

Study of vocal fold vibration using M-mode ultrasound: a proof of concept

Juliette Dindart¹, Agnès Rouxel^{1,2}, Crystal Lin³, Trung Kien Bui¹, Muriel Lefort⁴, Claire Pillot-Loiseau⁵, Christophe Trésallet², Frédérique Frouin¹

¹LITO, Inserm, Institut Curie, Université Paris-Saclay, France

²Departments of Nuclear Medicine and Surgery, Avicenne Hospital, AP-HP, Université Sorbonne Paris Nord, France

³Mindray France SARL, France

⁴LIB, CNRS, Inserm, Sorbonne Université, France

⁵LPP UMR 7018, CNRS, Sorbonne Nouvelle, France

Abstract

Ultrasound has recently been suggested as an alternative to laryngoscopy for checking vocal fold movement after neck surgery. We propose here to use M-mode ultrasound (MUS) to study vocal fold vibration in the left and right hemilarynges. MUS is acquired along a 1D line in each hemilarynx for about 5 seconds during phonation of a vowel. Post-processing has been developed to estimate spatio-temporal maps of fundamental and second harmonic frequencies. To validate our method, a total of 108 recordings (MUS and voice) from 12 subjects with no vocal pathology were acquired. We compared the fundamental frequency obtained by MUS with that provided by voice analysis. The median fundamental frequency was estimated by MUS with high accuracy ($y=0.997x+0.293$ with $r=0.999$). In our current trial, estimated frequencies are limited to 250 Hz. In future work, frequency will be increased to 1 kHz to avoid aliasing effect at high pitch, and MUS will be tested on patients suffering from vocal pathologies.

Index Terms: vocal fold vibration, M-mode ultrasound

1. Introduction

After thyroid surgery, due to the closeness of recurrent nerve to the thyroid, 3 to 5 % of patients may experience recurrent nerve damage [1]. This damage can cause vocal fold paralysis (VFP). According to current recommendations, flexible laryngoscopy should be performed before and after surgery to check on the mobility of the vocal folds [2]. However this procedure isn't always part of routine care due to patient discomfort and practicality. The lack of compliance with recommendations can lead to delayed or missed diagnosis of VFP. To overcome the disadvantages of laryngoscopy, some teams have recently proposed the detection of vocal fold palsy using dynamic trans-laryngeal ultrasound, based on the acquisition of dynamic B-mode (2D+t) images during free-breathing [3], [4], [5], [6], [7]. Even if the vocal folds are not always visible, their mobility can be assessed through the observation of arytenoid cartilages to which the vocal folds are connected. This ultrasound procedure has proved useful for detecting VFP and could be routinely proposed to patients, due to its low cost and non-invasive nature.

Conventional medical ultrasound systems can also acquire M-mode data which involves recording the ultrasound signal along a specific 1D line at a very high sampling rate. This mode is frequently used in echocardiography (see for example [8], [9]), but it has also proved useful to estimate the onset time of deep muscle activity [10]. Thus M-mode is relevant when a high temporal resolution is crucial.



FIGURE 1 – Experimental setup : simultaneous ultrasound acquisition and voice recording

For this study, considering dynamic trans-laryngeal ultrasound as a relevant clinical tool for assessing vocal fold mobility, we hypothesized that the use of M-mode ultrasound could complement this examination by studying vocal folds vibration. In this initial study, we aim to evaluate the vocal fundamental frequency from M-mode ultrasound. To validate this measurement, audio recordings were simultaneously acquired. Our first results demonstrate the interest of M-mode ultrasound as a new approach to evaluate vocal folds vibration.

2. Methods

2.1. Translaryngeal ultrasound

All the examinations were performed using a TE7 (Mindray) point of care ultrasound system with a L12-3RCs linear array transducer, having a frequency range between 3 and 12.8 MHz. The subjects were lying on their back with their neck slightly extended (see Figure 1). The M-mode acquisition (see below 2.2 for its definition) was combined with B-mode to make the positioning of the 1D line easier. Probe was positioned to reveal the right (then the left) hemilarynx, with vocal fold, false ventricular fold and arytenoid cartilage.

Then subjects were asked to hold a vowel (either /a/, /i/, or /u/) or the consonant /z/ for at least five seconds, and the practitioner moves the probe to follow the phonation-induced displacement of anatomical structures.

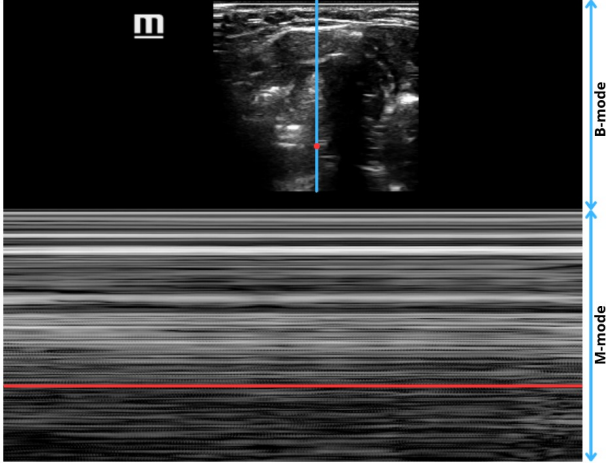


FIGURE 2 – Example of B-mode (top) and M-mode (bottom) images, with M-mode recorded along a vertical line (in blue on the B-mode image); the M-mode part represents ultrasound signal values along the vertical line according to time, in the horizontal direction. The signal evolution over time corresponding to the red point on the B-mode image is displayed along the red line in the M-mode zone.

2.2. M-mode recording

In ultrasound, B-mode images are 2D and videos are dynamic 2D images that are acquired at a frame rate of 30 images per second. Using M-mode, it is possible to largely increase the acquisition rate since the ultrasound signal is recorded along a single 1D line, over a length varying from 3.5 to 5 cm, depending on the subject's anatomy. B-mode ultrasound provides a 2D image with J rows and K columns, while M-mode is a specific mode that acquires the signal along a single column. For B-mode, the acquisition process is repeated over time for each of the columns. For M-mode, $K=1$, so time resolution can be considerably increased (by up to two orders of magnitude). In our device, the signal is acquired at a rate of 500 points per second ($F_r=500$ Hz). Figure 2 illustrates an extract of a B-mode/M-mode recording. The recordings lasted between 5 and 10 seconds during vowel phonation.

2.3. M-mode processing

Figure 3 illustrates the vibration process captured by the M-mode ultrasound. Let $s(j, t)$ be the M-mode ultrasound signal, measured at position j on the vertical line displayed on the B-mode image, at time t . Let $\mathbf{v}^{jt}(\tau_n)$ be the vector of N consecutive time values of $s(j, t)$, centered at time t and $V^{jt}(f_k)$ the Discrete Fourier Transform (DFT) of $\mathbf{v}^{jt}(\tau_n)$:

$$\mathbf{V}^{jt}(f_k) = \mathcal{F}(\mathbf{v}^{jt}(\tau_n)), \quad (1)$$

where \mathcal{F} is the DFT operator. The peak frequency f_0 at which $|\mathbf{V}^{jt}(f_k)|$ is maximum, over all frequencies, provides the estimation of the mean fundamental frequency. It is computed for each location j and time t , and is thus noted f_0^{jt} .

Considering a sliding time window of 30 ms, a parametric map K is computed as the collection of f_0^{jt} values for each pair (j, t) . To reduce the impact of low frequencies, which can generate noise in parametric maps, a frequency window $[f_{min}, f_{max}]$ was defined to constrain the peak frequency f_0^{jt} to belong to this interval. The same procedure was applied to

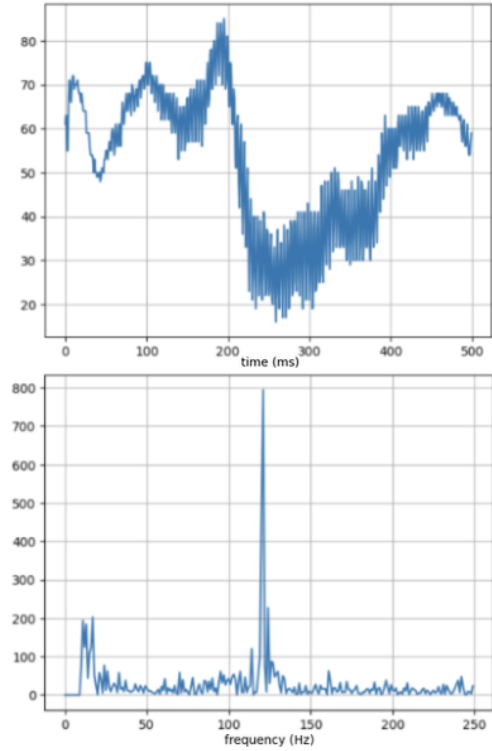


FIGURE 3 – Vibration process as captured by the M-mode signal, $s(j, t)$ (top), and its Discrete Fourier Transform (bottom). The peak frequency is equal to 120 Hz.

estimate the second harmonic map K_2 , when Shannon-Nyquist conditions were valid, i.e. when the fundamental frequency was below 125 Hz. Typical values of f_{min} were between 10 and 30 Hz for the calculation of K and between 130 and 160 Hz for the calculation of K_2 (Figure 4), f_{max} being equal to 250 Hz in all cases.

The number of points N defined for estimating the fundamental frequency f_0^{jt} was chosen to be 500. Finally a temporal window $[t_{min}, t_{max}]$ and a spatial window $[j_{min}, j_{max}]$ were defined to robustly compute the median frequency on each parametric map. This value is further noted f_0^{US} for K . Figure 5 is an illustration of two parametric maps, K and K_2 associated with one M-mode ultrasound record.

3. Experimental validation

3.1. Database

To test the feasibility of the M-mode ultrasound for assessing laryngeal vibrations, a database of 108 recordings from twelve subjects (8 males, and 4 females), was compiled (see Table 1). Ages ranged between 23 and 59 years. All subjects had no known vocal pathology and had not undergone cervical surgery.

3.2. Voice recordings and analysis

To validate the fundamental frequency estimation f_0 from ultrasound, simultaneously to ultrasound acquisition, the subject's voice was recorded using Octa-Capture UA 1010 soundcard (Roland) with C520 condenser microphone (AKG) at a sampling rate of 45 kHz. As seen on Figure 1, the subject was

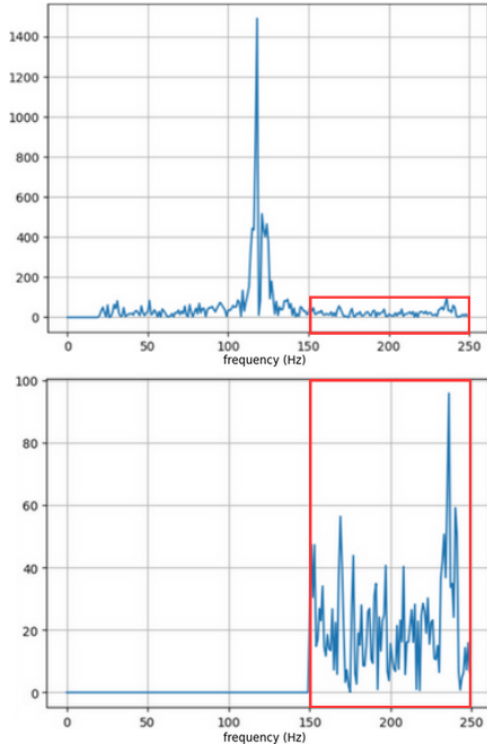


FIGURE 4 – Discrete Fourier Transform of the larynx vibration captured by the M-mode. The second harmonic is visible at 237 Hz with window filtering to keep only frequencies > 150 Hz.

TABLE 1 – Composition of the database according to the sounds held and the sex of the subjects

Sex	/a/	/i/	/u/	/z/	Total
Males (8)	26	22	25	4	77
Females (4)	13	8	8	2	31
Total (12)	39	30	33	6	108

lying down and recorded simultaneously with the microphone and M-mode ultrasound while maintaining a prolonged vowel vocalization.

Voice parameters were then extracted from the voice recordings using Praat software [11], which provided the time evolution of the fundamental frequency (f_0). Spectrograms were also extracted from the recordings with Parselmouth [12], the python extension of Praat software. As a pitch jump effect can be observed in the spectrogram and can induce biased values of the mean fundamental frequency [13], this value was replaced by the median fundamental frequency. It is further noted f_0^{VA} , the superscript VA referring to Voice Analysis.

4. Results

4.1. Spatio-temporal visualization of vibrations

The spatio-temporal maps K and K_2 , shown in Figure 5, illustrate the process of vibration in the larynx region, defined by the vertical line on the B-mode ultrasound. The maps show the time evolution along the horizontal axis and the spatial variations along this vertical line of the fundamental frequency (K) and of the second harmonic (K_2). For this example, the me-

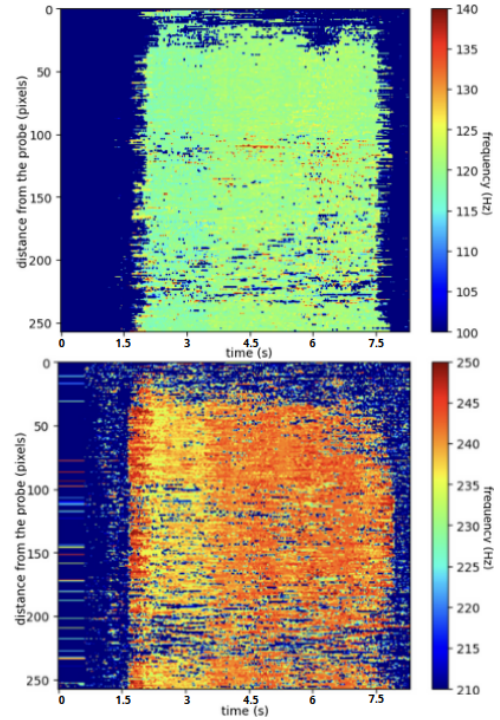


FIGURE 5 – Spatio-temporal maps showing fundamental frequency (K : top) and second harmonic (K_2 : bottom). Colors indicate frequency values in Hz. The horizontal axis corresponds to the temporal axis while the vertical axis corresponds to the 1D spatial axis (the vertical line defined in the larynx on the B-mode image).

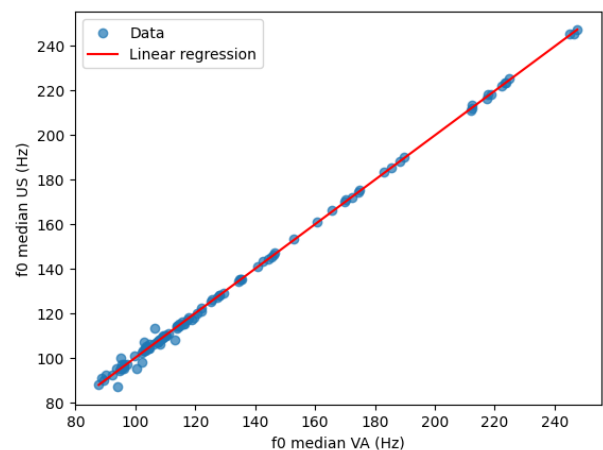


FIGURE 6 – Scatter plot of median f_0 values issued from voice recordings and M-mode ultrasound.

dian values are 118 Hz on the fundamental harmonic map and 237 Hz on the second harmonic map, which corresponds to the expected value of $2 \cdot f_0^{US}$. Spatial homogeneity of frequencies on the line drawn along the larynx seems to be higher than the temporal homogeneity where the temporal drift is clearly perceptible on the K and K_2 maps. This drift is coherent between the two maps, showing the relevance of our M-mode ultrasound approach. From the maps, we can also notice that the estimation of frequencies is out of the bounds retained for the display (blue points on the maps), particularly when the distance to the ultrasound probe is large, and the attenuation of the ultrasound signal is high.

4.2. Comparison of M-mode ultrasound to reference value

Figure 6 illustrates the high similarity between the median fundamental frequency as estimated by the ultrasound M-mode approach and the one estimated by the analysis of voice recording.

The linear regression between f_0^{US} and f_0^{VA} is given by :

$$f_0^{US} = 0.997 \cdot f_0^{VA} + 0.293 \quad (2)$$

with a correlation coefficient of 0.999.

The difference between the two types of measurement is less than 2 Hz in 92% of recordings, the highest difference being 7 Hz.

5. Discussion

In this work, we explored the use of M-mode ultrasound as a non-invasive method for analyzing vocal fold vibration. This technique captures high speed motion of reflectors, enabling highly temporal visualization of vibration in the laryngeal region defined by a 1D line. The frequency maps generated provide a spatio-temporal representation of f_0^{US} variations over time, and reveal patterns of vibration consistency.

Comparison between the fundamental frequency measured with M-mode ultrasound and the one obtained by simultaneous voice recording demonstrated a high degree of similarity between the two methods.

In the setup used for the current study, the sampling rate F_r was 500 Hz. On high pitched voices, for which $f_0 > 250$ Hz, we were confronted with aliasing effects. For example, in the case illustrated in Figure 7, the mean fundamental frequency, f_0^{US} , was found to be 220 Hz, following the M-mode ultrasound. However, the voice analysis shows a fundamental frequency f_0^{VA} equal to 280 Hz. The peak found at 220 Hz is clearly the impact of aliasing. Indeed, the spectral folding should be $F_r - 280$ Hz, i.e. 220 Hz. This occurred in 4 recordings (2 /i/ for a female subject, 2 /u/ for another female subject). These acquisitions were excluded from the database. However, in all 4 cases, the spectral aliasing was correctly estimated. Thus, for high pitched voices, due to the Shannon-Nyquist theorem, there may be ambiguities in the estimation of the fundamental frequency, using M-mode ultrasound.

Thus, the next step is increasing the sampling rate F_r of the M-mode ultrasound acquisition up to 2000 Hz, which will allow an estimation of frequencies up to 1000 Hz. This next step is crucial for obtaining accurate values and avoiding the aliasing effect on the first harmonics. As a result, we will study a wider scope of voices, and also assess the higher harmonics. New acquisitions at a sampling rate of 2000 Hz are underway, and preliminary results confirm the interest of this improved acquisition.

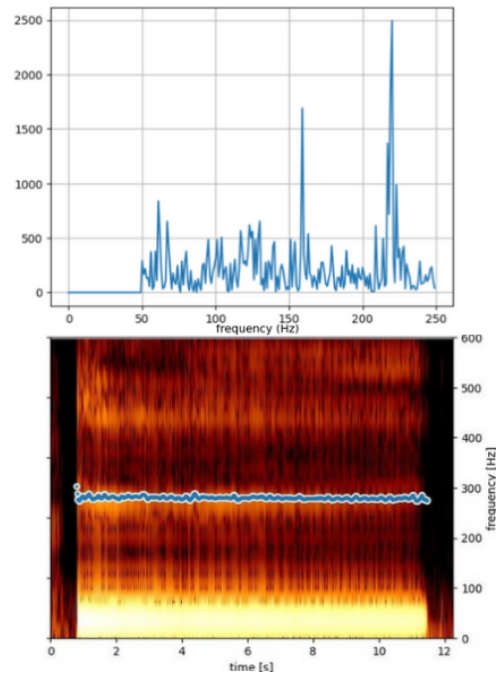


FIGURE 7 – Example of an aliasing effect obtained for a high-pitched voice, at 220 Hz. Corresponding spectrogram of the record is shown at the bottom of the figure, showing a fundamental frequency of 280 Hz.

By improving the proposed technique, M-mode ultrasound combined with B-mode ultrasound could provide a low-cost and non-invasive approach for studying vocal fold dysfunction, including vibration and mobility. It could provide an alternative tool to laryngoscopy before and after neck surgery.

6. Acknowledgments

The authors thank the French National Research Agency (ANR) for supporting the current project (VOCALISE, ANR-22-CE19-0035)

The Interspeech 2025 organisers would like to thank ISCA and the organizing committees of past Interspeech conferences for their help and for kindly providing the previous version of this template.

7. References

- [1] M. Heikkinen, K. Mäkinen, M. Qvarnström, T. Kemppainen, H. Löppönen, and J. Kärkkäinen, "Incidence, risk factors, and natural outcome of vocal fold paresis in 920 thyroid operations with routine pre- and postoperative laryngoscopic evaluation," *World Journal of Surgery*, vol. 43, no. 9, pp. 2228–2234, 2019.
- [2] M. Steurer, C. Passler, D. M. Denk, B. Schneider, B. Niederle, and W. Bigenzahn, "Advantages of recurrent laryngeal nerve identification in thyroidectomy and parathyroidectomy and the importance of preoperative and postoperative laryngoscopic examination in more than 1000 nerves at risk," *The Laryngoscope*, vol. 112, no. 1, pp. 124–133, January 2002.
- [3] C.-P. Wang, T.-C. Chen, T.-L. Yang, C.-N. Chen, C.-F. Lin, and P.-J. Lou, "Transcutaneous ultrasound for evaluation of vocal fold movement in patients with thyroid disease," *European Journal of Radiology*, vol. 81, no. 3, pp. e288–e291, March 2012.
- [4] K.-P. Wong, K.-P. Au, S. Lam, and B. H.-H. Lang, "Lessons learned after 1000 cases of transcutaneous laryngeal ultrasound

(tlusg) with laryngoscopic validation : Is there a role of tlusg in patients indicated for laryngoscopic examination before thyroidectomy ?” *Thyroid*, vol. 27, no. 1, pp. 88–94, January 2017.

- [5] D. Carneiro-Pla, B. S. Miller, S. M. Wilhelm, M. Milas, P. G. Gauger, and M. S. Cohen, “Feasibility of surgeon-performed transcutaneous vocal cord ultrasonography in identifying vocal cord mobility : A multi-institutional experience,” *Surgery*, vol. 156, no. 6, pp. 1597–1604, December 2014.
- [6] P. Knyazeva, V. Makarin, B. Seeliger, R. Chernikov, I. Sleptsov, and A. Semenov, “Transcutaneous laryngeal ultrasonography (tlus) as an alternative to direct flexible laryngoscopy (dff) in the perioperative evaluation of the vocal cord mobility in thyroid surgery,” *Langenbeck’s Archives of Surgery*, vol. 403, no. 8, pp. 1015–1020, December 2018.
- [7] D. S. Lazard, H. Bergeret-Cassagne, M. Lefort, L. Leenhardt, G. Russ, and F. Frouin, “Transcutaneous laryngeal ultrasonography for laryngeal immobility diagnosis in patients with voice disorders after thyroid/parathyroid surgery,” *World Journal of Surgery*, vol. 42, no. 7, pp. 2102–2108, July 2018.
- [8] S. Carerj, A. Micari, A. Trono, G. Giordano, M. Cerrito, C. Zito, F. Luzzza, S. Coglitore, F. Arrigo, and G. Oreto, “Anatomical m-mode : An old–new technique,” *Echocardiography : A Journal of Cardiovascular Ultrasound and Allied Techniques*, vol. 20, no. 4, pp. 357–361, May 2003.
- [9] P. Ballo, A. Bocelli, A. Motto, and S. Mondillo, “Concordance between m-mode, pulsed tissue doppler, and colour tissue doppler in the assessment of mitral annulus systolic excursion in normal subjects,” *European Journal of Echocardiography*, vol. 9, no. 6, pp. 748–753, May 2008.
- [10] A. V. Dieterich, C. M. Pickard, L. E. Deshon, G. R. Strauss, W. Gibson, P. Davey, and J. McKay, “M-mode ultrasound used to detect the onset of deep muscle activity,” *Journal of Electromyography and Kinesiology*, vol. 25, no. 2, pp. 224–231, April 2015.
- [11] P. Boersma and D. Weenink, “Praat : doing phonetics by computer [computer program],” 2025, version 6.4.27, retrieved 27 January 2025 from <http://www.praat.org/>.
- [12] Y. Jadoul, B. Thompson, and B. de Boer, “Introducing Parselmouth : A Python interface to Praat,” *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.
- [13] R. Vaysse, C. Astésano, and J. Farinas, “Performance analysis of various fundamental frequency estimation algorithms in the context of pathological speech,” *Journal of the Acoustical Society of America*, vol. 152, no. 5, pp. 3091–3101, 2022.