked leak (caused by abduction) and a parallel chink is the way in which the spectral balance is affected. For a parallel chink, the spectral level in the lower part of the spectrum is decreased relatively more than the higher frequencies. The slope at the high end of the spectrum remains largely unchanged in comparison with the no-leakage condition. For a linked leak it is the other way around: the higher frequencies are attenuated most, and the first few harmonics have a larger amplitude. These observations can be related directly to the way in which glottal opening and closure takes place. Whereas abduction causes the opening and closing gestures of the folds to become more zipper-like, so that the derivative of the glottal flow are less impulse like and therefore the spectral slope becomes increasingly steep, the presence of a parallel chink hardly affects the time-varying part of the glottal flow (except for the extra ripple). Moreover, the peak-to-peak value of the sinusoidal part of the area becomes larger when abduction increases, causing an increase of the energy levels of the first few harmonics.

In the time-domain the increased interaction caused by a leak becomes visible as formant ripple in the glottal flow. This ripple may be appreciable both during the "open glottis" interval and the "closed glottis" interval. Our simulations indicate that an appreciable ripple in the "closed glottis" interval must be expected whenever glottal closure is abrupt. Clearly, a strong excitation of the acoustic load will cause large ripples in the transglottal pressure and consequently may induce ripple in the flow. This is the reason why for a given dc offset flow the ripple caused by a parallel chink is more pronounced than the ripple caused by a linked leak. Generally, we found that an increase of the ripple in the closed glottis interval is accompanied by a decrease of the ripple in the open glottis interval (as compared to the no-leakage condition).

The fact that the spectrum of the voice source is affected differently for a parallel chink and a linked leak has some interesting consequences. First it must be expected that voices which are characterized by the same amount of dc-offset flow but in which the type of leak is different, are clearly distinguishable perceptively. The higher harmonics of the source have different amplitudes and must be expected to mask (or to be masked by) any noise present to some degree. Hence, the role of dc-offset flow in voice efficiency measures, breathiness prediction, etc. cannot be defined in a straightforward manner. A second and unfortunate consequence is that our extension of Titze's model decouples to a certain extent flow and contact area simulations. EGG simulations



only involve the membranous part of the glottis, whereas glottal flow waveforms are influenced both by the membranous glottis and the parallel chink. In our view, this makes the simultaneous modelling of EGG and flow, as proposed in Titze (1984), less feasible in general.

Finally, we want to point out that the interaction ripples in the ‘‘closed glottis’’ interval of our simulated glottal flow waveforms are in many respects very similar to what is often observed in real inverse filtered flow waveforms (cf. Fig. 3 and Fig. 7). A better understanding of the criteria (like flatness of the closed glottis interval, absence of ripple, etc) that should be applied in order to decide whether an inverse filtering was successful or not seems needed.

## Acknowledgements

The major part of the work reported in this paper has been carried out at the Acoustics Research Dept. of AT&T Bell Laboratories. Furthermore, we would like to thank the anonymous reviewers for their thorough work. In particular the comments of one of them has substantially improved the readability of this paper.

## Appendix A. Conversion to a discrete time-domain and linearization

Going to the discrete time domain (index  $n$  denoting the current sample) and using the same scheme adopted in Sondhi and Schroeter (1987) for calculating the convolution in terms of reflectances instead of impedances, and applying the following approximations to linearize second-order terms:

$$\begin{aligned} U_g^2(n) &\approx 2U_g(n)U_g(n-1) - U_g^2(n-1), \\ U_c^2(n) &\approx 2U_c(n)U_c(n-1) - U_c^2(n-1), \\ U_{\text{tot}}^2(n) &\approx 2U_{\text{tot}}(n)U_{\text{tot}}(n-1) - U_{\text{tot}}^2(n-1), \\ U_g(n)U_c(n) &\approx U_g(n)U_c(n-1) + U_c(n)U_g(n-1) - U_g(n-1)U_c(n-1), \end{aligned} \quad (20)$$

Eqs. (16)–(19) can be rewritten as

$$\begin{aligned} P_{\text{tr}}(n) &= P_{\text{lunq}}(n) - \hat{P}_1(n) - Z_{\text{tot}}(n)U_{\text{tot}}(n) - 2[K_{it1}(n) + K_{et1}(n)]U_{\text{tot}}(n-1)U_{\text{tot}}(n) \\ &\quad + [K_{it1}(n) + K_{et1}(n)]U_{\text{tot}}^2(n-1) - 2K_{et2}(n)U_g(n-1)U_g(n) + K_{et2}(n)U_g^2(n-1) \\ &\quad - 2K_{et3}(n)U_c(n-1)U_c(n) + K_{et3}(n)U_c^2(n-1), \end{aligned} \quad (21)$$

$$\begin{aligned} P_{\text{tr}}(n) &= 2K_{ig1}(n)U_{\text{tot}}(n-1)U_{\text{tot}}(n) - K_{ig1}(n)U_{\text{tot}}^2(n-1) + 2K_{ig2}(n)U_g(n-1)U_g(n) \\ &\quad - K_{ig2}(n)U_g^2(n-1) + 2K_{ig4}(n)U_g(n-1)U_g(n) - K_{ig4}(n)U_g^2(n-1) \\ &\quad + K_{ig4}(n)U_c(n-1)U_g(n) + K_{ig4}(n)U_g(n-1)U_c(n) - K_{ig4}(n)U_c(n-1)U_g(n-1) \\ &\quad + \{R_{gv}(n) + 2\tilde{L}_g(n) - \tilde{L}_g(n-1)\}U_g(n) - \tilde{L}_g(n)U_g(n-1) + 2K_{kg}(n)U_g(n-1)U_g(n) \\ &\quad - K_{kg}(n)U_g^2(n-1), \end{aligned} \quad (22)$$

$$\begin{aligned} P_{\text{tr}}(n) &= 2K_{ic1}(n)U_{\text{tot}}(n-1)U_{\text{tot}}(n) - K_{ic1}(n)U_{\text{tot}}^2(n-1) + 2K_{ic2}(n)U_c(n-1)U_c(n) \\ &\quad - K_{ic2}(n)U_c^2(n-1) + 2K_{ic4}(n)U_c(n-1)U_c(n) - K_{ic4}(n)U_c^2(n-1) \\ &\quad + K_{ic4}(n)U_g(n-1)U_c(n) + K_{ic4}(n)U_c(n-1)U_g(n) - K_{ic4}(n)U_g(n-1)U_c(n-1) \end{aligned}$$

$$+ \{R_{cv}(n) + 2\tilde{L}_c(n) - \tilde{L}_c(n-1)\}U_c(n) - \tilde{L}_c(n)U_c(n-1) + 2K_{kc}(n)U_c(n-1)U_c(n) - K_{kc}(n)U_c^2(n-1), \quad (23)$$

with

$$\tilde{L}_g(n) = L_g(n)/t_s; \quad t_s \text{ sampling time interval,}$$

$$\tilde{L}_c(n) = L_c(n)/t_s,$$

$$P_1(n) = \hat{P}_1(n) + z_{\text{tot}}(n)U_{\text{tot}}(n),$$

$$\hat{P}_1(n) = \frac{1}{1 - r_{\text{in}}(0)} \sum_{k=1}^{N-1} r_{\text{in}}(k, n) [P_1(n-k) + z_0 U_{\text{tot}}(n-k)],$$

$r_{\text{in}}(m, n)$  inverse Fourier transform of input reflectance (assumed to vary slowly with time) with  $m$  denoting the lag index,

$$R_{\text{in}}(\omega, t) = [Z_{\text{in}}(\omega, t) - Z_0(\omega, t)] / [Z_{\text{in}}(\omega, t) + Z_0(\omega, t)],$$

$$Z_{\text{in}}(\omega, t) = Z_{\text{sub}}(\omega, t) + Z_{\text{tract}}(\omega, t),$$

$$z_{\text{tot}}(n) = \frac{1 + r_{\text{in}}(0, n)}{1 - r_{\text{in}}(0, n)} z_0(n),$$

$z_0(n) = \rho c A_1(n)$  characteristic impedance of the first section of the vocal tract model,

$$K_{i1}(n) = \frac{(\eta_i + 1)\rho}{2 A_i(n)^2}, \quad K_{e1}(n) = \frac{\rho}{A_1(n)^2},$$

$$K_{e2}(n) = -\frac{\rho}{A_{g2}(n) A_1(n)}, \quad K_{e3}(n) = -\frac{\rho}{A_{c2}(n) A_1(n)},$$

$$K_{ig1}(n) = \frac{(\eta_{ig} - 1)\rho}{2 A_i(n)^2}, \quad K_{ig2}(n) = \frac{(\eta_{ig} + 1)\rho}{2 A_{g1}(n)^2}, \quad K_{ig4}(n) = -\frac{\eta_{ig} \rho}{A_{g1}(n) A_i(n)},$$

$$K_{kg}(n) = \frac{\rho}{2} \left\{ \left[ \frac{1}{[A_{g2}(n) + A_{l2}(n)]^2} - \frac{1}{[A_{g1}(n) + A_{l1}(n)]^2} \right] + \eta_{l2}(n) \left[ \frac{1}{A_{g2}(n) + A_{l2}(n)} - \frac{1}{A_{g1}(n) + A_{l1}(n)} \right]^2 \right\},$$

$$K_{ic1}(n) = \frac{(\eta_{ic} - 1)\rho}{2 A_i(n)^2}, \quad K_{ic3}(n) = \frac{(\eta_{ic} + 1)\rho}{2 A_{c1}(n)^2}, \quad K_{ic4}(n) = -\frac{\eta_{ic} \rho}{A_{c1}(n) A_i(n)},$$

$$K_{kc}(n) = \frac{\rho}{2} \left\{ \left[ \frac{1}{A_{c2}^2(n)} - \frac{1}{A_{c1}^2(n)} \right] + \eta_{l2}(n) \left[ \frac{1}{A_{c2}(n)} - \frac{1}{A_{c1}(n)} \right]^2 \right\}.$$

These equations are now linear in  $U_g(n)$ ,  $U_c(n)$ ,  $U_{\text{tot}}(n)$  and  $P_{\text{tr}}(n)$ . By elimination of  $P_{\text{tr}}(n)$  and substitution of  $U_{\text{tot}}(n)$  by  $U_g(n) + U_c(n)$ , Eqs. (21)–(23) can be rewritten as two equations with two unknowns ( $U_g(n)$  and  $U_c(n)$ ) which can be solved using Cramer's rule.

## References

- L.L. Beranek (1986), *Acoustics* (Acoustical Society of America, New York).
- B. Cranen (1990), "Interpretation of EGG and glottal flow by means of a parametric glottal geometry model", *Proc. Internat. Conf. on Spoken Language Processing*, Kobe, Japan, pp. 65–68.
- B. Cranen and L. Boves (1985), "Pressure measurements during speech production using semiconductor miniature pressure transducers", *J. Acoust. Soc. Amer.*, Vol. 71, pp. 1543–1551.
- B. Cranen and L. Boves (1987), "The acoustic impedance of the glottis. Modelling and measurements", in: T. Baer, C. Sasaki and K. Harris, Eds., *Laryngeal Function in Phonation and Respiration* (College-Hill, Boston, MA), pp. 203–218.
- B. Cranen and F. de Jong (1995), "Glottal leakage studied by means of simultaneous video-stroboscopy and flow measurement", in: K. Elenius and P. Branderud, Eds., *Proc. XIIIth Internat. Conf. on Phonetic Sciences 1995* (Stockholm), Vol. 2, pp. 626–629.
- B. Cranen and J. Schroeter (1995), "Modeling a leaky glottis", *J. Phonetics*, Vol. 23, pp. 165–177.
- S. Gupta and J. Schroeter (1993), "Pitch synchronous frame-by-frame and segment based articulatory analysis-by-synthesis", *J. Acoust. Soc. Amer.*, Vol. 94, pp. 2517–2530.
- E.B. Holmberg, R.E. Hillman and J.S. 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.
- K. Ishizaka (1983), "Air resistance and intraglottal pressure in a model of the larynx", in: I.R. Titze and R.C. Scherer, Eds., *Vocal Fold Physiology* (Denver Center for the Performing Arts, Denver), pp. 414–424.
- K. Ishizaka (1989), Unpublished work.
- K. Ishizaka and J. Flanagan (1972), "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell Syst. Techn. J.*, Vol. 51, pp. 1233–1268.
- J. Schroeter and M.M. Sondhi (1991), "Speech coding based on physiological models of speech production", in: S. Furui and M.M. Sondhi, Eds., *Advances in Speech Signal Processing* (Marcel Dekker, New York), pp. 231–268.
- M. Södersten (1994), Vocal Fold Closure during Phonation, *Studies in Logopedics and Phoniatrics*, No. 3, Stockholm.
- M.M. Sondhi and J. Schroeter (1987), "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-35, pp. 955–967.
- I.R. Titze (1984), "Parameterization of the glottal area, glottal flow, and vocal fold contact area", *J. Acoust. Soc. Amer.*, Vol. 75, pp. 570–580.