PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

On the production mechanisms of the singer's formant

Bernd J. KRÖGER

Department of Phoniatrics, Pedaudiology and Communication Disorders, RWTH Aachen University, Germany

ABSTRACT

A small extra tube directly located above the glottis (laryngeal tube), which is about six times smaller in cross-sectional area than the pharyngeal-oral tube section and which has a length of about 1/6 of the overall length of the vocal tract tube seems to be responsible for producing a broad (hyper-)formant at about 3 kHz, called the singer's formant. From an acoustic viewpoint the singer's formant has been described as a clustering of the third, fourth and fifth formant. Using a computer simulation model for sound production, which includes a self-oscillating vocal fold model (source model), an acoustic wave propagation and radiation model (tube model or filter model) as well as source-filter interaction mechanisms it is possible to show that a further mechanism is responsible for increasing the amplitude of partials in the frequency region of 3 kHz. Our simulations indicate that in the case of normal to high subglottal pressure (normal speaking to loud singing voice) a specific source-filter interaction mechanism is an important contributor for increasing partials in the region of 2 kHz to 5 kHz.

Keywords: Music, Singing, Voice, Singer's Formant, Synthesis

1. INTRODUCTION

It is well known that female as well as male opera singers are able to form a singer's formant (1, 2, 3). This formant is located at about 3 kHz and is strong in intensity. The singer's formant leads to a much smaller decay of the spectrum between 2 and 5 kHz as in the case of speaking voices or in the case of singing by amateurs or in case of singing in other singing styles. The huge energy concentration in the frequency region between 2 kHz and 5 kHz allows the opera singer to appear with a very sustainable voice. The singer is thus well perceivable without microphone and in case of an accompanying classical orchestra.

From the production perspective the source of the singer's formant seems to be a small extra tube right above the glottis (laryngeal tube, see Sundberg (1)) together with an abrupt change in cross sectional area at the end of this tube, i.e., at the transition of this laryngeal tube to the pharyngeal cavity. It is assumed that the cross sectional area of the laryngeal tube is about 6 times less than the mean cross sectional area of the rest of the vocal tract (pharyngeal and mouth cavity), thus forming an extra formant because of its acoustic mismatch (high change in acoustic impedance) between this small laryngeal tube and the main tube representing the vocal tract (pharyngx and oral region). Further studies were focused on the acoustic characteristics of the vocal tract transfer function in the case of opera singers and postulate a clustering of the third, fourth and fifth formant of the vocal tube as origin of the singer's formant as well as interaction effects between F1 and partials of F0 (2, 3).

The experiment described in this paper as well underpins the importance of a small laryngeal cavity followed by a wide pharyngeal cavity but emphasizes the fact that the increase of formants in the region between 2 kHz to 5 kHz is mainly related the so called "source-filter interaction" or "glottis-vocal tract interaction" (4, 5, 6, 7).

Source filter interaction can be differentiated in three subcases, i.e., the case of acoustic-aerodynamic, mechanic and no interaction. Acoustic interaction reflects the fact that the backward travelling acoustic waves reflecting the standing waves (formants) influence the glottal flow at the level of the glottal tube. This interaction mainly causes formant ripple within the opening phase of the glottal pulse (6). A major aerodynamic interaction effect is that the peak of the glottal flow pulse is shifted to the right side in comparison to the peak of the glottal area pulse shape. This effect results from the inertia of the subglottal-supraglottal air column and generates a specific flow shape mainly in the closing portion of the glottal cycle, i.e., more abrupt cutting off of glottal flow at the moment of glottal closure within a glottal cycle. Mechanical interaction mainly occurs in case of an obstruction of airflow in the vocal tract and thus occurs during consonantal full closures of the vocal tract. This decreases the glottal flow and leads to a reduction up to a stop of vocal fold vibration during the time interval of vocal tract closure.

In this study the importance of the acoustic-aerodynamic source-filter interaction for the generation of a singer's formant will be demonstrated using an articulatory-acoustic simulation model (8) capable of simulating glottal source and vocal tube aerodynamics, acoustics, and all of its interactions as described above.

2. METHOD

2.1 Model

The articulatory-acoustic simulation model used here (8, 9) comprises a subglottal, glottal and supraglottal component (Fig. 1, left). The subglottal system models the lungs and the tracheal tube. Lung or pulmonary pressure is inserted at the bottom of the tracheal tube and the subglottal pressure results underneath the glottis. The vocal fold system includes a self-oscillating two-mass model modeling the vibrating part of the glottis (Fig.1, right) and an extra shunt arranged in parallel modeling the cartilagenous part of the glottal slit (not shown in Fig. 1). The supraglottal part comprises an acoustic model of the pharyngeal, oral and nasal tube including all occurring mechanical, aerodynamic and acoustic loss mechanisms, including a noise frication generator (e.g. for producing fricatives) and including a radiation model for sound radiation at the mouth and at the nostrils. Vocal tract shapes can be formed using an articulatory model as described in Kröger et al. (8).

Three different versions of the model exist: i) direct insertion of a prescribed time-varying glottal flow in the vocal tract tube (see Veldhuis (10), no source-filter interaction, this variant of the model is called model 1). ii) Direct insertion of a prescribed time-varying glottal area A_1 (see Fig. 1; this variant of the model is called model 2). Here, glottal flow results from the aerodynamic-acoustic laws. Source-filter interaction (11, 7) occurs as a byproduct in this model because glottal flow is directly part of the subglottal-glottal-supraglottal vocal tube (see Fig. 1). iii) Self-oscillating glottis model (model 3, physiological model, see Pelorson (12)): the aerodynamic-acoustic part is same as model 2 but in addition the vibration of the vocal folds and thus glottal area as function of time is calculated from aerodynamic and mechanical forces acting on the two masses (13, 12, 11, 7).

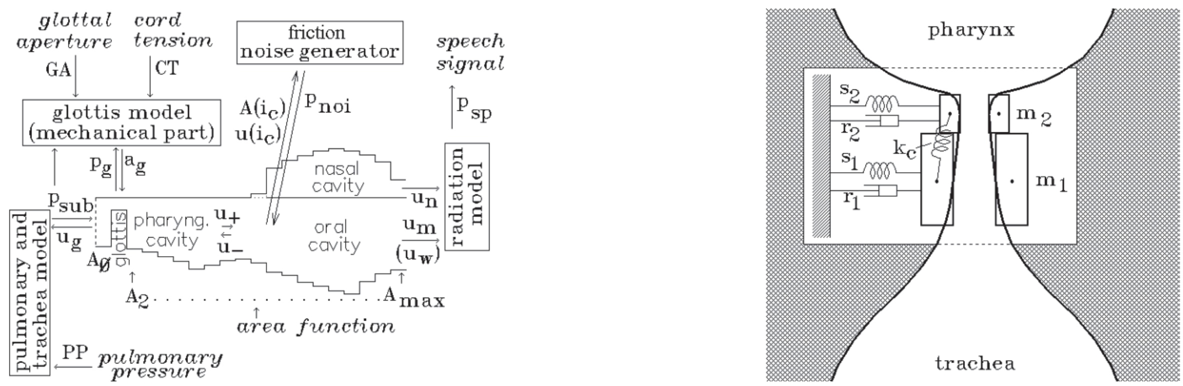

Figure 1 –The articulatory-acoustic simulation model (left side) and the mechanical part of the self-oscillating glottis model (right side; part of model 3). Left side: u_+ and u_- represent forward and backward traveling partial flow waves, u_g represents glottal flow, u_m and u_n represent air flow at mouth and nostrils, p_{sub} represents subglottal pressure and p_{sp} represents radiated sound pressure from the mouth and nostrils. GA, CT and PP represent glottal control parameters glottal aperture, cord tension and pulmonary (or lung) pressure. A_0, A_1 , A_2, … A_{max} represent cross-sectional area of the subglottal tube section, of the glottal tube section, as well as of the supraglottal tube sections. Noise is inserted if a narrow constriction together with a high airflow occurs in the oral tube (noise frication generator). Right side: A two-mass model represents the mechanical part of the self-oscillating glottis model and comprises two damped spring-mass pairs (m_i, r_i, s_i, i=1,2) as well as a coupling spring with spring constant k_c. Model 2: A_1 is directly controlled by an input time function. Model 1: Glottal area is set to a low value in comparison to A_0 and A2 in order to allow a small backward travelling wave u_{-1} as it occurs in the open phase of glottal cycle. Glottal flow is inserted in tube section 1 (glottal tube section) as forward traveling partial wave u_{+1}.

2.2 Simulations

In this study we simulated combinations of (i) two different lung pressure levels, (ii) two different types of source-filter interactions (no interaction vs. acoustic-aerodynamic interaction), (iii) different vocal tract shapes (schwa-sound an /a/-sound) and (iv) occurrence or absence of the narrow laryngeal tube, i.e. wide (or no) and narrow laryngeal tube. Two different values of lung pressure were set, representing a normal loud voice, here called speaking voice (pulmonary (or lung) pressure PP = 600 Pa) and a loud voice, here called singing voice (PP = 1200 Pa). Two different types of interactions were modeled in order to differentiate the cases no interaction vs. full acoustic-aerodynamic source-filter interaction. These different cases or models are labeled here as “model 1” (directly imposed glottal flow, no source-filter interaction) and as “model 2” (directly imposed glottal area, source-filter interaction). Two types of vocal tract shapes were used, i.e. that of a schwa-sound (constant cross sectional shape of vocal tract tube, Fig. 2) and that of a “technical /a/-sound” (see Fig. 2). The technical /a/ comprises two sections or tubes with constant cross section, i.e. a relatively narrow and a wide tube. The claim for producing an acoustically correct “cardinal /a/” is reached, if both tubes are of comparable length (14, 15). In addition each cross sectional tube was realized with a wide (i.e., no) and with a narrow laryngeal tube. These 8 cases used in the simulation experiment are subsumed in table 1. The self-oscillating glottis model is not used in this simulation study.

Table 1 – Simulation cases for each model (model 1 without source-filter interaction and model 2 including source-filter interaction)

case	lung pressure	vocal tract shape	laryngeal tube
1	low (speech)	schwa	wide
2	low (speech)	schwa	narrow
3	low (speech)	/a/	wide
4	low (speech)	/a/	narrow
5	high (singing)	schwa	wide
6	high (singing)	schwa	narrow
7	high (singing)	/a/	wide
8	high (singing)	/a/	narrow

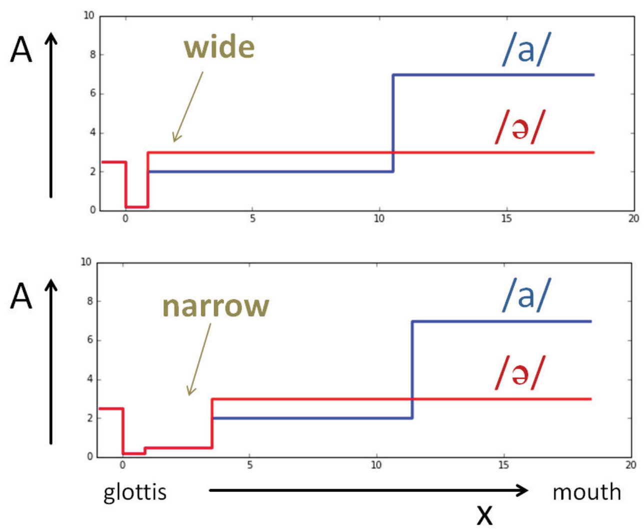

Figure 2 – The area function for schwa vowel (vocal tract tube with constant cross-sectional shape of 3 cm^2, red) and for the technical /a/-vowel (blue) with wide (top) and with narrow laryngeal extra-tube (bottom). The nasal tract is not coupled. All vowel productions are non-nasalized.

Figure 3 shows typical results for simulations using no source-filter interaction (model 1) and full acoustic-aerodynamic source-filter interaction (model 2). It can be seen that source-filter interaction leads to a modification of glottal flow, which can be found in a pronounced way in the first time derivative of glottal flow. This acoustic-aerodynamic interaction causes a formant ripple, i.e., an additional ripple in the time signal caused by the pressure and flow variations of the vocal tract which result from the vocal tract resonances (4, 5, 6). We can clearly see that the ripple occurring during the opening phase of the glottal cycle is in a frequency region of about 1000 Hz to 1500 Hz which covers the region of the second formant in case of the /a/.

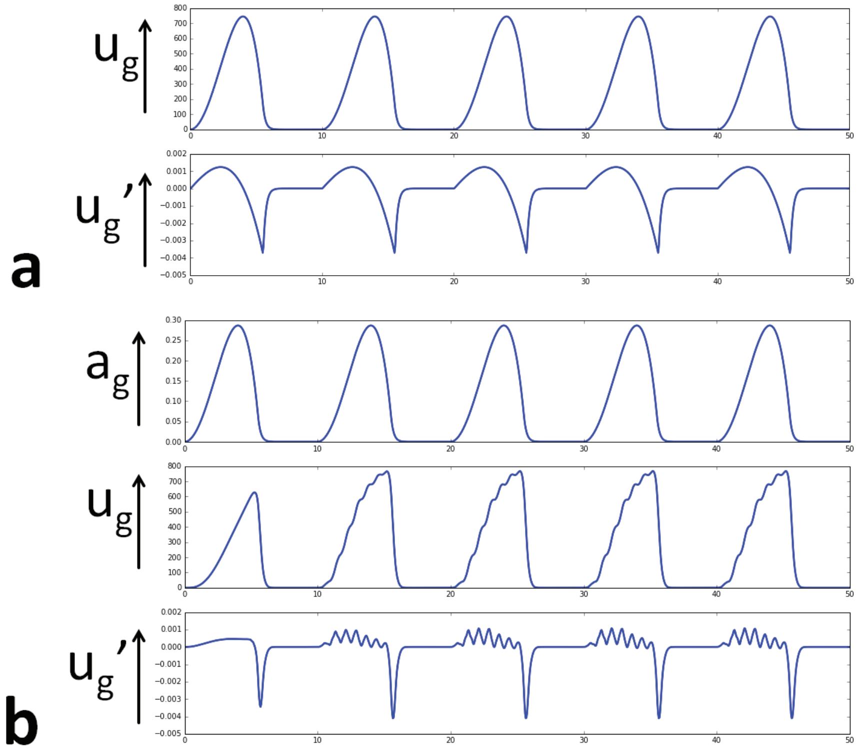

Figure 3 – Time signals for glottal flow u_g and its time derivative u_g' in a model without (model 1, top: a) and with source-filter interaction (model 2, bottom: b) for vowel /a/ and pulmonary pressure PP = 1200 Pa. In case of model 2 the glottal area function is plotted in addition to flow and time derivative of flow. The time axis is 50 msec and thus 5 glottal cycles are given (F0 = 100 Hz).

3. RESULTS

The analysis of the 8 spectra for the 8 simulations displayed in Fig. 4 reveals that in the case of no source-filter interaction (model 1) a fourth formant occurs in the case of narrow larynx-pharynx tube which replaces the fourth and fifth formant in the case of wide larynx-pharynx tube and the mean amplitude between 2 kHz and 5 kHz is a little increased by about 5 dB. This holds for speech and singing in the case of the simulation of the schwa vowel while no such increase in is observed in the case of simulation of the vowel /a/.

Analyzing the 8 spectra for the 8 simulations displayed in Fig. 5 we can see that in this case of acoustic-aerodynamic source-filter interaction (model 2) a strong increase in amplitude occurs within the frequency region of 2 kHz to 5 kHz in both cases, i.e. in case of schwa-vowel as well as in the case of vowel /a/. Thus, the occurring narrow laryngeal tube together with an occurring source-filter interaction seems to increase the mean amplitude in this frequency range by about 15 dB to 20 dB for speech and singing. In the case of the schwa vowel two formants replace the fourth formant in case of existence of the narrow laryngeal cavity while in case of /a/ vowel not such a clear situation can be observed. Here no extra formant occurs but a strong increase in amplitude in the frequency region between 2 kHz and 5 k Hz occurs as well.

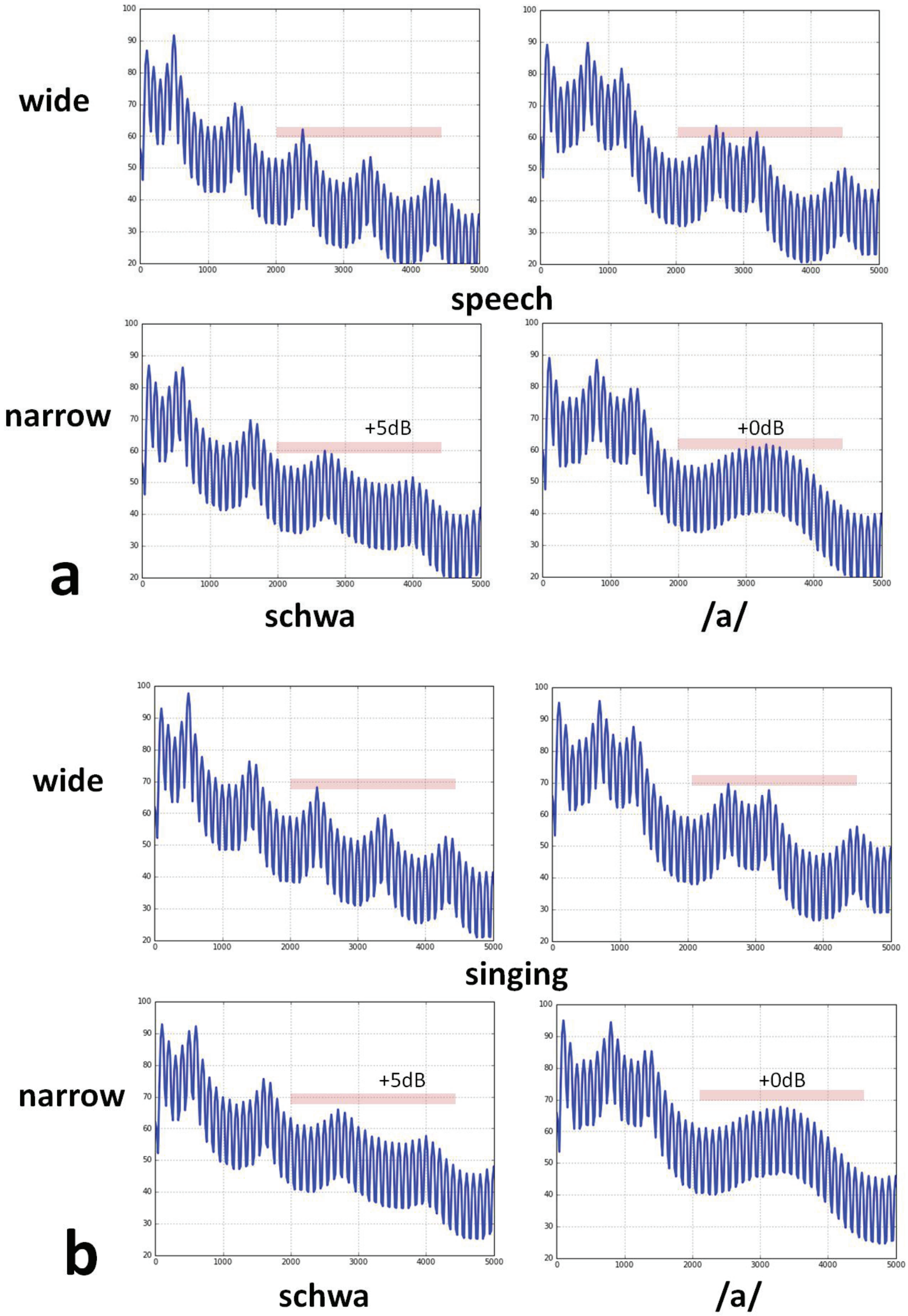

Figure 4: Spectra for schwa- and /a/-vowel and both laryngeal tube sizes using the model without acoustic-aerodynamic interaction (model1) for speech (a: PP = 600 Pa) and singing (b: Pa = 1200 Pa) (cases 1 to 8 in table 1). Frequency scale: 0 to 5 kHz; relative amplitude scale covers a range of 80 dB.

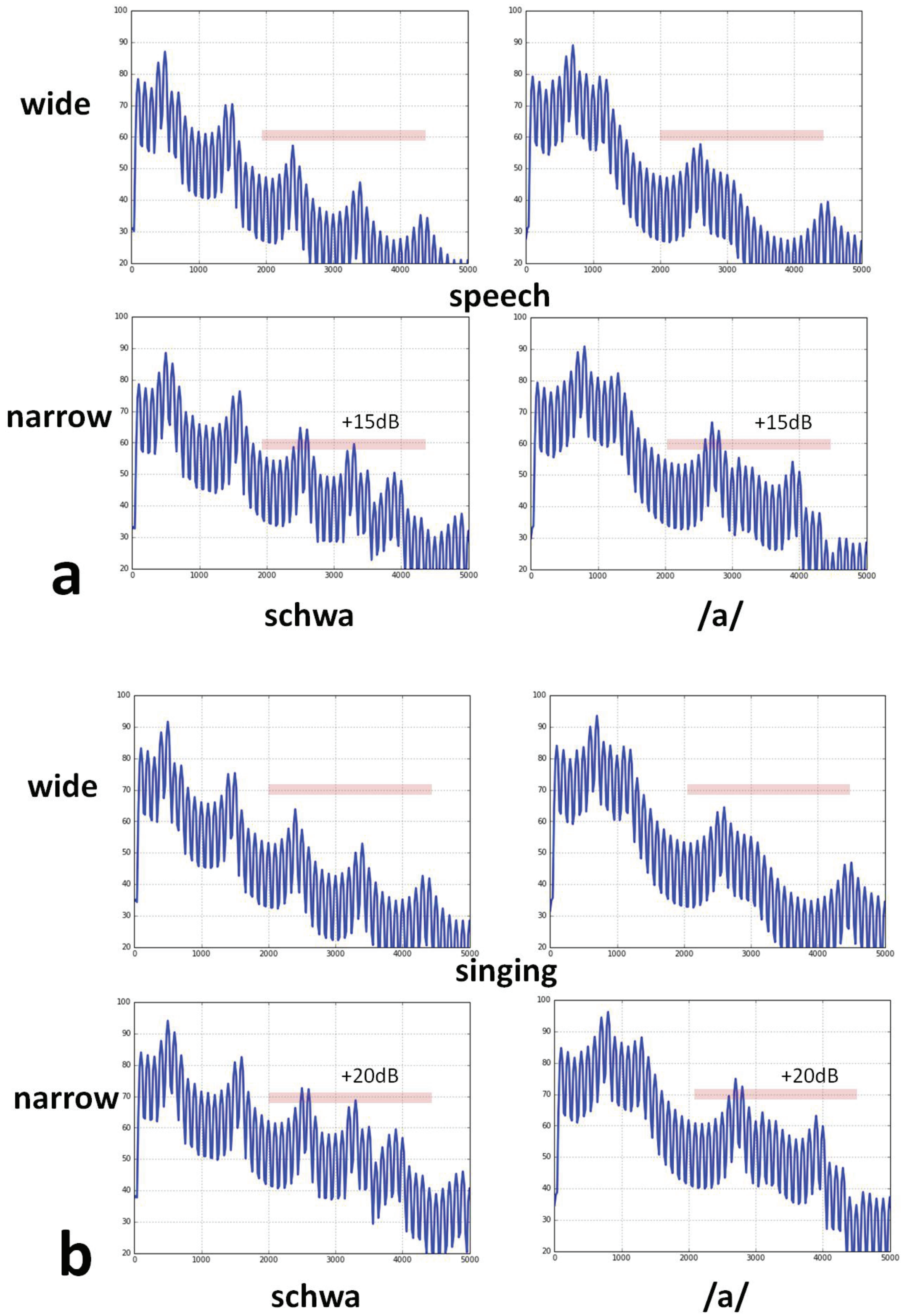

Figure 5: Spectra for schwa- and /a/-vowel and both laryngeal tube sizes using the model with acoustic-aerodynamic interaction (model 2) for speech (a: PP = 600 Pa) and singing (b: p_{sub} = 1200 Pa) (cases 1 to 8 in table 1). Frequency scale: 0 to 5 kHz; relative amplitude scale covers a range of 80 dB.

4. DISCUSSION

No clear effect can be found for the modification of formants by changing from a wide to a narrow larynx-pharynx tube in case of low subglottal pressure and in the case of a simple articulatory-acoustic model which does not include acoustic-aerodynamic source-filter interaction. Thus, the change of formant intensity in the region of 2 kHz to 5 kHz mainly depends on the model used, i.e. whether the model with or without source-filter interaction is used. The intensity in this region increases by 15 dB to 20 dB in case of source-filter interaction (model 2) while it increases in maximum by 5 dB in case of no source-filter interaction (model 1). Furthermore the change of pulmonary pressure from 600 Pa (normal pulmonary pressure as occurs during speaking) to 1200 Pa (high pulmonary pressure as may occur e.g. in opera singing) has no strong effect on increase in intensity within the singer's formant frequency region in case of no source-filter interaction. But in the case of source-filter interaction this leads to a small effect, i.e., to an increase of 5 dB in that frequency range.

Thus this simulation study illustrates that source-filter interaction has a considerable effect on establishing a singer's formant in combination with the occurrence of a narrow laryngeal tube. But it is not clear whether this effect is caused by the clearly visible effect of formant ripple in the opening phase of the glottal cycle (Fig. 3) or by other effects resulting from acoustic-aerodynamic source-filter interaction. Thus it should be checked whether the source-filter interaction effect influences the closing phase parameters of glottal flow, i.e. the parameters Ee, Td and Ta in terms of the LF-model (16, 17). More simulations and analyses of simulated as well as of natural data are needed in order to confirm or reject this hypothesis.

5. CONCLUSIONS AND FURTHER WORK

It can be concluded that the occurrence of source-filter interaction beside insertion of a narrow laryngeal extra tube initiates the effect on increase in intensity of about 15 dB in the frequency region of the singer's formant. But no clear or rule-based effect can be observed for a shift of formants or an insertion of an extra formant in the case of an insertion of the laryngeal narrow tube on the basis of our simulation experiments. Like in other studies here a clear separation of cross sectional area of the laryngeal tube (narrow tube) and the further pharyngeal tube (wide tube, wide sinus Morgagni, see Sundberg (1, 2) and Sundberg et al. (3)) needs to occur within the area function of the vocal tract and thus the articulatory maneuvers which initiate such an occurrence of a narrow laryngeal extra tube need to be discovered and learned by singers. This may be the reason why only trained opera singers are capable of producing the singer's formant.

The effects speculated on in this paper need a further and more detailed investigation. It is necessary to check the increase in intensity in the frequency range of the singer's formant by doing simulations with more vocal tract shapes (more vocalic configurations beside schwa- and /a/-sound) as well as to check the effects ad different F0. The simulations introduced here are done only on schwa vowel and technical /a/ vowel at fundamental frequency F0 = 100 Hz. Thus a lot of combinations of F0 values with different vocal tract shapes should be checked in further studies in order to evaluate the effect of source-filter interaction and of insertion of a narrow laryngeal tube section in more detail.

ACKNOWLEDGEMENTS

This paper is dedicated to all classic style singers and especially to Rauf Berman (Maastricht, Netherlands).

REFERENCES

1. Sundberg J (1974) Articulatory interpretation of the "singing formant". Journal of the Acoustical Society of America 55:838-844
2. Sundberg J (2001) Level and center frequency of the singer's formant. Journal of Voice 15: 176-186
3. Sundberg J, Lã FMB, Gill BP (2011) Professional male singers' formant tuning strategies for the vowel /a/. Logopedics Phoniatrics Vocology 36: 156-167
4. Bavegard M, Fant, G (1994) Notes on glottal source interaction ripple. Department for Speech, Music and Hearing at KTH Stockholm: Quarterly Progress and Status Report STL-QPSR 35(4): 63-78
5. Fant G, Lin Q (1987) Glottal source -vocal tract acoustic interaction. Department for Speech, Music and Hearing at KTH Stockholm: Quarterly Progress and Status Report STL-QPSR 28(1): 13-27
6. Fant G (1986) Glottal flow: models and interaction. Journal of Phonetics 14: 393- 399

7. Titze I (2008) Nonlinear source–filter coupling in phonation: Theory. Journal of the Acoustical Society of America 123: 2733-2749
8. Kröger BJ, Bekolay T, Eliasmith C (2014) Modeling speech production using the Neural Engineering Framework. Proceedings of CogInfoCom 2014 (Vetri sul Mare, Italy) pp. 203-208 (ISBN: 978-1-4799-7279-1) and IEEE Xplore Digital Library DOI=10.1109/CogInfoCom.2014.7020446
9. Liljencrants J (1985) Speech Synthesis with a Reflection-type Line Analog. Department of Speech Communication and Music Acoustoustics, Royal Institute of Technology, Stockholm. Department for Speech, Music and Hearing at KTH Stockholm: Quarterly Progress and Status Report STL-QPSR 36(2-3): 119-156
10. Veldhuis R (1998) A computationally efficient alternative for the Liljencrants–Fant model and its perceptual evaluation. Journal of the Acoustical Society of America 103: 566-571
11. Lous NJC, Hofmans GCJ, Veldhuis RNJ, Hirschberg A (1998) A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design. Acta Acustica united with Acustica 84: 1135-1150
12. Pelorson X, Hirschberg A, van Hassel RR, Wijnands APJ (1994) Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. Journal of the Acoustical Society of America 96: 3416-3431
13. Ishizaka K, Flanagan JL (1972) Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Labs Technical Journal 51: 1233-1268
14. Stevens KN (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In: David EE & Denes PB (eds.) Human communication: A unified view (McGraw Hill, New York), pp. 51-66
15. Stevens KN, Keyser SJ (2010) Quantal theory, enhancement and overlap. Journal of Phonetics 38: 10-19
16. Fant G, Liljencrants J, Lin Q (1985) A four-parameter model of glottal flow. Department for Speech, Music and Hearing at KTH Stockholm: Quarterly Progress and Status Report STL-QPSR 26(4): 1-13
17. Fant G (1995) The LF-model revisited. Transformations and frequency domain analysis. Department for Speech, Music and Hearing at KTH Stockholm: Quarterly Progress and Status Report STL-QPSR 36(2-3): 119-156