OBJECTIVE DETECTION AND TRACKING OF VOCAL FOLDS AUTO-OSCILLATION

Raphaël Chottin, Mohammad Ahmad, Didier Demolin, Xavier Pelorson, Annemie van Hirtum

forum acusticum 2023

OBJECTIVE DETECTION AND TRACKING OF VOCAL FOLDS AUTO-OSCILLATION

Raphaël Chottin[1] **Mohammad Ahmad**[1] **Didier Demolin**[2]
Xavier Pelorson[1] **Annemie Van Hirtum**[1*]
[1] CNRS-Université Grenoble Alpes/LEGI UMR 5519, Grenoble, France
[2] CNRS-Université Sorbonne-Nouvelle/LPP UMR 7648, Paris, France

ABSTRACT

A method for the objective detection of oscillations in physical signals is presented. It is based on YIN, a well-known fundamental frequency estimator often used in speech sciences. A dataset of physical signals such as the upstream pressure, generated using flow-induced auto-oscillating vocal folds replicas, is used to assess the method's efficiency in comparison to manual oscillation feature detection. The fluid-structure interaction obtained for six different deformable mechanical vocal folds replicas, either molded silicone composites or pressurized latex tubes, allows obtaining a database of auto-oscillation signals. Signals depend on the structure of the mechanical replicas and might exhibit complex behaviors such as subharmonic generation or time-changing fundamental frequency. The method's key parameters are discussed, evaluated and fine-tuned. It provides an objective, robust and accurate detection of the oscillation onset and offset characteristics. The method is thus able to detect fundamental oscillation frequencies and related pressures such as associated with onset, offset and steady oscillation regimes.

Keywords: *YIN algorithm, (ab-)normal vocal folds structure, (ab-)normal auto-oscillation, vocal fold replicas/human speech*

***Corresponding author**: *annemie.vanhirtum@univ-grenoble-alpes.fr.*

1. INTRODUCTION

A speech sound is voiced if its production includes auto-oscillation of the vocal folds. Its fundamental frequency, f_0, which will be perceived as the pitch, can be tracked. The close relation between the tasks of detecting voicing and tracking f_0 led to the creation of a wide variety of algorithms able to perform both, either in the frequency-domain [1, 2] or in the time-domain [3, 4]. Among the latter ones exists a method called YIN [5]. Its easy and intuitive implementation associated with its good performance in noisy conditions [6] made it popular in speech science. YIN is largely used on acoustic signals associated to speech or bio-acoustics, but rarely on other physical signals. In this paper, we attempt to show that the use of YIN for non-acoustic signals such as upstream airflow pressure or vocal folds displacement is of interest. Physical signals (pressure *etc.*) are obtained from physical experiments with mechanical vocal folds replicas without or with a structural abnormality, which represents either a local rigidification within the vocal fold structure or a growth on its surface. In this case complex vibration dynamics can occur involving a frequency variation in time or even the presence of subharmonics. YIN algorithm and its application in this context is outlined and results are presented and discussed. Finally, an application of YIN as a voicing detector is illustrated using human speech.

2. METHODS

2.1 Physical data

Physical data are obtained from controlled fluid-structure interaction experiments. Flow-induced vibration of vocal folds (VF) is mimicked using deformable mechanical

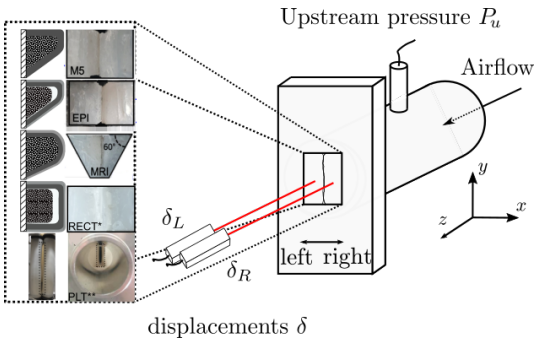
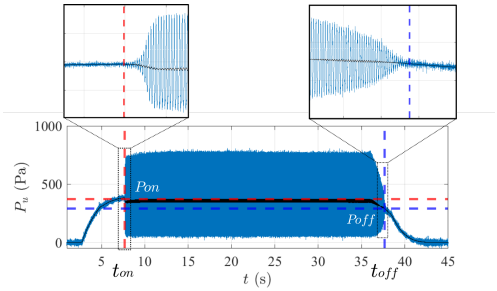
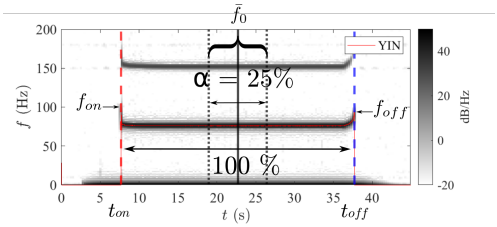


Figure 1: Schematic of the experimental setup with airflow supply and assessed vocal folds replicas: M5, EPI, MRI, rectangular with inclusion oriented parallel ($R^{\parallel}$) or serial ($R^{\perp}$) to the left-right direction and PLT without or with a surface growth [13, 14]. Upstream pressure $P_u(t)$ and vocal folds displacement $\delta(t)$ on the left and/or right vocal folds are measured. *Rectangular replicas displacement is gathered on the left vocal fold only. **No vocal fold displacement is measured for the PLT replica.



(a) upstream pressure time signal, $P_u(t)$



(b) P_u spectrogram and $f_0(t)$ trace from YIN analysis

Figure 2: Typical: a) time signal $P_u(t)$, b) spectrogram and YIN of a flow-induced auto-oscillation with the EPI replica. α indicates percentages setting the extent (arrows) of the steady oscillation portion used to estimate steady frequency $\bar{f}_0$. YIN is performed over a 5 ms window with threshold $\sigma = 0.1$.

VFs replicas, *i.e.* composite silicone-molded replicas or a Pressurized Latex Tubes replica (PLT). Five multi-layer silicone VF replicas are used, labeled M5 [7], MRI [8], EPI [9] and rectangular with an embedded stiff inclusion which is oriented either parallel ($R^{\parallel}$) or serial ($R^{\perp}$) to the main left-right vibration direction along the x-axis [10]. The replicas are composed of respectively, two (M5), three (MRI) and four (EPI, $R^{\parallel}$, $R^{\perp}$) overlapping silicone layers with different elasticity [10, 11]. The PLT replica is assessed without and with a surface growth as outlined in [12]. Two spherical growths are assessed, *i.e.* L1 with diameter 2.75 mm and L2 with diameter 3.75 mm. Replicas are mounted to an experimental airflow supply setup [10, 13] illustrated in Fig. 1. The setup allows measuring pertinent physical quantities in controlled conditions, *i.e.* upstream pressure $P_u(t)$ (pressure transducer Endevco 8507C-5) and displacement $\delta(t)$ (optic measurement with laser transceiver Panasonic HL-G112-A-C5, wavelength 655 mm) of the left ($\delta_L(t)$) or right ($\delta_R(t)$) VF in the main airflow direction along the z-axis. For the rectangular replicas with embedded inclusion ($R^{\parallel}$ and $R^{\perp}$) displacement is measured on the left VF only. For the PLT replica no VF displacement is measured and only auto-oscillation onset is experimentally assessed. Whereas the elasticity of molded composite silicone replicas is constant, the rigidity of the PLT replica is controlled by varying internal pressure P_w between 10 up to 60 cm H_2O.

During each experiment, the upstream pressure is gradually increased until the initiation of the replicas' auto-oscillation at an onset pressure, P_{on} a time, t_{on} and with a frequency, f_{on}. After a steady oscillation regime is established and maintained at a frequency, $\bar{f}_0$, the upstream pressure is reduced and the replicas cease to vibrate at a pressure P_{off}, a time t_{off} and with a frequency f_{off}. All of those informations can be found manually on the time-trace (tm) or spectrogram (sm) of physical quantities as illustrated for upstream pressure in Fig. 2. All physical signals are sampled at 10 kHz. A typical harmonic upstream pressure signal measured during a fluid-structure experiment with any mechanical vocal fold replica representing a normal vocal fold (M5, MRI, EPI or PLT without surface growth) is plotted in Fig. 2. Upstream pressure P_u

is plotted as a function of time t in Fig. 2(a) and its frequency (f) variation is illustrated considering the spectrogram in Fig. 2(b).

2.2 YIN algorithm

The YIN instantaneous fundamental frequency (f_0) estimator developed by De Cheveigné and Kawahara [5] is widely used in speech processing. The YIN algorithm is based on six steps that aim to improve the performance of autocorrelation-based techniques. The originality of YIN compared to other f_0-estimators based on auto-correlation lies in its second step. In the general case, the estimate of the period of a periodic signal s is the first (resp. last) non-zero time lag (τ) maximum of the auto-correlation function (YIN step 1). In the case of YIN, a difference function $d_t(\tau)$ at time index t,

$$d_t(\tau) = \sum_{j=1}^{W} (s_j - s_{j+\tau})^2 \tag{1}$$

is defined with integration window size W, which can be expressed in the form of a sum of auto-correlations (YIN step 2). The estimated period T is then the first non-zero lag minimum of the difference function which reduces errors (f_0 underestimation) related to amplitude changes.

Eq. 1 is then normalized by its cumulative mean to prevent extreme low values around the zero-lag (f_0 overestimation), e.g. due to imperfect periodicity or due to a strong resonance at the first harmonic of f_0, to be interpreted as period T. The resulting cumulative mean normalized difference function (cmndf) (YIN step 3) is proportional to the aperiodic/total power ratio and defined as:

$$d'_t(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ d_t(\tau) \Big/ \left[\frac{1}{\tau} \sum_{j=1}^{\tau} d_t(j) \right] & \text{otherwise.} \end{cases} \tag{2}$$

Following that step, an absolute cmndf threshold σ is defined (YIN step 4). The lag associated with a local cmndf minimum will be considered as a genuine period estimate only if its value $d'_t(\tau)$ is below this threshold. This step has the effect of preventing to interpret subharmonics as the fundamental period (f_0 underestimation) by giving an importance to the amplitude of the matching periods in the difference function. Next, a parabolic interpolation is performed around the dip where the period has been identified to increase accuracy of the estimation (YIN step 5). Finally, the best local estimate is sought for every time t (YIN step 6). In the vicinity of every point, the algorithm looks for the lowest value of the normalized difference function associated to a period T in an interval of $[t - T_{max}; t + T_{max}]$ with T_{max} the maximum period expected and associates T to the time t.

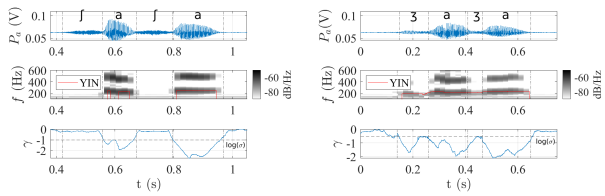
In order to perform YIN, the time signal has to be windowed. The size of that window W defines the maximum period T_{max} that can be observed. That feature, if tuned correctly using a prior knowledge of the signal, can be an advantage, but has to be chosen carefully. Another important parameter is threshold σ of step 4. A wrong tuning of this parameter can lead to a sub-harmonic being mistaken as the fundamental period (if too high) or even to no period detected at all (if too low). The latter case tends to happen when the signal-to-noise ratio (SNR) is low. Generally, a lower SNR implies the necessity of using a higher threshold to detect a period and thus frequency f_0.

When no frequency is detected, *i.e.* when cmndf remains above threshold σ, the algorithm returns a zero-frequency value. This point is important, because YIN can then not only be used to track the frequency of the oscillation, but also detect the existence of it. That characteristic leads to YIN as a voiced/unvoiced classifier,which, as a reference, is illustrated in Section 3.1.

3. YIN APPLICATION

3.1 Human speech sound as a reference

Fig. 3 illustrates two successive consonant-vowel (CV) or CVCV utterances by a male adult native French speaker [15]. Uttered consonants are either unvoiced postalveolar fricative /ʃ/ as in **sh**oe or its voiced variation /ʒ/ as in trea**s**ure. The vowel is /a/ as in h**a**t. It is clearly shown that YIN is able to track voicing with a continuous detection of fundamental frequency on the voiced CVCV and an interruption during the unvoiced part of unvoiced CVCV. In addition, YIN fundamental frequency detection criterium expressed as $\gamma(t) \leq \log(\sigma)$ with $\gamma = \log(\min(d'_t(\tau)))$ and threshold $\sigma = 0.3$ is indicated. Plotted γ values tend to be lower during the vowel than during the voiced fricative. This indicates that turbulent flow characterising this consonant appears as noise in acoustic pressure P_a, which tends to 'reduce' its apparent periodicity. Such property, if studied more thoroughly, might open fields of application related, for example, to the study of voiced fricatives. An additional step would be to apply YIN on physical signals associated with speech production. This is firstly assessed using mechanical VF



(a) with unvoiced fricative (b) with voiced fricative

Figure 3: Acoustic pressure $P_a(t)$, spectra with YIN $f_0(t)$ and $\gamma(t)$ with threshold $\log\sigma$ for CVCV utterances with: a) unvoiced fricative, b) voiced fricative.

replicas	normal				abnormal		both
	MRI & M5		EPI		$R^{\parallel}$ & $R^{\perp}$		PLT
signal	P_u	δ	P_u	δ	P_u	δ	P_u
$\sigma^{\diamond}$	0.2	0.3	0.3	–†	0.1	0.3	0.1
β	3	5	3	–‡	3	6	3

$^{\diamond}$As a reference, $\sigma = 0.3$ for speech in section 3.1
† assessed range $0.1 \leq \sigma \leq 0.6$.
‡ assessed range $1 \leq \beta \leq 7$.

Table 1: Overview of varied β-YIN parameters for VF replicas with normal and abnormal structure.

replicas as outlined in the next section 3.2.

3.2 Auto-oscillation on VF replicas: YIN to β-YIN

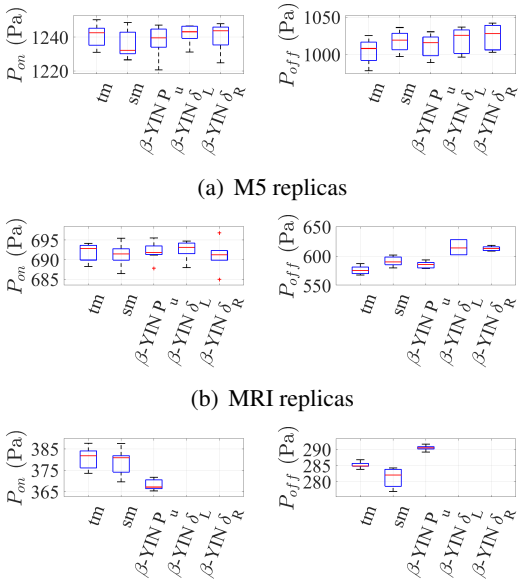
In order to obtain an objective measurement of both onset and offset pressures P_{on} and P_{off}, YIN has been applied with an additional step. Indeed, YIN is a fundamental frequency estimator, thus, if the signal is not periodic, no frequency (or rather a zero frequency) should be detected by YIN. Although the last step of YIN aims to prevent this, it may happen that a frequency is detected on a short-period of time (or even one unique point) due to either a physical reality (short vibration of the replica that ceases after a few periods) or due to an unfortunate effect of the noise. In order to avoid detecting those events as the oscillation start, the onset (resp. offset) time t_{on} (resp. t_{off}) is defined as the first non-zero point followed (resp. preceded) by βW non-zero points with $\beta \geq 0$ being a multiplicative coefficient of the window W on which YIN is applied. The idea behind using β is to find the first oscillation lasting for at least a given minimum number of periods. The onset and offset pressures can then be obtained by taking the value of the pressure at the onset and offset time averaged over a duration T_{avg} around that time, here 10 ms. This adapted version of YIN is called β-YIN. An advantage of using β-YIN is the automatic access to onset (resp. offset) frequency f_{on} (resp. f_{off}). This information is not accessible immediately when performing a manual estimation on the time signal, and may vary in the case of a manual detection on a spectrogram. We have chosen to use a window centered around the point of analysis. Other options of backward or forward windows are expected to produce slightly better results for either onset or offset but dramatically worse ones for the other one. For example, a forward window would spot oscillation onset earlier (even maybe too early depending on the size of window) but would also detect oscillation offset earlier and thus likely increasing the discrepancy when comparing to manual estimates of *e.g.* threshold pressures at onset and offset. This issue has, however, limited consequences on the pressure estimation in the case of slow variation of the oscillation amplitude at onset or offset. The steady oscillation frequency $\bar{f}_0$ is estimated as the averaged $f_0(t)$ around $t_{st} = (t_{on} + t_{off})/2$ on a time interval corresponding to a percentage α of the duration of oscillation. A constant value of $\alpha = 25\%$ is used in this work. Estimated features will thus depend on two analysis parameters σ and β as $T_{max} = 17ms$ (which can also be expressed as $f_{min} = 1/T_{max}$) and $\alpha = 25\%$ are held constant. An overview of the non-constant β-YIN analysis parameters σ and β for all VF replicas with normal or abnormal structure for all assessed signals is provided in Table 1. Parameter values σ and β are set in order to favour the agreement between extracted features, either manually or using β-YIN.

4. RESULTS

4.1 Auto-oscillation for normal VF replicas

4.1.1 Silicone replicas: M5, MRI and EPI

β-YIN has been applied over a set of seven repeated fluid-structure interaction experiments for EPI, M5 and MRI replicas. Onset and offset threshold values t_{on}, t_{off}, P_{on} and P_{off} are firstly detected on P_u using manual estimation on the time signal (tm), spectrogram (sm) and β-YIN analysis (β-YIN P_u). In addition, the β-YIN approach is applied to detect onset and offset times on left (β-YIN δ_L) or/and right (β-YIN δ_R) displacement signals. Associated onset and offset pressures with these

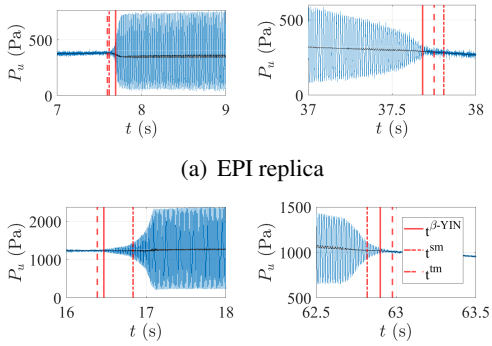


(a) M5 replicas

(b) MRI replicas

(c) EPI replicas

Figure 4: Boxplots of P_{on} and P_{off} for a) M5, b) MRI and c) EPI replicas: either manually on $P_u(t)$ (tm) and on its spectrogram (sm) or using β-YIN on $P_u(t)$ (β-YIN P_u), on left (β-YIN δ_L) and right (β-YIN δ_R) VF displacement.



(a) EPI replica

(b) M5 replica

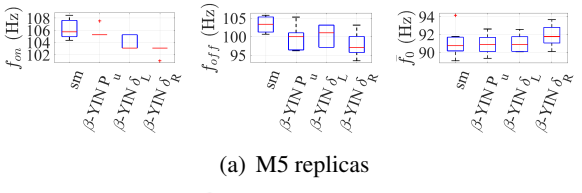
Figure 5: Manual and β-YIN oscillation onset and offset detection on $P_u(t)$ for replica: a) EPI and b) M5. Lines indicate t_{on} and t_{off} from β-YIN ($t^{\beta-YIN}$, full), manual spectrogram (t^{sm}, dashed-dotted) and manual time signal (t^{tm}, dashed).

times are then determined on the smoothed pressure signal as $P(t = t_{on})$ and $P(t = t_{off})$. Statistics of detected on and offset pressures P_{on} and P_{off} are plotted in Fig. 4 showing boxplots indicating the median value, interquartile range between the first and third quartile and extrema.

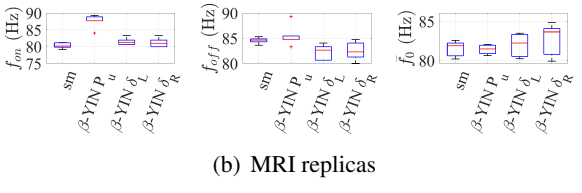
Overall, all five detection methods result in threshold pressures P_{on} and P_{off} of the same order of magnitude, which decreases between the M5, MRI and EPI replicas. For the M5 and MRI replicas all P_{on} estimates are within the same interval. For the EPI replica, β-YIN P_u slightly (10 Pa for the median value) underestimates manual estimates (tm and sm), whereas due to the low signal-to-noise ratio β-YIN could not detect periodic oscillation on the displacement signals. Considering P_{off}, manual estimates (tm and sm) and β-YIN P_u are again within the same interval for the M5 and MRI replicas whereas for the EPI replica β-YIN P_u slightly (10 Pa for the median value or up to 4%) overestimates manual estimates. Again due to the lower signal-to-noise ratio β-YIN $\delta_{L,R}$ overestimates (20 Pa for the median values for all replicas or up to 8%) P_{off} estimates obtained on P_u and the interquartile range associated with β-YIN $\delta_{L,R}$ tends to increase compared to the range observed for β-YIN P_u, which yields up to 10 Pa.

The increased discrepancy between manual and β-YIN estimates for the EPI replica compared to the M5 and MRI replica can be understood considering the gradient of the pressure envelope which is slowest for the EPI replica and steepest for the M5 replica as this gradient reflects the growth of the fluid-structure interaction instability or thus the auto-oscillation amplitude. This is further illustrated in Fig. 5 where a typical measurement of $P_u(t)$ is plotted near oscillation onset t_{on} and offset t_{off} for both the EPI and M5 replicas. A slow gradient (EPI replica) tends to delay/advance β-YIN detection compared to manual detection associated with oscillation onset/offset compared to a steeper gradient (M5 replica). This illustrates a windowing effect of β-YIN. Indeed, in order to detect oscillation, *i.e* cmndf (YIN step 3) to be smaller than absolute threshold σ (YIN step 4), the windowed signal on which β-YIN is applied has to be composed of a sufficient number of periods with sufficient amplitude.

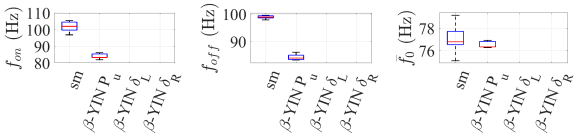
Statistics of detected frequencies associated with oscillation onset f_{on}, steady oscillation $\bar{f}_0$ and oscillation offset f_{off} for all three replicas are plotted in Fig. 6. For manual detection the frequency is obtained from the spec-



(a) M5 replicas



(b) MRI replicas



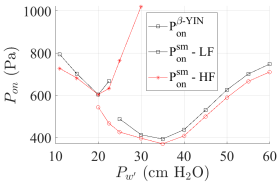
(c) EPI replicas

Figure 6: Onset f_{on}, offset f_{on} and steady $\bar{f}_0$ oscillation frequencies detected manually on the spectrogram (sm) or using β-YIN (β-YIN P_u, β-YIN δ_L, β-YIN δ_R) for replicas: a) M5, b) MRI and c) EPI.

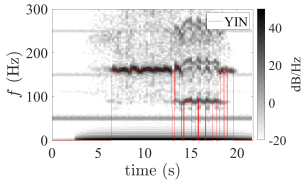
trogram (sm) whereas β-YIN estimates are obtained for each of the cases outlined before, *i.e.* β-YIN P_u, β-YIN δ_L and β-YIN δ_R. For all replicas, manual and β-YIN estimates of $\bar{f}_0$ are in good agreement as median values differ up to 4 Hz at most. The discrepancy between different estimates of f_{on} and f_{off} increases as median values differ up to 20 Hz. Moreover the discrepancy is larger for the EPI replica (up to 18%) than for the MRI and M5 replicas (up to 10%). The increased difference between manual and β-YIN detected values is again attributed to delayed/advanced detection of oscillation onset and offset (see Fig. 5).

4.1.2 Pressurized Latex Tubes replica

The upstream pressure is analysed using β-YIN with parameters given in Table 1. β-YIN estimated oscillation onset features t_{on}, f_{on} and P_{on} are compared to manually detected values on the spectrogram (sm) reported in [12]. Results for P_{on} as a function of internal pressure P_w are plotted in Fig. 7(a). Two oscillation regimes with low and high f_{on}, labelled LF and HF respectively, have been observed manually as illustrated in Fig 7(b). For $P_w > 30$ cm H_2O, only the HF regime is manually observed. In this P_w-range, β-YIN and manual estimates are in good agreement as β-YIN consistently overestimates manual P_{on} estimates with 5%. As outlined for silicone replicas in Section 4.1.1 (Fig. 5), the overestimation is again due to a delayed auto-oscillation onset detection with β-YIN compared to manual detection. For $P_w \leq 30$ cm H_2O, β-YIN detects a single oscillation frequency whereas for $20 \leq P_w \leq 30$ cm H_2O both oscillation regimes are manually detected (Fig. 7(b)).



(a) $P_{on}(P_w)$



(b) $P_w = 25$ cm H_2O

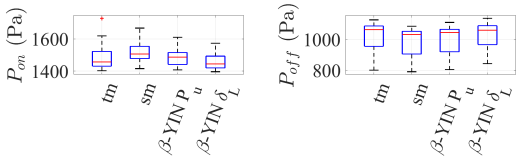
Figure 7: a) Manual (superscript sm) and β-YIN (superscript β-YIN) P_{on} estimates for the PLT replica as a function of internal pressure P_w: β-YIN regime ($P_{on}^{\beta-YIN}$) and manual low (P_{on}^{sm} - LF) and high (P_{on}^{sm} - HF) frequency regimes. b) spectrogram of P_u for $P_w = 25$ cm H_2O.

4.2 Auto-oscillation for abnormal VF replicas

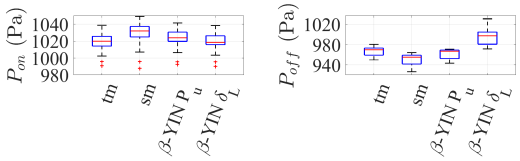
4.2.1 Rectangular silicone replicas with inclusions

β-YIN has been applied over a set of 19 repeated fluid-structure interaction experiments with both rectangular composite silicone replicas embedding respectively a perpendicular ($R^{\perp}$) or a parallel ($R^{\parallel}$) inclusion with respect to the main left-right auto-oscillation direction along the x-axis (Fig. 1). As outlined for the silicone composite replicas without inclusion M5, MRI and EPI in section 4.1.1, β-YIN is applied to the upstream pressure (β-YIN P_u) and to the left (β-YIN δ_L) and right (β-YIN δ_R) vocal folds displacement signals and associated oscillation features are compared to manually detected values on the time-trace (tm) and spectrogram (sm) of measured upstream pressures P_u. Statistics on detected onset and offset pressures P_{on} and P_{off} are plotted in Fig. 8. Overall P_{on} estimated for $R^{\perp}$ are greater (median values increase with > 400 Pa or 30% at least) than values found for $R^{\parallel}$. The same is observed for P_{off}, but the decrease in reduced to about 100 Pa for median values (or about 10%). Furthermore, the interquartile range between the first and

third quartile observed for both P_{on} and P_{off} yields up to 200 Pa for $R^{\perp}$ whereas for $R^{\parallel}$ it is limited to about 20 Pa. This suggests that a serial inclusion introduces more perturbations resulting in harmonic distortion or frequency variation than observed for a parallel inclusion. Moreover, the 20 Pa range observed for $R^{\parallel}$ is similar to the 10 Pa range found for P_{on} on replicas M5, MRI and EPI without inclusion (Section 4.1.1). Median values of P_{on} and P_{off} obtained using manual estimates (tm and sm) are in good agreement with β-YIN P_u and β-YIN $\delta_{L,R}$ as the discrepancy is limited up to 2% and up to 5%, respectively. Compared to the steady harmonic spectra characterising



(a) Rectangular with serial inclusion, $R^{\perp}$



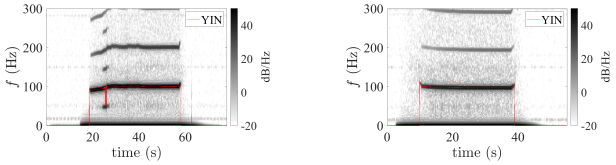
(b) Rectangular with parallel inclusion, $R^{\parallel}$

Figure 8: Boxplots of P_{on} and P_{off} for rectangular replicas with inclusion: a) $R^{\perp}$ and b) $R^{\parallel}$: either manually on $P_u(t)$ (tm) and on its spectrogram (sm) or using β-YIN on $P_u(t)$ (β-YIN P_u), on left (β-YIN δ_L) and right (β-YIN δ_R) VF displacement.

replicas M5, MRI and EPI without inclusion illustrated in Fig. 2(b), more complex time-frequency behaviour is observed for silicone replicas with inclusions. Indeed, the fluid-structure interaction can lead to a time-varying fundamental frequency or generate sub-harmonics as illustrated in Fig. 9 for $R^{\perp}$. Overall, β-YIN tracks $f_0(t)$ accurately and an appropriate tuning of the threshold σ avoids confusing with sub-harmonics as long as these are not too important. In any case, β-YIN detects oscillation onset and offset.

4.2.2 *Pressurized latex tubes with growths*

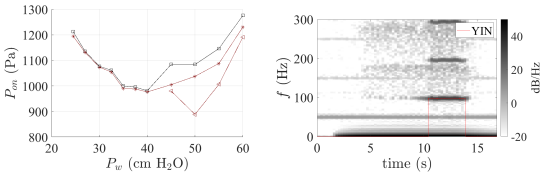
β-YIN estimated oscillation onset features t_{on}, f_{on} and P_{on} are compared to manually detected values on the



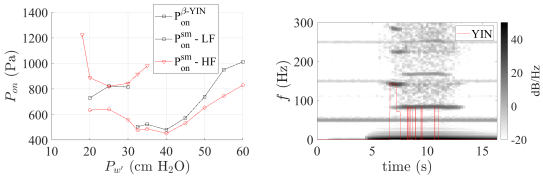
(a) $R^{\perp}$ (b) $R^{\parallel}$

Figure 9: Spectrograms and YIN for $R^{\perp}$ and $R^{\parallel}$.

spectrogram (sm) given in [12]. Results for $P_{on}(P_w)$ are plotted in Fig. 10. The same findings as for the PLT replica without a growth (Section 4.1.2) holds as β-YIN detection is limited to a single regime whereas manually two regimes can be observed on the spectrogram in the case of growth L2 and $P_w = 20$ cm H_2O.



(a) L1 growth, spectrogram for $P_w = 32.5$ cm H_2O



(b) L2 growth, spectrogram for $P_w = 20$ cm H_2O

Figure 10: $P_{on}(P_w)$ for manual (sm) regimes (LF and HF) and β-YIN regime of the PLT replica with surface growth: a) L1 and b) L2. Spectrograms are illustrated.

5. CONCLUSION

The YIN fundamental frequency tracking algorithm is extended to β-YIN by introducing an additional parameter. This parameter allows to identify and quantify the auto-oscillation onset and offset on physical signals by imposing a minimum threshold length on a sequence of non-zero

YIN frequencies. The ability of β-YIN to objectively detect, track and quantify auto-oscillation features is then assessed on physical signals measured using different deformable vocal fold replicas. Spectra of measured signals are either harmonic or reflect more complex oscillation behaviour such as sub-harmonics generation, frequency jumps or even non-harmonic behaviour. Overall, it is shown that estimated oscillation features for harmonic signals using β-YIN are in good agreement with manually extracted features as long as the noise level is low. As the noise level is elevated such as on the displacement signals the discrepancy with manually detected reference values increases or even fail to detect auto-oscillation. Weak sub-harmonics do not prohibit the efficiency of β-YIN, however as strong sub-harmonics are present β-YIN tends to identify the sub-harmonic frequency instead. The same limitation holds for non-harmonic behaviour exhibiting two different frequencies. Moreover, frequency jumps, indicating more than one auto-oscillation regime, are currently not detected as only a single steady oscillation frequency is sought for. Thus current results encourage to further extend the YIN approach in the case of noisy signals or in the case of signals exhibiting complex time-frequency behaviour. The use of probabilistic YIN [16] (PYIN) may be an efficient alternative as it associates a probability to potential fundamental frequencies. Moreover, it may be interesting to compare β-YIN to standard methods in signal processing such as Kalman filter for tracking [17] and MUSIC for frequency estimation. This work opens the door to application of YIN on non-acoustical signals associated to speech production measured on vocal folds replicas or on human speakers whether or not suffering from a vocal fold pathology.

6. ACKNOWLEDGEMENTS

This project has received financial support from the CNRS through the 80|Prime program.

7. REFERENCES

[1] A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41, no. 2, pp. 293–309, 1967.

[2] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 124, no. 3, 2008.

[3] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 1, pp. 24–33, 1977.

[4] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, pp. 353–362, 1974.

[5] A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, p. 14, 2002.

[6] L. Sukhostat and Y. Imamverdiyev, "A comparative analysis of pitch detection methods under the influence of different noise conditions," *J. Voice*, vol. 29, no. 4, pp. 410–417, 2015.

[7] R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," *J. Acoust. Soc. Am.*, vol. 109, no. 4, pp. 1616–1630, 2001.

[8] B. A. Pickup and S. L. Thomson, "Flow-induced vibratory response of idealized versus magnetic resonance imaging-based synthetic vocal fold models," *J. Acoust. Soc. Am.*, vol. 128, no. 3, pp. EL124–EL129, 2010.

[9] P. R. Murray and S. L. Thomson, "Vibratory responses of synthetic, self-oscillating vocal fold models," *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. 3428–3438, 2012.

[10] M. Ahmad, *Study of the Influence of Structural Properties on the Fluid-Structure Interaction of Artificial Vocal Folds*. PhD thesis, Université Grenoble Alpes, 2023.

[11] P. R. Murray and S. L. Thomson, "Synthetic, multi-layer, self-oscillating vocal fold model fabrication," *JoVE*, no. 58, p. 3498, 2011.

[12] P. Luizard and X. Pelorson, "Threshold of oscillation of a vocal fold replica with unilateral surface growths," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. 3050–3058, 2017.

[13] A. Bouvet, I. Tokuda, X. Pelorson, and A. Van Hirtum, "Influence of level difference due to vocal folds angular asymmetry on auto-oscillating replicas," *J. Acoust. Soc. Am.*, vol. 147, no. 2, pp. 1136–1145, 2020.

[14] A. Van Hirtum and X. Pelorson, "High-speed imaging to study an auto-oscillating vocal fold replica for different initial conditions," *Int. J. Appl. Mechanics*, vol. 09, no. 05, p. 1750064, 2017.

[15] D. Demolin, S. Hassid, C. Ponchard, S. Yu, and R. Trouville, "Speech aerodynamics database," 2019.

[16] M. Mauch and S. Dixon, "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP*, (Firenze, Italy), pp. 659–663, IEEE, 2014.

[17] A. Dardanelli, S. Corbetta, I. Boniolo, S. Savaresi, and S. Bittanti, "Model-based kalman filtering approaches for frequency tracking," *IFAC Proceedings Volumes*, vol. 43, no. 10, pp. 37–42, 2010.