WRESTLING THE TWO-MASS MODEL TO CONFORM WITH REAL GLOTTAL WAVE FORMS

Inger Karlsson and Johan Liljencrants

Dept of Speech Communication and Music Acoustics, KTH,
Box 70014, S-100 44 Stockholm, Sweden

ABSTRACT

This paper reports on some experiments to simulate female glottal wave shapes using a mechanical glottal model. The model is a modified mechanical two-mass model [9]. The simulated wave forms were compared to natural inverse filtered speech from one female speaker. Physiological data for the informant and from the literature were inserted in the model. Some properties of the model proved to be essential, for example bulging vocal cords and incomplete closure. The correspondence between simulated and natural wave forms is considerable.

INTRODUCTION

We will present a comparison between wave-forms produced by a mechanical model of the glottis and data from real speech. The mechanical glottal model used is inspired by the classical Ishizaka-Flanagan two-mass model [5], but the basic degrees of freedom are viewed as a lateral translation and a coronal rotation of the vocal cord body rather than lateral movements of successive caudal sections of it. This layout is easily interpreted as the motions of the cord body and of the mucosal cover which then governs the fundamental shape of the glottal passage as divergent or convergent. It is a distinctive feature of the model that these two movement components are mechanically uncoupled and independent. Instead they are coupled via the aerodynamics since the lateral movement is driven by the average glottal pressure and the rotation by the pressure gradient. The aerodynamic part of the modelling is based on classical theory and includes the Bernoulli effect and the influences on flow from viscosity and inertia in the airflow.

The control parameters for this model, like lung pressure, vocal cord masses and tensions can now be adjusted to accommodate various speakers and glottal articulations. A slightly earlier version of the model has been used for modelling male speakers [10]. The parameter values that were obtained in that study is compared to what this study has produced.

In this study the model is compared to observed aerodynamic data from female speech. As a first step towards imitating female glottal wave forms the dimensions of the model was altered to conform with typical female values. This includes realistic vocal cord length and mass. Further manipulations are performed to achieve as good fits with the real wave forms as possible, keeping in mind the particular vocal tract shape that generated the natural wave. Data on female speech has been collected from read sentences using inverse filtering of the differentiated air flow. The data also includes speech from speakers with different voice qualities. Air flow and subglottal pressure recordings are also available for the same speakers. So far, only one speaker has been modelled.

VOCAL CORD MODEL

The model used here is a variation on the classical two-mass model of Ishizaka and Flanagan [5] and was developed with the particular aim to control boundary movements in a simulation of glottal aerodynamics [8,9], and purports to be anatomically slightly more realistic than the Ishizaka-Flanagan model. A distinction between the two models is that the two resonators are here organised as a translational system and a superimposed rotational system, as shown in Fig. 1.

The primary model element is thus a mechanical resonator comprising a compliance $C=1/k$, a mass M, and a damping resistance R (not shown in the figure). For the translation we also include one more degree of freedom to allow for movement in the flow direction. The mass centres thus follow essentially elliptical paths, lateral and axial relative to the airflow. A more novel feature is the second resonance mechanism which is taken to be a rotational oscillation of the same mass M around its centre of gravity. Physiologically there may be some little rotation in the body of the vocal cord around its axis, but the explicit pur-

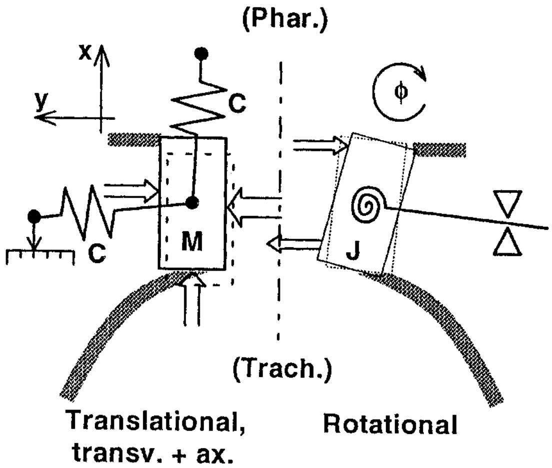

Fig. 1. Basic mechanical elements of the glottal model. Both sides have translational and rotational components, here separated for clarity.

pose of the rotor is here rather to model the wave motion in the cord cover in a maximally simple way. We can visualise the model mass as a translating and rotating rectangle, while the physiological counterpart rather is a translating mass with a surface layer sloshing axially back and forth to modulate convergence or divergence of the glottal passage. Our translational movement is in the terminology of Titze [12] called the 10 mode and our rotational movement corresponds to the 11 mode.

Control parameters

Among the parameters used to specify the model a first group relates to the cord anatomy such as the length measures of Fig. 2. The complete model also includes other parameters not used here which account for left-right asymmetry and second harmonic movement seen in the z direction.

A second parameter group relates to glottal articulation. Here the tension measure T relates to the cord modelled as an ideal string of length L and total mass M which at resonance at F would correspond to a stiffness $k = (2\pi F)^2 M = \pi^2 T/L$. In conjunction a ratio r of rotational to translational resonance frequencies defines a torsional stiffness assigned to the rotation. The rest gap V implements abduction gestures. The lung pressure P could also be included in this group rather than in the next.

A third parameter group describes oral articulation in terms of first and second formant frequencies and bandwidths, complemented with an oral area to define the load impedance level. A subglottal formant is specified for the trachea to introduce a source impedance for the lung pressure. The bandwidth of this formant is 0 for a fully closed glottis and increases with glottal opening.

The basic mechanical variables computed during simulation are the instantaneous deflections x, y, and rotation φ. From these instantaneous spatial amplitudes the active glottal surfaces are visualised as two ribbons with sinusoidal edges, moving like skipping ropes with a superimposed twisting, as in Fig. 2. The space between these ribbons defines the volume to be considered for the aerodynamics. In the present version of the model, the ribbons are also given a static bulge along the middle, shown in Fig. 2. The shape of the bulge is sinusoidal, the amplitude B decides the amount of bulging. The bulge will make the vocal cords

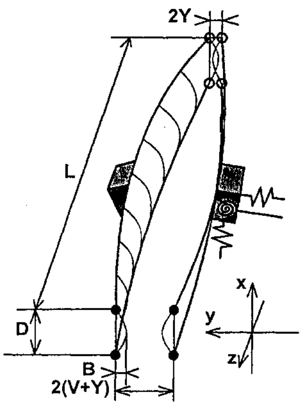

Fig. 2. Geometric development of glottis as the space between two sinusoidally edged ribbons.

roll rather than clap shut, this will result in a more rounded differentiated flow curve. This bulge is similar but not identical to one used by Titze [13] for modelling the vocal cord contact area. Apart from its influence on air flow, it was here also used to find a contact area which in turn controlled the amount of mechanical damping.

The driving input to the cord model follows from the aerodynamics. The aerodynamic component of the model was based on general knowledge of the pressure profile from classical work like that of Ishizaka and Flanagan [5], supplemented with experiences from detailed treatment with the Navier-Stokes equations, [9].

For studies of the model it has been installed in a test program which can scan arbitrary selected pairs of parameters. For each set of values the model is operated from a rest position until it has worked a specified time, and then measurements are done on multi-track records of the model behaviour. When the scan is complete these measurements can be summarised by iso-contours in the plane of the two parameters. Measurements performed include amplitude, pitch, jitter, shimmer, flow and flow derivative peak amplitudes, and open quotient.

INFORMANT DATA

The data from natural speech concern mainly one female speaker. All essential data were not available for this speaker, some data have been collected from the literature, while some had to be guessed.

Air flow measurements were performed for the female informant. Utterances of the syllable [pa] was recorded using a Rothenberg mask and simultaneously the subglottal pressure was measured during [p], [6]. The syllable was uttered in a normal voice. The glottal air flow was obtained by the inverse filtering. For the informant, signified by F1 in [6], the subglottal pressure was 5.7 cm H_2O and the peak glottal flow 200 ml/s.

The speech samples matched in the modelling experiments where taken from recordings that have been used for inverse filtering [7]. The samples were stressed short /a/ and /i/ vowels followed by /d/. The recordings were made in an anechoic chamber using phase-true equipment and sampled with 16 kHz. The speech was inverse filtered to get the differentiated glottal air flow, examples are given in Fig. 3. The formants and bandwidths for the two examples discussed in this paper were obtained with the inverse filter program. Resulting signal bandwidth for the inverse filtered wave form was 25-4000 Hz.

The subglottal formant was estimated to be slightly higher than the value measured for male speakers during the closed part of the glottal cycle, that is, 700 Hz as compared to the male value of about 550 Hz [1].

A realistic vocal cord length has been estimated from published data. In [4] the average vocal cord length for middle-aged Japanese women were measured as about 10 mm. S Hertegård and co-workers measured the vocal cord length during phonation at different pitches using tomography [3]. Their results vary between 11 and 12 mm for different female speakers and fundamental frequencies. A comparison between his speakers' sizes and fundamental frequencies and the present speaker gave a probable vocal cord length of 11 mm for the modelled speaker.

Other physiological parameters needed for the model are the mass and the thickness of the vocal cords. These are impossible to measure on a live, phonating subject. In the Japanese study mentioned above [4] the length ratio between male and female vocal cords was found to be about 2 to 1 and the thickness of the mucosa was about 10% higher in the male vocal cords. In conjunction with an estimated vocal cord height of 2 mm as compared with 3 mm used for the male vocal cord model, this gave an estimated vocal cord mass of about 0.05 gram as compared with 0.2 grams for the male model.

For most female speakers, the vocal cords do not close completely during phonation [11]. The speaker modelled here has, as can be seen from the air flow data in [6] a constant leakage of air. This indicates an incomplete vocal cord closure which has also been confirmed with fibre-optics registrations (not published). We chose to start with a abduction gap V of 0.2 mm. This seems to be a fairly good estimate for a normal speaker speaking at normal loudness according to [11].

EXPERIMENTS

These parameter values together with surmised values for other parameters were used as a starting point for the simulations. For some parameters, for example r, the male values were used.

The different parameters, normally two at a time, were then made sweeps of. The values that best fitted both the measured F0 and air flow and gave the most look-alike wave shape were chosen. This procedure had to be run through some times as different combinations of parameter values needed to be tested. Some unexpected differences between the male and the female values also cropped up, and, as we started with male values on some parameters, many sweeps had to be re-run.

Some major modifications compared to the version in [10] had to be introduced in the model, namely bulging vocal cords and varying damping factor at closure. These modifications make it possible to create wave shapes looking like natural glottal wave shapes.

The vocal cord thickness was first set to 4 mm as we felt that this would be equivalent to 2 mm thickness of non-bulging ribbon-like vocal cords. This thickness proved to be too large and was consequently lowered to 2 mm. The combination of bulging vocal cords and a rest gap V proved to be of crucial importance. So was the increase in energy damping during vocal cord closure. In the male case referred in [10] this factor was typically 0.2 (0 indicates no damping, 0.7 critical damping). In this experiment the damping factor varied with the area of vocal cord contact, at complete closure it reached the specified closure value, 0.5.

The subglottal formant frequency, about 700 Hz, proved to be critical especially for the [a] simulation where the first formant, 770 Hz, is close to the subglottal formant in frequency and both are near a harmonic of F0. F0 is about 250 Hz. The changes in the wave shape when the subglottal and the first formants are changed slightly are illustrated in Fig. 4. For [ɪ] the subglottal formant was set higher than in [a], 725 Hz compared with 700 Hz.

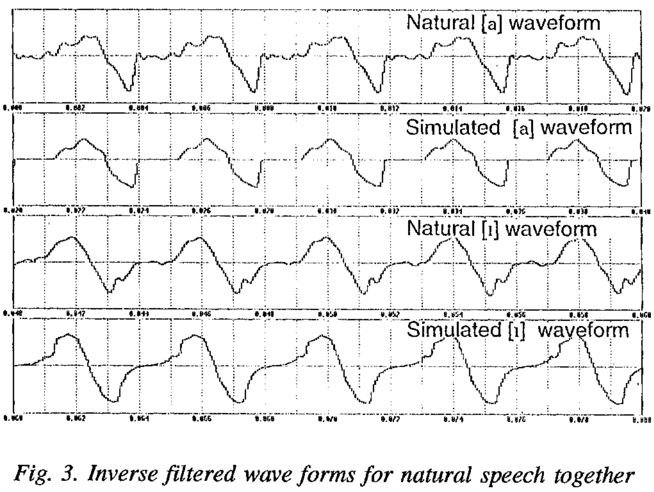

Fig. 3. Inverse filtered wave forms for natural speech together with simulated glottal wave forms. The largest differences can be seen for the closed part in [a] and during closing in [ɪ].

RESULTS

The end results of the simulations are shown in Fig. 3 and the parameter settings for these are given in Table 1. As can be seen, the simulations show a good agreement with the natural shapes, but there still exist some discrepancies especially in the closed part. We have tried to reproduce the crinkles in the closed part of the [a] and the small spike during the return phase of [ɪ], but this has so far proved impossible. A possible candidate for the explanation of the spike in [ɪ] is the parallel chink suggested by Cranen and Schröter [2]. We intend to study these problems further and will present results at the conference.

Male - female differences

A preliminary male-like wave-form simulation was performed using experiences gained in [10] and the present study. The different parameter settings are given in Table 1. Apart from the overt differences due to the gender-related size difference in the structures there are other parameter differences. The reasons for these differences will be studied further, they may have their origin in the physiology.

One such parameter was r, the ratio of rotational to translational resonance frequencies. r defines a torsional stiffness assigned to the rotational movements of the cords. For the simulations of male glottal wave forms [10], r was always around 2 but for the female speaker modelled in our

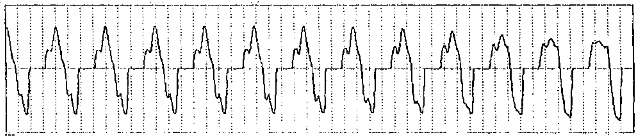

Fig. 4. Glottal wave forms for a sweep of the subglottal formant and the first formant in [a]. The subglottal formant sinks from 775 Hz at the left to 725 Hz at the right edge and the first formant increases from 725 Hz at the left to 775 Hz at the right edge.

Table 1. Parameter values used to produce natural-looking glottal waveforms for two vowels.

	unit	female [a]	female [ɪ]	preliminary male [ɪ]
damping at closure	prop.	0.5	0.5	0.5
damping at open	prop.	0	0	0
rotary to translational resonance frequency, *r*	prop.	1	1	2
gyration radius	mm	0.5	0.5	0.5
vocal fold tension, *T*	N	0.06	0.06	0.2
adduction force, fraction of P·D·L	prop.	-1	-1	0
lung pressure, *P*	Pa	600	600	1000
vocal fold mass, *M*	mg	30	30	200
vocal fold length, *L*	mm	10	10	20
vocal fold height, *D*	mm	2	2	3
rest gap, *Y*	mm	0.1	0.1	0.1
bulging, *B*	mm	0.2	0.2	0.1
abduction gap, *V*	mm	0.1	0.3	0.05
vocal tract mean area	cm^2	2	1	4
tracheal mean area	cm^2	2	2	4
subglottal formant	Hz	700	725	600
first formant	Hz	725	390	300
Q-value for this	-	10	5	5
second formant	Hz	1400	2400	2000

experiment the optimal value for *r* turned out to be 1.

Another example was the non-closing part, *V*. It was possible to model male wave forms assuming nearly complete vocal cord closure, *V* equals 0.05. To achieve a female wave form shape, *V* had to be larger.

CONCLUSION

The experiments here to simulate female glottal wave forms have been fairly successful. Some discrepancies between natural and simulated wave forms remain to be explained and modelled. In a comparison between male and female model parameter values, most of the differences can be explained by gender-related differences in size.

ACKNOWLEDGEMENT

This work has been supported by NUTEK, HSFR, KTH, FRN and the European CEC-ESPRIT project SPEECH MAPS.

REFERENCES

[1] Cranen, B. and Boves, L. (1987): "On subglottal formant analysis", JASA, Vol. 81, pp. 743-746

[2] Cranen, B. and Schröter, J. (1993): "Physiologically motivated modeling of the voice source in articulatory analysis/synthesis", Proc. Eurospeech '93, Berlin, pp. 95-98

[3] Hertegård, S., Håkansson, A., and Torstensson, Ö. (1993): "Vocal fold length measurements with computed tomography," Scand. J. of Logopedics and Phoniatrics, Vol. 18, pp. 57-63

[4] Hirano, M., Kurita, S., and Nakashima, T. (1981): "Growth, development and aging of human vocal folds," Proc. Vocal Fold Phys. Conf., eds D. Bless, J. Abbs, College Hill Press, pp. 22-43

[5] Ishizaka, K., Flanagan, J. L. (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell Syst. Tech. J., Vol. 51, pp. 1233-1268.

[6] Karlsson, I. (1988): " Glottal waveform parameters for different speaker types," Proc. of 7th FASE symposium, Edinburgh, pp. 225-231

[7] Karlsson, I. (1990): "Voice source dynamics for female speakers" Proc. 1990 ICSLP, Kobe, pp. 69-72.

[8] Liljencrants, J. (1989): "Numerical simulation of glottal flow", Proc. Vocal Fold Physiology, eds. J.Gauffin, B. Hammarberg, S. Felicetti, Singular Press, pp. 99-104

[9] Liljencrants, J. (1991): "A translating and rotating mass model of the vocal folds", STL-QPSR 1, pp. 1-18.

[10] Liljencrants, J. (1994): "Control of voice quality in a glottal model," Proc. Vocal Fold Physiology Conference, Kurume, ed. O Fujimura, forthcoming

[11] Södersten, M., Hertegård, S. and Hammarberg, B. (1994): "Glottal closure, airflow and voice quality in middle-aged women as related to changes in loudness", Phoniatric and Logopedic Progress Report, Dept. of Logopedics and Phoniatrics, Huddinge University Hospital, Vol. 9, pp. 3-20

[12] Titze, I. R. (1977): "On the mechanics of vocal-fold vibration", JASA, Vol. 60, pp. 1366-1380

[13] Titze, I. R. (1989): "A four-parameter model of the glottis and vocal fold contact area", Speech Communication, Vol. 8, pp. 191-201