A GLOTTAL CHINK MODEL FOR THE SYNTHESIS OF VOICED FRICATIVES

Benjamin Elie[1], Yves Laprie[1]

[1]LORIA, INRIA/CNRS/Université de Lorraine, Nancy, France
benjamin.elie@inria.fr, yves.laprie@loria.fr

ABSTRACT

This paper presents a simulation framework that enables a glottal chink model to be integrated into a time-domain continuous speech synthesizer along with self-oscillating vocal folds. The glottis is then made up of two main separated components: a self-oscillating part and a constantly open chink. This feature allows the simulation of voiced fricatives, thanks to a self-oscillating model of the vocal folds to generate the voiced source, and the glottal opening that is necessary to generate the frication noise. Numerical simulations show the accuracy of the model to simulate voiced fricative, and also phonetic assimilation, such as sonorization and devoicing. The simulation framework is also used to show that the phonatory/articulatory space for generating voiced fricatives is different according to the desired sound: for instance, the minimal glottal opening for generating frication noise is shorter for /z/ than for /ʒ/.

Index Terms: Articulatory synthesis, Vocal folds, Glottal chink, Voiced fricatives

1. INTRODUCTION

Articulatory synthesis [1–3], or artificial talkers, are a useful technique to study acoustic, articulatory or physiological phenomena involved in speech production. They are generally based on articulatory models [1, 3, 4], or waveguide modulation models [2], that enable the evolution of the vocal tract geometry to be accurately defined using a few parameters driving the position and the shape of the articulators (jaw, tongue, lips, velum...). In the aforementioned models, little attention is paid to the glottal source model: the glottal opening area is commonly imposed at the vocal tract input. This strong assumption restrains the consideration of the coupling between the vocal folds and the vocal tract. Hence the need of self-oscillating models for the vocal folds.

However, models for simulating the self-sustaining oscillations of the vocal folds [5–7] are generally based on the assumption that vocal folds vibrate along their whole length. Although this assumption suffices for numerous applications, it cannot account for partial closure of the glottis, which can be an issue for the study of voiced fricatives or breathy voice. Indeed, in these cases, the vocal folds vibrate solely along a portion of their length, producing a glottal leakage (or *glottal chink*) [8, 9].

This paper proposes to divide the glottal length into two parallel components: self-sustaining vocal folds and the glottal chink. To illustrate the method, the paper uses a 2×2 spring-mass system with smooth contours and a mobile separation point [6, 7]. The acoustic propagation is computed thanks to the single-matrix formulation [10]. Details are given in Sec. 2.

In comparison with previous works, which either used a parametric model of glottal opening area [9], or were connected to very simplified models of the vocal tract [11], restraining the model to be used in continuous speech synthesis context, the presented simulation framework uses a more realistic model of the glottis, and offers the possibility to connect the glottal system to realistic time-varying vocal tract geometries, including side cavities and length variation, to be used in continuous speech synthesis. This approach is intended to better account for the simultaneous production of frication noise and voiced source in voiced fricatives.

The consideration of the glottal partial closure in parallel with normally self-oscillating vocal folds is detailed in Sec. 3. It consists in connecting an acoustic side branch in parallel to the glottis, as suggested in [12]. Examples of synthesis of voiced sibilant fricatives are presented in Sec. 4.

2. THEORETICAL BACKGROUND

2.1. Self-oscillating model of the vocal folds

The model used to generate the self-sustained oscillations of the vocal folds (VF) is based on two spring-mass systems, representing the rear and front ends of the VF. It stems from the basic two-mass model [5]: it considers recent improvements, such as smmoth contours, allowing a mobile separation point [6, 7], and it adds corrective terms to take into account the viscous losses and the unsteady flow effects [13, 14].

At the flow separation point x_s, the pressure drop is given by Eq. (1).

$$P_{sub}(t) - P_{sup}(t) = R_b(t)U_g^2(t) + R_v(t)U_g(t) + \frac{\partial}{\partial t}\left[L_g(t)U_g(t)\right],$$

(1)

where

$$R_b(t) = \frac{\rho}{2l_g^2}\left[\frac{1}{h^2(x_s,t)} - \frac{1}{h^2(x_0,t)}\right], \qquad (2a)$$

$$R_v(t) = \frac{12\mu}{l_g}\int_{x_0}^{x_s}\frac{dx}{h^3(x,t)}, \qquad (2b)$$

$$L_g(t) = \frac{\rho}{l_g}\left[\int_{x_0}^{x_s}\frac{dx}{h(x,t)}\right], \qquad (2c)$$

are respectively the steady term of the Bernoulli equation, the Poiseuille corrective term and the unsteady term of the Bernoulli equation, $h(x)$ is the glottal opening along the x coordinates, l_g is the length of the vibrating part of the vocal folds. The position of the flow separation point x_s varies with the glottal constriction geometry:

- if $1.2h_1 > h_2$, $x_s = x_2$

- if $1.2h_1 < h_2$, x_s is such that $h_s = h(x_s) = 1.2h_1$

The value 1.2 is an *ad-hoc* criterion [6, 7].

2.2. Acoustic propagation

To compute the acoustic propagation inside the vocal tract (VT), the latter is modeled as a concatenation of cylindrical tubes (or *tubelets*). The length and the cross-sectional areas of the tubelets are such that they approximate the VT geometry. In this paper, the *transmission line circuit analog model* approach [10, 15, 16] is preferred to the *reflection type line analog models* approach [2, 17, 18], mainly because the latter does not include the possibility to easily deal with length variations of the VT. This is an important issue in realistic continuous articulatory synthesis since the length of the VT varies during natural speech production. Maeda [15], then Mokhtari *et al.* [10], showed that the acoustic propagation inside the vocal system seen as a waveguide network may be written into a matrix form

$$\mathbf{f} = \mathbf{Z}\mathbf{u}, \tag{3}$$

where $\mathbf{f} \in \mathbb{R}^{(N+1)}$ is a vector containing pressure forces, $\mathbf{Z} \in \mathbb{R}^{(N+1)\times(N+1)}$ is a tridiagonal matrix containing impedance and loss terms associated to each tubelets, and $\mathbf{u} \in \mathbb{R}^{N+1}$ is the vector containing the volume velocities inside each tubelet.

Considering the vocal folds model presented in Sec. 2.1, a quadratic term should be added to the set of linear equations to account for the Bernoulli resistance. The matrix formulation is then rearranged so that the quadratic equation depends only on U_1:

$$\mathbf{Z}^{-1}\mathbf{f} = \mathbf{I}\mathbf{u}_Z + \mathbf{Z}^{-1}\mathbf{Q}\mathbf{u}_Q, \tag{4}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{Q}$ is a square matrix the same size as $\mathbf{Z}$ having only one non-zero element, that is $Q_{(1,1)} = R_b$, and $\mathbf{u}_Q \in \mathbb{R}^{(N+1)} = [U_1^2, U_2^2, \ldots, U_N^2]^T$ is the vector containing the square power of the volume velocities. Eq. (4) is also valid in the case of a waveguide network, since it does not directly modify the coupling equations between different side cavities modeling the VT.

2.3. Frication noise generation model

The model to generate frication noise is important to approach the naturalness of the synthesized speech. The method presented in this paper uses bandpass filtered Gaussian white noise sources [19]. The amplitude of the noise source P_{n_i} at section i is

$$P_{n_i} = \max\left\{0, \xi w \left(Re^2 - Re_c^2\right) \frac{U_{DC}^3}{a_{i-1}^{3/2}}\right\}, \tag{5}$$

where ξ is an arbitrarily adjustable real constant used to control the noise level, and w is a random value between 0 and 1. Re is the Reynolds number of the air flow inside the vocal tract, Re_c is a threshold above which the air flow is turbulent [20], and consequently, above which the frication noise is generated. U_{DC} is the air flow volume velocity inside the vocal tract, and a_{i-1} is the area of the upstream tubelet.

The choice of Re_c is arbitrary. For this study, it is set to 1700. The computation of U_{DC} follows the method proposed by Maeda [19]. Note that to generate frication noise with high amplitude, the vocal tract should exhibit a narrow constriction (high Reynolds number), U_{DC} should be large (open glottis), and the noise source should be located directly downstream the constriction (small a_{i-1}).

3. A GLOTTAL CHINK MODEL

3.1. Presentation

Fig. 1 represents the geometrical model of the glottal chink, as well as the corresponding electric-acoustic analogy. In the model, the glottis partial closure is due to a partial abduction of the vocal folds and the chink is linked to the mobile part of the glottis. However, since the movements of the amplitude of the vocal folds oscillation is assumed to be small in relation to the partial abduction, the parallel branch assumption can be considered as valid, *i.e.* the geometry of the glottal chink should not be disturbed by the vocal folds oscillations. In Fig. 1, it is clear that the glottal chink area is $a_{ch} = l_{ch}h_{ab}$, where l_{ch} is the length of the glottal chink and h_{ab} is the abduction of the vocal folds.
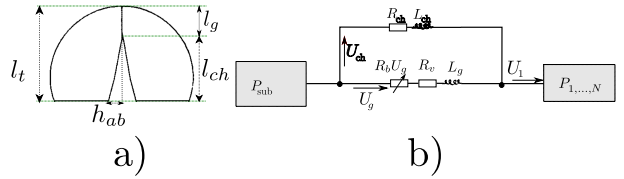
Fig. 1. a) View of the partial closure model of the glottis, after Cranen and Schroeter [8]. In this model, the partial closure is due to a partial abduction of the vocal folds. l_g is the length of the vibrating part of the vocal fold, l_{ch} is the length of the glottal chink and $l_t = l_g + l_{ch}$ is the total length of the vocal folds. the abduction of the vocal folds is assumed to be constant and is denoted by h_{ab}. b) Electric-circuit analogy of the partially closed glottis. U_{ch}, R_{ch}, and L_{ch} are the volume velocity through the glottal chink, the energy loss, and the air inertance inside the glottal chink, respectively.

3.2. Equations

The presence of the glottal chink modifies the first line of the system defined in Eq. (4). Indeed, one can write

$$P_1 - P_{sub} = R_b U_g^2 + [R_v + R_1] U_g + \frac{\partial}{\partial t}[L_g + L_1] U_g \tag{6}$$

$$+ R_1 U_{ch} + \frac{\partial}{\partial t} L_1 U_{ch} \tag{7}$$

$$P_1 - P_{sub} = [R_{ch} + R_1] U_{ch} + \frac{\partial}{\partial t}[L_{ch} + L_1] U_{ch} \tag{8}$$

$$+ R_1 U_g + \frac{\partial}{\partial t} L_1 U_g \tag{9}$$

$$P_1 = b_1 [U_g + U_{ch} - U_2 + S_1], \tag{10}$$

where R_1, L_1, b_1 and S_1 are the air resistance and the air inertance, wall losses and integrative constants, respectively, associated to the first tubelet of the vocal tract, *i.e.* at the glottal junction. One may refer to previous papers [10, 15] for details about their numeric computation, and to [12] for details about the integration of the glottal leakage side branch.

This yields to the following system

$$\left[\frac{\mathbf{f}_{VT}}{f_{ch}}\right] = \left[\frac{\mathbf{Z}_{VT} \mid C_{(VT,ch)}}{C_{(VT,ch)} \mid Z_{ch}}\right] \cdot \left[\frac{\mathbf{u}_{VT}}{U_{ch}}\right] + R_b U_g^2, \tag{11}$$

where $C_{(VT,ch)} = b_1 + R_1 + \frac{\partial}{\partial t}L_1$, and

$$Z_{ch} = b_1 + R_{ch} + R_1 + \frac{\partial}{\partial t}(L_{ch} + L_1),$$

is the impedance term associated to the coupling between the glottal chink and the vocal tract. The index VT denotes the vocal tract seen as a waveguide network. This method is compatible with the single-matrix formulation [10], as shown in [12].

4. NUMERICAL SIMULATIONS

4.1. Data used for the simulations

The contours of the vocal tract in the mid-sagittal plane were derived from X-ray films from the DOCVACIM database [21]. The area functions were obtained by dividing the vocal tract shape in tubelets perpendicular to the vocal tract centerline, and then applying α β transformations to recover the area [22]. A specific algorithm was designed purposely for determining the centerline [23].

4.2. Synthesis of French sibilant voiced fricatives

4.2.1. Synthesis of /z/

Fig. 2 shows the spectrum, the simulated output pressure, the glottal flow, the motion of the vocal folds and the Reynolds number computed during the simulation of a sustained /z/, in 3 configurations corresponding to 3 different glottal chink openings.
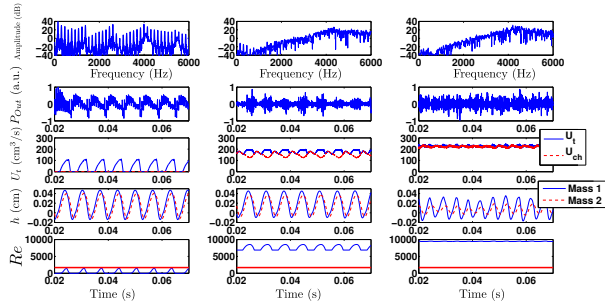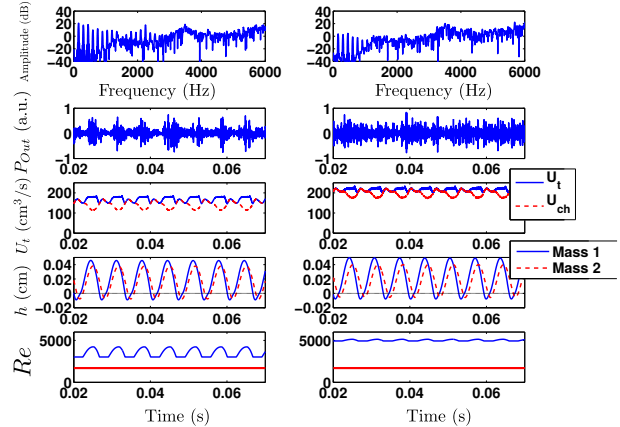
Fig. 2. Results of the simulation of a sustained /z/ for different glottal chink opening l_{ch}: 0 cm (left column), 0.3 cm (center), and 0.5 cm (right column). From top to bottom: narrow-band spectrum, output signal pressure P_{Out} radiated at the lips, glottal flow (solid line) and flow through the chink U_{ch} (dashed line), opening height of the vocal folds at x_1 (solid line) and x_2 (dashed line), and Reynolds number, with indication of the threshold (straight horizontal line).

The effect of the glottal chink can then be highlighted: the DC component of the glottal flow due to the chink raises the Reynolds number. When the glottal flow is sufficiently high, the Reynolds number is constantly above the threshold Re_c. The output pressure P_{Out} that is radiated at the lips is then the superimposition of the voiced source, generated by the glottal flow through the vibrating part of the vocal folds, and the frication noise source. This can be seen on the spectrum of the simulated audio signal: the voiced source, represented by harmonics, is predominant in the low-frequency domain, whereas the noise source is predominant in the mid- and the high-frequency domains, where the harmonic nature of the spectrum has disappeared. When the glottal chink length is sufficiently long, 0.5 cm in this case, the frication noise source is so predominant that the voiced source has almost no influence on P_{Out}. The simulated utterance sounds somehow like a devoiced /z/, namely /s/: the spectrum no longer exhibits harmonics, and the oscillating nature of some quantities, such as U_t and Re, disappears.

Note that the vocal folds still vibrate, but the oscillating motion is more erratic and has a smaller amplitude.

4.2.2. Synthesis of /ʒ/

Fig. 3 shows the spectrum, the simulated output pressure, the glottal flow, the motion of the vocal folds and the Reynolds number computed during the simulation of a sustained /ʒ/, in configurations corresponding to 2 different glottal chink openings.

Fig. 3. Results of the simulation of a sustained /ʒ/ for different glottal chink opening l_{ch}: 0.3 cm (left column), and 0.5 cm (right column). From top to bottom: narrow-band spectrum, output signal pressure P_{Out} radiated at the lips, glottal flow (solid line) and flow through the chink U_{ch} (dashed line), opening height of the vocal folds at x_1 (solid line) and x_2 (dashed line), and Reynolds number, with indication of the threshold (straight horizontal line).

Similarly to the simulation of /z/, the glottal chink opening impacts the voicing intensity of the simulated /ʒ/. Influence is quantitatively similar in both cases: the wider the glottal chink opening, the larger the noisiness. Nonetheless, there are quantitative differences: for a glottal chink opening of 0.5 cm, the spectrum of the simulated /ʒ/ still exhibits salient harmonics, while the latter were no visible in the spectrum of the simulated /z/. Therefore, for similar phonatory conditions, the sibilant fricative /z/ seems to tend to produce more frication noise, in relation to the voiced source, than the sibilant voiced fricative /ʒ/. Thus, when the glottal chink opening is too wide, the voiced fricatives tend to sound like the associated voiceless fricative. These results suggest that the partial abduction of the vocal folds during the production of voiced fricatives may be a cause for devoicing if the glottal gesture is not sufficiently well coordinated with the vocal tract configuration. Besides, this phenomenon seems to depend on the produced fricative, since preliminary results show that the devoicing of /z/ seems to be more likely than that of /ʒ/. This hypothesis is discussed in the next section via a more detailed investigation.

4.3. Voicing and noise onsets

This section presents a short study about the required glottal conditions to generate voiced fricatives. Based on a set of several area functions (17 for /z/ and 9 for /ʒ/) that have been computed from Xray images of the DOCVACIM database [21], the minimal length of the glottal chink for generating frication noise has been computed

as a function of the minimal constriction cross-section area. This quantity, denoted l_1, corresponds to the minimal length of the glottal chink from which the simulated Reynolds number is larger than the threshold, so that frication noise is generated. A second quantity, denoted l_2, is also computed: it is the minimal length of the glottal chink from which the simulated noise level is larger than the energy level generated by the voiced source. Finally, the voicing range $\Delta l = l_2 - l_1$ is computed. It is the range length of the glottal chink for which the generated utterance is considered as a voiced fricative.
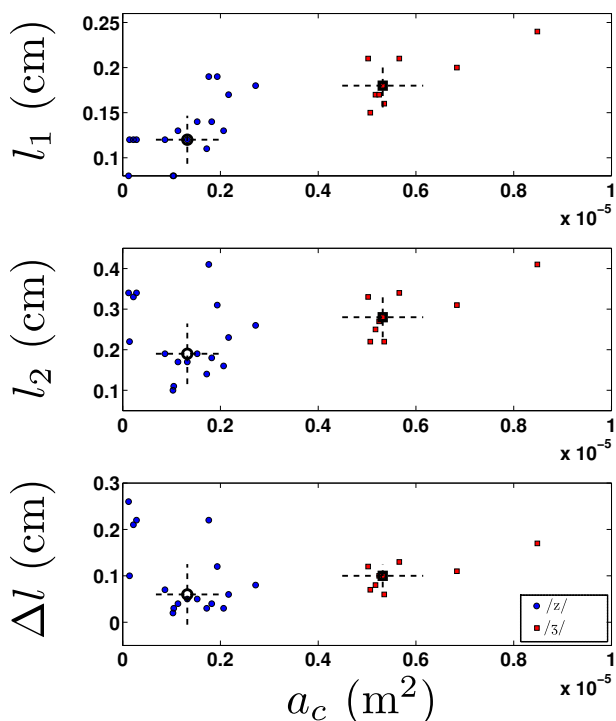
Fig. 4. Minimal length of the glottal chink for frication noise generation (l_1), minimal length of the glottal chink for predominant frication noise over voiced source (l_2), and length range of the glottal chink for voiced fricative generation (Δl), as a function of the minimal constriction area. Values obtained for different area functions corresponding to /z/ ('o'), and /ʒ/ ('□'). The median values of both group of fricative, as well as the median absolute deviations (dashed lines), are displayed.

Fig. 4 shows the values of l_1, l_2, and $\Delta l = l_2 - l_1$, as a function of the constriction area for both places of articulation, /z/ and /ʒ/. The cross-section area of the constriction seems to have an important influence on the frication onset l_1. Indeed, small constriction areas lead to small values of l_1. As a consequence, the minimal length l_1 is significantly smaller for /z/ than for /ʒ/, since the latter articulation exhibits a less narrow constriction.

Basically, the same phenomenon occurs for the minimal length l_2: the required length of the glottal chink so that the frication noise becomes predominant over the voiced source is shorter when the supraglottal constriction area is smaller. This leads to smaller values of l_2 for /z/ than for /ʒ/. Yet, it is not true for a few cases, especially when the constriction area is very small (less than 0.3 mm^2).

Results also show that the length range for generating voiced

sibilant fricatives is somewhat very small (usually smaller than 0.15 cm). Our simulations reveal that the range of the glottal chink length for generating a voiced fricative depends on the place of articulation and on the supraglottal constriction geometry. Consequently, if partially opening the glottis with a glottal chink may be considered as an articulatory gesture to generate voiced fricative, it should be done in a very precise way so that it generates frication noise without masking the voiced source. This tend to confirm that bad coordinations between the supraglottal constriction geometry and glottal configuration may be a cause for phonetic assimilation, such as sonorization and devoicing of fricatives.

5. CONCLUSION AND FURTHER WORKS

The paper has presented a glottis model for simulating voiced fricatives. It consists in an extension of classic self-oscillating models of the vocal folds to include the possibility of a partial glottal closure. The glottal leakage, or glottal chink, is acoustically modeled as a side branch in parallel with the vibrating part of the vocal folds. The model supports the connection with anatomically realistic models of vocal tract. This is a major contribution for realistic continuous speech synthesis.

The interest of the method has been shown via numerical simulations of sustained voiced fricatives /z/ and /ʒ/. The addition of the glottal leakage together with the self-sustaining model of the vocal folds enable the frication noise production mechanism to be realistically modeled. Indeed, the incomplete closure of the glottis imposes a DC component of the Reynolds number. If the latter is larger than the chosen threshold, then frication noise is generated aside the voiced source generated by the vibrating vocal folds.

Simulations also showed that the opening of the glottal chink has a great influence on the degree of voicing of fricatives. If there is not enough leakage, the frication noise may not be generated. On the contrary, a too wide open glottis may annihilate the voiced source of the fricative. Consequently, the incomplete glottal closure may be a cause of assimilation, whether sonorization, voicing or devoicing of fricatives. This result is confirmed by a short study revealing that the minimal length of the glottal chink from which the frication noise is predominant in the resulting speech signal is shorter for /z/ than for /ʒ/. This suggests that /z/ is more likely to be devoiced than /ʒ/.

The glottis model presented in this paper is thus a base to investigate more thoroughly the acoustic impact of the partial closure of the glottis. For instance, in the next future, the boundaries of the phonatory/articulatory space that generates voiced and voiceless fricatives might be better defined thanks to this model.

6. REFERENCES

[1] P. Birkholz, D. Jackèl, and B. J. Kröger, "Simulation of losses due to turbulence in the time-varying vocal system," *IEEE Tansaction on Audio, Speech and Language Processing*, vol. 15(4), pp. 1218–1226, 2007.

[2] B. H. Story, "Phrase-level speech simulation with an airway modulation model of speech production," *Computer Speech & Language*, vol. 27(4), pp. 989–1010, 2013.

[3] Y. Laprie, R. Sock, B. Vaxelaire, and B. Elie, "Comment faire parler les images aux rayons X du conduit vocal (How to make X-ray images speak)," in *SHS Web of Conferences*. EDP Sciences, 2014, pp. 1285–1298.

[4] S. Maeda, "Un modele articulatoire de la langue avec des composantes linéaires (An articulatory model of the tongue with linear components)," in *Actes 10emes Journées d'Etude sur la Parole*, 1979, pp. 152–162.

[5] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.*, vol. 51(6), pp. 1233–1268, 1972.

[6] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, and Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.*, vol. 96(6), pp. 3416–3431, 1994.

[7] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, and A. Hirschberg, "A symetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prothesis design," *Acta Acustica*, vol. 84, pp. 1135–1150, 1998.

[8] B. Cranen and J. Schroeter, "Modeling a leaky glottis," *Journal of Phonetics*, vol. 23, no. 1–2, pp. 165 – 177, 1995.

[9] ——, "Physiologically motivated modelling of the voice source in articulatory analysis/synthesis," *Speech Communication*, vol. 19(1), pp. 1–19, 1996.

[10] P. Mokhtari, H. Takemoto, and T. Kitamura, "Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches," *Speech Communication*, vol. 50(3), pp. 179 – 190, 2008.

[11] R. Wilhelms-Tricarico, "A modified two-mass model of the vocal folds with a chink and gradual closure," 1994, speech Communication Group Working Papers.

[12] B. Elie and Y. Laprie, "Extension of the single-matrix formulation of the vocal tract: consideration of bilateral channels and connection of self-oscillating models of vocal folds with glottal chink," Sep. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01199792

[13] C. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems, "Experimental validation of a quasi-steady theory for the flow through the glottis," *J. of Sound and Vibration*, vol. 276(3–5), pp. 475 – 490, 2004.

[14] L. Bailly, X. Pelorson, N. Henrich, and N. Ruty, "Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation," *J. Acoust. Soc. Am.*, vol. 124(5), pp. 3296–3308, 2008.

[15] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Communication*, vol. 1, pp. 199–229, 1982.

[16] P. Birkholz and D. Jackèl, "Influence of temporal discretization schemes on formant frequencies and bandwidths in the time-domain simulation of the vocal tract system." in *Proc. of the Interspeech 2004-ICSLP*, 2004, pp. 1125–1128.

[17] J. L. Kelly and C. C. Lochbaum, "Speech synthesis," in *Proceedings of the Fourth International Congress on Acoustics*, 1962, pp. 1–4.

[18] P. Meyer, R. Wilhelms, and H. W. Strube, "A quasiarticulatory speech synthesizer for german language running in real time," *J. Acoust. Soc. Am.*, vol. 86(2), pp. 523–539, 1989.

[19] S. Maeda, "Phoneme as concatenable units: VCV synthesis using a vocal tract synthesizer," in *Sound Patterns of Connected Speech: Description, Models and Explanation, Proceedings of the symposium held at Kiel University, Arbeitsberichte des Institut für Phonetik und digitale Spachverarbeitung der Universitaet Kiel:31*, 1996, pp. 145–164.

[20] M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust. Speech Sig. Process.*, vol. 35(7), pp. 955–967, 1987.

[21] R. Sock, F. Hirsch, Y. Laprie, P. Perrier, B. Vaxelaire, G. Brock, F. Bouarourou, C. Fauth, V. Hecker, L. Ma, J. Busset, and J. Sturm, "DOCVACIM an X-ray database and tools for the study of coarticulation, inversion and evaluation of physical models," in *The Ninth International Seminar on Speech Production - ISSP'11*, Canada, Montreal, 2011, pp. 41–48.

[22] A. Soquet, V. Lecuit, T. Metens, and D. Demolin, "Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI," *Speech Communication*, vol. 36(3), pp. 169–180, 2002.

[23] Y. Laprie, M. Loosvelt, S. Maeda, E. Sock, and F. Hirsch, "Articulatory copy synthesis from cine X-ray films," in *Interspeech 2013 (14th Annual Conference of the International Speech Communication Association)*, Lyon, France, 2013, pp. 1–5.