



HAL
open science

Practical Guidelines for Implementing Vocal Tract Resonances Characterization with Excitation at the Lips

Timothée Maison, Baptiste Allain, Patrick Hoyer, Fabrice Silva, Philippe Guillemain, Nathalie Henrich Bernardoni

► **To cite this version:**

Timothée Maison, Baptiste Allain, Patrick Hoyer, Fabrice Silva, Philippe Guillemain, et al.. Practical Guidelines for Implementing Vocal Tract Resonances Characterization with Excitation at the Lips. MAVEBA 2021 - 12th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications, Dec 2021, Florence, Italy. hal-03485436

HAL Id: hal-03485436

<https://hal.science/hal-03485436>

Submitted on 17 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PRACTICAL GUIDELINES FOR IMPLEMENTING VOCAL TRACT RESONANCES CHARACTERIZATION WITH EXCITATION AT THE LIPS

T. Maison¹, B. Allain¹, P. Hoyer², F. Silva¹, P. Guillemain¹, N. Henrich Bernardoni³

¹ Aix-Marseille Univ., CNRS, Centrale Marseille, LMA, Marseille, France

² Fraunhofer Headquarters, Hansastr, 27c, 80686 Munich, Germany

³ Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

timothee.maison@ens-lyon.org, baptiste.allain@centrale-marseille.fr, patrick.hoyer@zv.fraunhofer.de, silva@lma.cnrs-mrs.fr, guillemain@lma.cnrs-mrs.fr, nathalie.henrich@gipsa-lab.fr

Abstract: The measurement of vocal tract resonances is crucial to understand voice acoustics. Their characterization using a broadband excitation signal and pressure measurements both at the lips is a good compromise between accuracy, speed and intrusiveness. In this paper we address some practical guidelines for performing such measurements in order to provide reliable estimates for resonance frequencies and quality factors. We also experiment the possibility to move away microphone from excitation at the lips, with a cylindrical waveguide as ideal vocal tract and its line-transmission model. Using a far-field model in which pressure decreases as the inverse of the distance from excitation at the lips, the microphone could be placed anywhere as soon as the Signal-to-Noise Ratio is good enough: measurements become more sensitive to interferences with other acoustics sources. Modal analysis is performed with a robust method using both amplitude and phase of measured functions. Resonance frequencies and quality factors estimations at distances up to 30 cm deviates by less than 0.2% and 10% respectively from reference measurement at the lips validating accurate characterization with a distant microphone from the lips.

Keywords: vocal tract, resonance, measurement, radiation, acoustics

I. INTRODUCTION

Vocal tract resonances are essentials for spoken communication and they enhance singing voice efficiency [1]. Their characterization has led to the development of several devices with a compromise between intrusiveness and accuracy. Indirect methods based on formants analyses in the voice signal (such as linear predictive coding) evidenced severe limitations for high pitches or closed vowels. We focus here on a non-invasive device based on a broadband excitation signal at the lips, first proposed by Epps [2]. It has

undergone several developments and variations such as the use of sine-sweep signals [3], a buzzer excitation device [4], or the direct estimation of pressure and velocity [5]. Recently the measurement model of this approach has been clarified, also showing its limitations [6]. While the measurement setup seems to be quite simple, accurate estimate of resonance frequencies and quality factors requires some precautions. From the hardware system implementation to the signal processing of a measurement and the estimation of modal parameters, each step needs precise adjustments. We first address a full methodology and some practical guidelines for performing such measurements in the best possible conditions (Sec. II), leaving aside for the moment the question of real-time estimation. Motivated by the high voice level at the lips that can saturate and damage the microphone (physical clipping), we explore the possibility to separate pressure measurement from excitation (traditionally both done at the lips), by means of experiments and radiation theory (Sec. III). We study here the unvoiced case. All results are expected to remain valid in the voiced case assuming a perfect separation between the excitation signal and the voice.

II. METHODS

A. Measurement device and theoretical context

The device consists of a flexible capillary (diameter 4 mm) connected to an impedance adapter on a loudspeaker (Beyma CP850Nd) with amplifier (Flying Mole DAD-M100 proII BB) and sound card (Focusrite 6i6), positioned at the inlet of a vocal tract model (an ideal open-closed cylindrical waveguide with length 15 cm and diameter 21 mm for validation purpose). The flexible capillary is coated by a thicker tube in order to remove its sides radiation (see Figure 1): the only acoustic source should be the output of the capillary at the lips – otherwise interferences may

occur. The loudspeaker and impedance adapter with the tube must also be acoustically isolated for the same reason.



Figure 1 – Photo of excitation at the waveguide inlet.

As usually done [1-3, 5, 6], the microphone is positioned close to the excitation system at the lips. It records the waveguide (vocal tract) responses to a broadband signal in a calibration condition (closed-mouth) and in a measurement condition (open-mouth). The pressure measurement model at the lips, valid for all kinds of excitation signal, has been described in [6]. The spectrum in open-mouth condition $P_{meas}(\omega)$ calibrated by closed-mouth spectrum $P_{cal}(\omega)$ gives access to a frequency response $H(\omega)$ characterizing the vocal-tract acoustics including its radiation (Equation 1). The excitation capillary is small enough (high impedance) to be assumed independent from the load (i.e. it provides the same source flow for open or closed vocal tract).

$$H(\omega) = \frac{P_{meas}}{P_{cal}} = \frac{Z_{VT}}{Z_{VT} + Z_R} \quad (1)$$

with Z_{VT} the vocal tract input impedance seen from the lips and Z_R the vocal tract radiation impedance. The resonances of $H(\omega)$ are identical to a usual vocal tract transfer function pressure at the lips over flow at the glottis ($P_{lips} / U_{glottis}$). The lips radiation is included, so these resonances corresponds to the acoustics of the radiating vocal tract.

B. Excitation signal and analysis method

A synchronized exponential swept-sine is used (frequency range 0,1-5,5 kHz, duration 1 s, with sinusoidal fade in and fade out). Deconvolution in Fourier domain method detailed in [7] is applied, which enables to properly separate non-linear contributions of the excitation (and measurement) chain from the waveguide linear impulse response. Briefly, each pressure measurement (closed and open mouth conditions) is convoluted in Fourier domain with inverse sweep, then the linear impulse is windowed in time domain. For the cylindrical waveguide in use here, the linear impulse is around 120 ms. A long-enough flat-top window is used to minimize distortion of the impulse response. Finally

the corrected frequency response $H(\omega)$ is calculated (Eq. (2)) with a regularized (parameter $\epsilon(\omega)$) spectrum inversion as discussed in [8].

$$H(\omega) = \frac{P_{meas} \cdot P_{cal}^*}{|P_{cal}|^2 + \epsilon(\omega)} \quad (2)$$

Time-domain signals are sampled at $f_s = 44100$ Hz, Fourier transforms are performed with the power of 2 greater than the length of signal to optimize computation with zero-padding. The final frequency resolution of $H(\omega)$ is less than 1 Hz. The excitation signal sound pressure level measured with a dBmeter (Nor131) reaches around 86 dB (LAF_{max}) at 9 cm from the source (capillary output and waveguide inlet). All measurements were made in the anechoic room of LMA laboratory.

C. Microphone position testbed

In order to explore the possibility of placing the microphone away from the lips, a microphone (GRAS 40PR/PL with conditioner MMF M108) is horizontally moved away between 0.4 cm and 38 cm from the waveguide inlet. For each microphone position, two records are made in 1) open-mouth condition and 2) closed-mouth condition. Root-Mean Square value of pressure level is computed over short time windows (ten periods long) along the sweep, with respect to the distance. It is compared with a monopolar far-field radiation model (RMS decreasing as the inverse of the distance).

In a second step, a 4-microphones ($M1, M2, M3, M4$) antenna (at distance $d = 0.5, 10, 20$ and 30 cm) allows simultaneous measurements (see Figure 2) of the frequency response at the waveguide inlet. Resonance frequencies and quality factors estimated at each microphone distance from the vocal tract inlet (0.5 cm reference) were compared with each other, and with a transmission line numerical model.

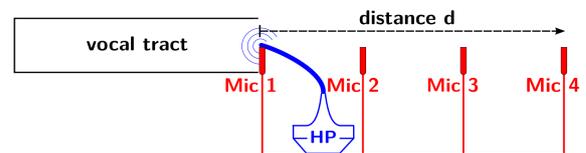


Figure 2 – Four microphones ($M1, M2, M3, M4$) antenna scheme at distances 0.5, 10, 20 and 30 cm.

D. Modal estimation

We focus on the resonances (amplitude maxima) of the calculated frequency response $H(\omega)$. The chosen method to determine a resonance frequency and quality factor should be robust to noise and to the proximity of nearby zeros (see Figure 4). Even if frequency is not

biased, quality factor estimation may suffer from the skewness of the amplitude peak and from the phase shift: bandwidth at -3dB of $H(\omega)$ resonances deviates from real bandwidth of vocal tract resonances. First the frequency areas of maxima are delimited (see Figure 3 for the first resonance case). The values of $H(\omega)$ in the considered frequency range are displayed on the complex plane and fitted on a geometric circle (so-called Kennelly circle). Then fitting the curve of angle relative to the circle's center enables the evaluation of the resonance frequency and of its quality factor [9].

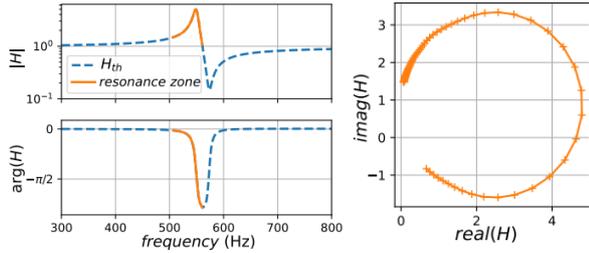


Figure 3 – Definition of first resonance zone and associated Kennelly circle for numerical model.

III. RESULTS

A. Far-field pressure decrease

Figure 4 displays the RMS pressure value (over ten periods) of recorded signals for the mid-frequency 2400 Hz with respect to the distance from the waveguide inlet for open and closed mouth conditions. Plots are normalized by the reference value at the inlet (0.4 cm). They are compared to the far-field model.

The sound pressure level decreases globally as the inverse of the distance, equivalent to a far-field monopolar radiation model on the considered frequency ranges, as soon as distance is superior to 2 cm from the inlet. For other studied frequencies, little variation is observed: the decrease in pressure remains proportional to the inverse of distance.

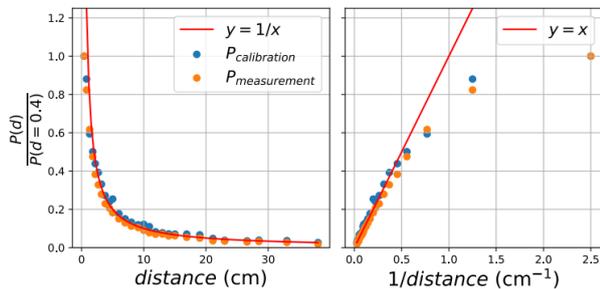


Figure 4 – Normalized RMS pressure levels at 2.4 kHz with distance (left) and with inverse of distance (right).

The decrease behavior is similar between closed and open conditions which indicates that the method can compensate the position of the microphone.

Therefore the measurement model (Eq. (1)) remains the same for remote microphones.

B. Far-field modal estimation

Figure 5 displays measurements made with the 4-microphones antenna with distances 0.5 (M1) , 10 (M2) , 20 (M3) and 30 cm (M4) . Modal estimations (frequencies in Table 1, quality factors in Table 2) obtained from measurements are compared with the one at 0.5 cm and with the frequency response at the waveguide inlet from transmission line model with visco-thermal losses. Model data are computed with the same modal analysis as Sec. II.D.

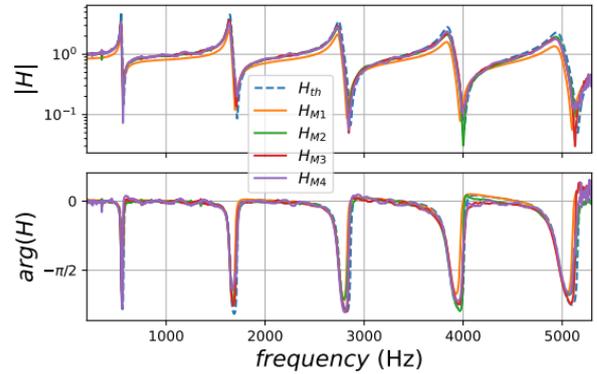


Figure 5 – Example of H measurements with distance for 4 microphones and numerical model (dashed-line)

Measurements at different distances are very similar and close to the numerical model. The distance of the microphone and the decrease in pressure is well compensated between calibration and measurement. Noise appear for far-field positions of microphone due to a lower Signal-to-Noise Ratio. The measurement at the lips ($M1$) has a lower amplitude probably due to the bias introduced by the thick cover used for closed mouth condition (1 mm thick rigid latex cut to the size of the cylindrical waveguide diameter).

f (Hz)	$R1$	$R2$	$R3$	$R4$	$R5$
Model	546.8	1645.4	2746.5	3851.2	4960.7
M1	543.8	1641.3	2736.9	3846.5	4947.1
(std)	± 0.1	± 0.4	± 0.6	± 1.0	± 1.6
M2	543.9	1641.7	2737.4	3845.6	4951.1
(std)	± 0.1	± 0.4	± 0.8	± 1.6	± 2.0
M3	544.0	1642.1	2737.2	3851.7	4947.5
(std)	± 0.1	± 0.5	± 1.1	± 3.0	± 3.3
M4	543.5	1641.2	2735.1	3847.8	4950.4
(std)	± 0.2	± 0.7	± 2.0	± 3.0	± 3.4

Table 1 – Mean estimations of resonance frequencies (and standard deviation) for 20 measurements (with new calibration for each one) and numerical model.

For all studied distances, resonance frequencies estimates deviate by less than 0.2 % relatively to the reference measurement at 0.5 cm (*M1*). They are underestimated by less than 0.6 % relatively to the computational model at the inlet. Standard deviation of frequency estimates slightly increases with the distance (microphone number) and with the frequency (resonance number).

<i>Q</i>	<i>R1</i>	<i>R2</i>	<i>R3</i>	<i>R4</i>	<i>R5</i>
Model	59.1	55.6	45.4	39.0	35.0
M1 (std)	46.6 ± 0.2	45.5 ± 1.2	39.9 ± 0.4	32.5 ± 1.2	30.8 ± 0.4
M2 (std)	46.9 ± 0.6	46.2 ± 1.2	39.9 ± 1.2	31.8 ± 1.8	32.0 ± 1.3
M3 (std)	47.0 ± 0.5	46.8 ± 1.1	39.2 ± 2.1	32.4 ± 3.0	31.9 ± 3.4
M4 (std)	47.1 ± 1.2	48.7 ± 0.8	38.7 ± 2.4	36.1 ± 3.5	31.9 ± 4.5

Table 2 – Mean estimations of resonance quality factors (and standard deviation) for 20 measurements (with new calibration for each one) and model

For all microphones, quality factors estimates deviate by less than 10 % relatively to the reference measurement (*M1*). They are underestimated compared with results from the model, particularly for the first resonance: measurement of a low-loss waveguide would require a longer excitation time to obtain a better estimate of amplitude and quality factor. This should not be an issue for the estimation of quality factors of real vocal tract whose low-frequency resonances have larger bandwidth [10]. Standard deviation increases with the frequency and the distance.

IV. CONCLUSION AND GUIDELINES

Results evidence that the microphone can be positioned away from the lips. If the study focuses on a horizontal motion of the microphone in order to limit the different radiation patterns of frequencies, the calibration and measurement are always performed at the same microphone position so that radiation pattern is compensated. Therefore the microphone could be placed anywhere. However, the transfer function is more sensitive to ambient noise as the distance of the microphone increases. For accurate measurement, any other acoustic sources than the excitation at the lips should be avoided or limited. Interferences between different sources would appear as noise or peaks on measurements.

A microphone placed at 30 cm distance as recommended by the Union of the European Phoniatricians [11] for stand mounted microphone voice measurements can be used without loss of precision. The device could also be based on head-mounted microphone on which an excitation system could be added. Those possibilities are advantageous for comfort of subjects and to encourage an ecological vocal gesture during studies.

REFERENCES

- [1] N. Henrich Bernardoni, J. Smith, and J. Wolfe, “Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones”, *J. Acous. Soc. Am.*, vol. 129, no 2, pp. 1024-1035, 2011.
- [2] J. Epps, J.R. Smith, and J. Wolfe, “A novel instrument to measure acoustic resonances of the vocal tract during phonation”, *Meas. Sci. and Tech.*, vol. 8, no 10, pp. 1112, 1997.
- [3] B. Delvaux, and D. Howard, “Sinesweep-Based Method to Measure the Vocal Tract Resonances”, *Proc. 9th MAVEBA*, September 2-4, 2015.
- [4] P. Hoyer, and S. Graf, “Adjustment of the vocal tract shape via biofeedback: a case study”, *J. Voice*, vol. 33, no 4, pp. 482-489, 2019.
- [5] M. Kob, and C. Neuschaefer-Rube, “A method for measurement of the vocal tract impedance at the mouth”, *Med. Eng. & Physics*, vol. 24, n 7-8, pp. 467-471, 2002.
- [6] T. Maison, F. Silva, Ch. Vergez, and N. Henrich Bernardoni, “Measuring vocal-tract impedance at the lips: model, hypotheses and limits”, *Proc. 12th ICVPB*, Grenoble, France, December 2020.
- [7] A. Novak, P. Lotton, and S. Laurent, “Synchronized swept-sine: Theory, application and implementation”, *J. Audio Engineering Society*, vol.63, no 10, pp. 786-798, 2015.
- [8] M. Rébillat, R. Hennequin, E. Corteel, and B. Katz, “Identification of cascade of Hammerstein models for the description of non-linearities in vibrating devices”, *J. Sound and Vibration*, vol.330, no 5, pp. 1018-1038, 2010.
- [9] A.E. Kennelly, and K. Kurokawa, “Acoustic impedance and its measurement”, *Proc. American Academy of Arts & Sciences*, vol.56, no 1, pp. 3-42, 1921.
- [10] N. Hanna, J.R. Smith, and J. Wolfe, “Low frequency response of the vocal tract: acoustic and mechanical resonances and their losses”, *Proc. Australian Acoustical Society*, Fremantle, Australia, pp. 317-323, 2012.
- [11] J.G. Svec, and S. Granqvist, “Guidelines for selecting microphones for human voice production research”, *Am. J. Speech-Language Pathology*, vol.19, no 8, pp. 356-369, 2010.