IS THE RELATIVE FUNDAMENTAL FREQUENCY AN ACOUSTIC CORRELATE OF LARYNGEAL TENSION IN PORTUGUESE SPEAKERS?

Luis M. T. Jesus[1,2], Sara Castilho[1], Andreia Hall[1,3]

[1] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, Portugal
[2] School of Health Sciences (ESSUA), University of Aveiro, Portugal
[3] Department of Mathematics (DMat), University of Aveiro, Portugal

lmtj@ua.pt, sara.castilho@ua.pt, andreia.hall@ua.pt

ABSTRACT

In this study the relative fundamental frequency (RFF) of healthy speakers (N=24) and speakers with voice disorders (vocal nodules N=9; Reinke's edema N=15) and its relation with laryngeal tension, were analysed. Ten VCV sequences from a phonetically balanced text were selected to calculate RFF values, and vowel [a] was used to estimate the mean fundamental frequency, jitter, shimmer, and mean harmonics-noise-ratio. Strain was also perceptually assessed with the GRBAS scale. Analysis-of-variance was used to compare the RFF values from the three different groups but no statistically significant differences were found. There were, however statistically significant differences between the cycles. The RFF values within each group were widely dispersed. Contrary to what has been previously claimed in the literature it was not possible to establish a correlation between RFF and laryngeal tension, and it is hard to see the applicability of the RFF measure in a clinical context.

Keywords: Voice; relative fundamental frequency; laryngeal tension; vocal nodules; Reinke's edema.

1. INTRODUCTION

Acoustic voice analysis is an effective and non-invasive tool that can be used to confirm an initial diagnosis and provide an objective identification of impairment. Despite the variety of models and methods developed by signal processing engineers, voice clinicians still express their disappointment with regard to the performance of the existing approaches for assessing voice quality.

Vocal hyperfunction, characterised by an increased muscle tension and effort [3], accounts for 10-40% of individuals who are referred to multidisciplinary voice clinics [17]. Excessive laryngeal and paralaryngeal tension may cause voice disorders with or without organic changes on the surface of the vocal folds. Vocal effort and overall severity are usually assessed through auditory-perceptual scales, but parameters derived from acoustic signals [14] have recently been shown to have great potential as diagnostic tools [15] and outcome measures of voice therapy [16]. There are also some recent results [4, 17] showing some degree of correlation between the perceived effort (strain) and the relative fundamental frequency (RFF) during the onset and offset of voicing, in patients with vocal hyperfunction.

Fundamental frequency (F0) in vowels surrounding voiceless consonants has been shown [4, 9, 15, 18] to increase before the voiceless consonants and decrease after voiceless consonants, when compared with F0 values at the steady-state of the same vowels. These results are based on data from healthy American English (AE) speakers and could be related to the control of voicing offset and onset.

The RFF has been defined as an objective measure of the F0 of vocal fold oscillation cycles immediately before and after production of a voiceless consonants, normalised to the steady-state F0 values of the vowels [15]. Values have been previously presented in semitones (ST) to allow the comparison between subjects with different ranges of F0. The ten cycles before the consonant have been designated as RFF offset and the ten cycles after the consonant as RFF onset [17].

Recent work suggests the potential use of RFF changes surrounding voiceless consonant production in the assessment of vocal hyperfunction, but the physiological bases of RFF are still unclear and it will not be clinically useful until it can be automatically and reliably calculated [4].

Tension is hypothesised to increase before, during and immediately following a voiceless consonant production. This is a possible mechanism to inhibit voicing [9]. With vocal hyperfunction and excessive laryngeal tension, there are smaller short-term variations of F0, resulting in lower RFF values [15].

In this study the behaviour of RFF in European Portuguese (EP) healthy speakers and EP speakers with voice disorders (vocal fold nodules and Reinke's edema), and the relation between this measure and laryngeal tension, were analysed. Ten within word VCV sequences were selected from a phonetically balanced text [8], and acoustic parameters that have been previously reported as having a potential to capture essential clinical features [20] of patients with vocal fold nodules (F0 and intensity ranges are typically reduced) and Reinke's edema (F0 is typically lower than average and diminished in the upper range; hoarseness or roughness are observed), have been extracted from vowels in real words.

This paper identifies the current limitations of using an acoustic measure as an objective assessment of laryngeal tension that has been previously shown to have a clinical application for AE speakers, with a different population (EP speakers).

2. METHOD

In this study 48 EP speakers (24 without voice disorders, 9 with vocal nodules and 15 with Reinke's edema) were selected from the first representative EP pathological voice database [7]. Participants gave their informed consent and were subject to otolaryngological examination by a specialist consultant, including videostroboscopy assessment of vocal fold closure, regularity, mucosal wave, and symmetry.

Participants readings of a Portuguese version of the phonetically balanced text "The North Wind and the Sun" [8] were recorded in a clinical setting, using *Praat* 5.3.56 (32-bit edition) [2]. A Behringer ECM8000 microphone and a Presonus AudioBox USB (16 bits and 48000 Hz) were used for all of the recordings. The subjects were seated and the microphone was aligned to the mouth at a distance of 30 cm.

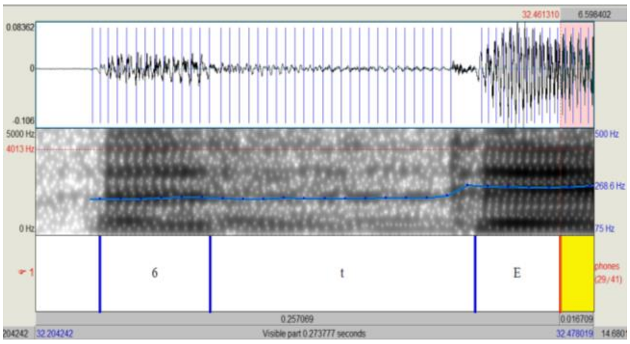
The present corpus was based on ten phonetic VCV sequences. Previous studies [15, 16] included VCV sequences between different words, VCCV sequences, and both stops and fricatives. In this study, only EP voiceless stops [p, t, k] and oral vowels [i, e, ɛ, a, ɐ, o, u] were selected. All the VCV sequences selected were produced within words, namely: 1×[uti] in [diʃkuˈtiɐ̃w̃]; 4×[apɐ] in [ˈkapɐ]; 1×[ɐko] in [ɐˈkordu]; 1×[ɐke] in [ɐˈkelɨ]; 1×[ɐtɛ] in [ɐˈtɛ]; 1×[atɐ] in [imediatɐˈmẽtɨ]; 1×[eku] in [Rekuɲɨˈseɾ]. It was not possible to calculate the RFF for some VCV sequences because there were less than ten cycles available in one of the vowels, the vowels had been reduced, or different phonemes had been produced. Still, RFF values were calculated from 331 out of 480 (68%) possible VCV sequences.

To compare the RFF with other measures that could potentially be correlated with laryngeal tension, the same ten cycles used to compute the RFF values for the vowel [a] in the word [imediatɐˈmẽtɨ] (the VCV sequence [atɐ] was also used to calculate the RFF) with *Praat* 5.3.56 (32-bit edition): mean F0, jitter (ppq5), shimmer (apq5) and the mean harmonics-to-noise (HNR).

2.1. RFF calculation

Each of the ten sequences of all subjects, prior to (RFF offset) and after (RFF onset) the voiceless consonant were manually annotated using *Praat* 5.3.56 (32-bit edition) and the pulse function was used to determine the beginning of each cycle, as shown in Fig. 1.

Figure 1: Annotation of one [ɐtɛ] sequence



The instantaneous fundamental frequency (f) of twenty cycles was extracted and then converted to semitones relative the fundamental frequency (fref) at the first cycle of voicing offset and at the tenth cycle of voicing onset using the following formula [1]:

$$ST = 39.86 * log10\left(\frac{f}{fref}\right) \qquad (1)$$

The RFF values for each subject were averaged across the ten offset and onset cycles [15].

2.2. Subjective laryngeal tension assessment

Subjective voice quality assessment of laryngeal tension with the GRBAS scale [6] was based on the expert opinion of three independent raters. Only the strained (S) parameter was considered because it is an auditory-perceptual quality related to vocal hyperfunction [18]. Listeners based their assessment on recordings of the "The North Wind and the Sun" passage.

2.3. Statistical analysis

Statistical analysis was performed using IBM SPSS software version 19.

Analysis-of-variance (ANOVA) was used to study the behaviour of RFF values.

The relation between RFF, mean F0, jitter, shimmer and HNR was analysed with Pearson's correlation coefficient. Normality was tested with the Kolmogorov-Smirnov (K-S) test. When there was not a normal distribution Spearman's coefficient was used to analyse these correlations.

The Kruskal-Wallis test was used to analyse the differences between the three groups of participants in terms of mean F0, jitter, shimmer, HNR, and S from the GRBAS scale.

A level of significance of 0.05 was used for all statistical analysis.

3. RESULTS AND DISCUSSION

Table 1 shows the mean RFF offset and onset values, and the range of RFF values from cycle 10 at offset and from cycle 1 at onset. These two instances in time have been previously suggested [16] as showing the greatest potential to differentiate degrees of vocal hyperfunction. They also correspond to the two cycles that are more distant from those used for the normalisation of data (cycle 1 for RFF offset and cycle 10 for RFF onset). Values shown in table 1 are quite different from those previously reported for AE speakers [16, 17] and data shows a much larger variance (see also Fig. 2). Previous studies by Stepp et al. [16, 17] had also reported a range of different values.

Table 1: Mean values of RFF offset and onset.

Diagnosis	Mean RFF offset (ST)	RFF offset - cycle 10 (ST)	RFF onset - cycle 1 (ST)	Mean RFF onset (ST)
Normal	-0.290	-2.225 to 1.776	2.324 to -0.658	0.216
Nodules	-0.176	-0.840 to 1.182	2.508 to -0.256	0.317
Reinke's Edema	-0.216	-2.923 to 1.567	0.977 to -2.491	0.022

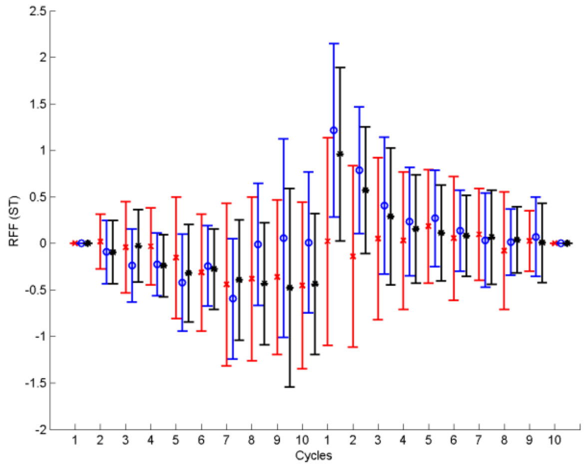
Analysis-of-variance (ANOVA) was used to study the effects of the variables *cycles* and *diagnosis*. Two ANOVAs were run, one for the RFF offset and the other one for the RFF onset.

At RFF offset there were statistically significant differences between cycles (p=0.017), but neither for the interaction (p=0.105) nor between the three groups of participants (p =0.888). At RFF onset there were significant differences between cycles (p<0.001) and for the interaction (p<0.001), but not in terms of diagnosis (p=0.201).

The two ANOVAs revealed statistical differences between cycles (p<0.001), which agrees with previous studies [16]. During RFF offset there were significant differences between cycle 2 and cycle 10 (p=0.0495), and during RFF onset the significant differences were found between cycle 1 and all other cycles (p<0.001).

Fig. 2 shows small differences between the three groups during RFF offset. However, RFF offset values tend to decrease in the group without voice disorders and in the Reinke's edema group. During RFF onset, the group with vocal nodules and without voice disorders have a similar behaviour, i.e., they both show an increase in the RFF during the first cycles, followed by a gradual decrease of those values along the remaining cycles. Contrary to what had been previously reported [5, 16, 18] the RFF onset values of the group with vocal nodules were higher than those of the group without voice disorders. Variability observed within Reinke's edema group could be related to the mucosa wave features found in this vocal pathology which range from hyperdynamic to hypodynamic vibratory patterns [19].

Figure 2: Mean RFF (cross; circle; asterisk) ± standard deviation of mean values in ST, for the three groups: Red – Reinke's edema; blue – vocal nodules; Black – without voice disorder. Left – RFF offset; right – RFF onset.



Although the ANOVAs did not reveal overall significant differences between the three groups, individual ANOVAs for each cycle were undertaken and the two first cycles of RFF onset revealed significant differences (p<0.001 in both cases). In these two cycles, post hoc tests showed that only the Reinke's edema group was significantly different from the other two, as expected from Fig. 2.

Previous studies results [15, 16], revealed some differences between groups, and like in the present study these differences only appeared in some of the cycles. In our data all groups had highly dispersed values (less so for the healthy speakers group) which clearly contributed to the non-significant differences between groups in most of the cycles and in the overall ANOVAs.

This variability in RFF values also results in mean RFF patterns which are quite different from other studies [5, 17], so previous interpretations of the data regarding the influence of laryngeal tension are not possible.

Table 2 shows the mean F0, jitter (ppq5), shimmer (apq5) and the mean harmonics-to-noise (HNR) for the three groups of participants.

Table 2: Fundamental frequency (F0), jitter, shimmer and the mean harmonics-to-noise (HNR) for the three groups of participants. The values shown correspond to (in this order): Mean; standard deviation of mean; median.

Diagnosis	Gender	F0 (Hz)	jitter (%)	shimmer (%)	HNR (dB)
Normal	Male (n=9)	127.73; 11.40; 128.96	0.61; 0.24; 0.65	10.23; 6.46; 8.90	6.69; 3.83; 5.94
Normal	Female (n=15)	237.90; 56.11; 230.08	0.33; 0.26; 0.29	8.22; 5.96; 6.68	10.11; 3.55; 10.16
Nodules	Male (n=1)	158.35	0.87	8.24	4.36
Nodules	Female (n=8)	200.85; 49.60; 216.56	0.50; 0.19; 0.47	0.07; 3.84; 6.42	9.54; 3.74; 8.01
Reinke's Edema	Male (n=8)	121.36; 13.65; 120.80	2.26; 3.96; 0.60	13.61; 8.82; 10.58	6.61; 3.47; 7.75
Reinke's Edema	Female (n=7)	182.46; 24.43; 193.00	1.35; 1.41; 1.04	9.37; 3.28; 10.60	7.45; 3.10; 8.63

In this study no significant correlations between RFF and jitter (R_s=-0.32; p=0.831), shimmer (R_s=-0.205; p=0.161), mean HNR (R_{xy}=-0.144; p=0.329), mean F0 (R_s=0.08; p=0.589) and parameter S from the GRBAS scale (R_s=-0.16; p=0.913) were found. All the corresponding scatter plots showed a great deal of dispersion.

A more controlled selection process of the VCV sequences (only voiceless stops and the VCV sequence was produced within words) than what had been previously used and a similar sample size to the most recent study [14], was presented in this paper. However, RFF values with greater dispersion were found, which could be attributed to the variability of clinical features in the patients group.

4. CONCLUSIONS

Relative fundamental frequency values in this study, when compared to previous studies, show variability not previously observed, which makes it difficult to find a pattern in RFF behaviour and discriminate pathology groups. This does not allow an adequate interpretation of the data collected with regard to the influence of laryngeal tension on RFF.

The RFF onset values of the group without voice disorders and with vocal nodules increases during the first cycles and decreases gradually along the remaining cycles. The Reinke's edema group did not follow this pattern which can be related with specific features of this vocal pathology.

The RFF values of the three groups were not significantly different. However, there were differences between the cycles during RFF offset and between the cycles during RFF onset.

The acoustic parameters jitter, shimmer, mean HNR and mean F0 were not significantly correlated with RFF. The RFF and laryngeal tension subjectively assessed did not have a correlation too.

Current understanding of aerodynamic, articulatory and acoustic interactions that govern the production principles involved in voicing of speech sounds, particularly voicing during consonant production, is still limited [10,13]. The use of stimuli in a rich variety of contexts (resulting in multiple within-word and across-word interaction effects) reveals details about production mechanisms resulting from real physiological conditions and requirements placed upon the speech system. Qualitatively and quantitatively defining non-modal voicing based on factors more closely related to phone production (laryngeal behaviour) than to the acoustic signal, could facilitate the exploration of relation between laryngeal activity and biomedical signals [13]. Recent findings on acoustic correlates of prosody related to voice quality and the role of the subglottal system in vocal fold vibration unveiled novel physiological and acoustical characteristics of the voice source. Still, literature is sparse concerning the different contributions (across languages) of acoustic parameters or auditory features for voicing distinction [11, 12].

Therefore, major issues that limit the effective assessment of vocal effort include: the consistency of measurements, language specific voicing strategies, objective interpretation of estimated features and correlation with perception.

These results question the relevance of RFF to assess laryngeal tension in different languages and clinical contexts.

5. ACKOWLEDGMENTS

This work was partially funded by National Funds through FCT - Foundation for Science and Technology, in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.

6. REFERENCES

[1] R. Baken and R. Orkiloff, *Clinical Measurement of Speech and Voice*, 2nd ed. San Diego: Singular, 2000.

[2] P. Boersma, “Praat, a system for doing phonetics by computer,” *Glot Int.*, vol. 5, no. 9/10, pp. 341–345, 2001.

[3] D. Boone, S. McFarlane, S. Berg, and R. Zraick, *The Voice and Voice Therapy*, 9th ed. Boston: Allyn & Bacon, 2013.

[4] T. L. Eadie and C. E. Stepp, “Acoustic correlate of vocal effort in spasmodic dysphonia,” *Ann. Otol. Rhinol. Laryngol.*, vol. 122, no. 3, pp. 169–176, Mar. 2013.

[5] A. M. Goberman and M. Blomgren, “Fundamental Frequency Change During Offset and Onset of Voicing in Individuals with Parkinson Disease,” *J. Voice*, vol. 22, pp. 178–191, 2008.

[6] M. Hirano, *Clinical Examination of Voice*. Vienna: Springer-Verlag, 1981.

[7] L. M. T. Jesus, “University of Aveiro’s Advanced Voice Function Assessment Databases (AVFAD),” *Rev. Saude Publica*, vol. 48, no. Special Number, p. 291, 2014.

[8] L. M. T. Jesus, A. Valente, and A. Hall, “Is the Portuguese Version of ‘The North Wind and the Sun’ Passage Phonetically Balanced?,” *J. Int. Phon. Assoc.*, vol. 45, no. 1, pp. 1–11, 2015.

[9] A. Löfqvist, T. Baer, N. McGarr, and R. Story, “The cricothyroid muscle in voicing control,” *J. Acoust. Soc. Am.*, vol. 85, no. 3, pp. 1314–1321, 1989.

[10] D. Pape and L. M. T. Jesus, “Stop and fricative devoicing in European Portuguese, Italian and German,” *Lang. Speech*, 2015.

[11] D. Pape and L. M. T. Jesus, “Cue-weighting in the perception of intervocalic stop voicing in European Portuguese,” *J. Acoust. Soc. Am.*, vol. 136, no. 3, pp. 1334–1343, 2014.

[12] D. Pape and L. M. T. Jesus, “Production and perception of velar stop (de)voicing in European Portuguese and Italian,” *EURASIP J. Audio, Speech, Music Process.*, vol. 2014, p. 6, 2014.

[13] C. M. R. Pinho, L. M. T. Jesus, and A. Barney, “Weak Voicing in Fricative Production,” *J. Phon.*, vol. 40, no. 5, pp. 625–638, 2012.

[14] C. E. Stepp, “Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson’s disease,” *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1637–1643, Mar. 2013.

[15] C. E. Stepp, R. E. Hillman, and J. T. Heaton, “The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset.,” *J. Speech. Lang. Hear. Res.*, vol. 53, pp. 1220–1226, 2010.

[16] C. E. Stepp, G. R. Merchant, J. T. Heaton, and R. E. Hillman, “Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction,” *J. Speech, Lang. Hear. Res.*, vol. 54, no. 5, pp. 1260–2266, Oct. 2011.

[17] C. E. Stepp, D. E. Sawin, and T. L. Eadie, “The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset,” *J. Speech, Lang. Hear. Res.*, vol. 55, no. 6, pp. 1887–1896, Dec. 2012.

[18] B. C. Watson, “Fundamental frequency during phonetically governed devoicing in normal young and aged speakers,” *J. Acoust. Soc. Am.*, vol. 103, no. 6, pp. 3642–3647, 1998.

[19] S. M. Zeitels, R. R. Casiano, G. M. Gardner, N. D. Hogikyan, J. A. Koufman, and C. A. Rosen, “Management of common voice problems: Committee report,” *Otolaryngol. - Head Neck Surg.*, vol. 126, pp. 333–348, 2002.

[20] K. Verdolin, C. A. Rosen, and R. C. Branski, Eds., *Classification manual for voice disorders–I*. Mahwah: LEA, 2006.