DETECTING ARYTENOID CARTILAGE MISPLACEMENT THROUGH ACOUSTIC AND ELECTROGLOTTOGRAPHIC JITTER ANALYSIS

Maurílio N. Vieira[1]~†, *Arnold G. D. Maran[2]*, *Fergus R. McInnes[1]*, *and Mervyn A. Jack[1]*
E-mail: maurilio, fmi, maj@ccir.ed.ac.uk

[1] CCIR, Dept. of Electrical Engineering
[2] Otolaryngology Unit, Royal Infirmary of Edinburgh
University of Edinburgh

ABSTRACT

This paper describes a comparative study of acoustic and electroglottographic (EGG) jitter extracted from simultaneous recordings of 45 dysphonic patients. The estimated jitter values agreed to within ±20% in 67% of the patients, the difference being above +20% and below −20% in 13% and 20% of the cases, respectively. In the group below −20%, EGG jitter increased due to an increase in EGG shimmer and videoendoscopic images revealed abnormal movements of the arytenoid cartilage caused, possibly, by minor selective paralysis of the intrinsic laryngeal muscles. This observation suggests a possible application of signal analysis techniques in support of the diagnosis of laryngeal disorders.

1. INTRODUCTION

Most laryngeal pathologies can be successfully diagnosed by means of videostroboscopic images [1], but inaccurate assessments can result due to poor images, misinterpretation, or an undetected relevant aspect. Objective (i.e. computerised) analysis of acoustic and electroglottographic (EGG) signals, which play only a modest role in the differential diagnosis, can, on the other hand, provide important "warning signs" if perturbation measures fall outside their expected normal ranges. Speculations about the underlying pathologies are also possible: a small (acoustic) signal-to-noise ratio, for example, suggests the existence of a glottal chink, while a low EGG closed quotient commonly indicates reduced medial contact due e.g. to paralysis or atrophy of the vocal folds. Furthermore, an abnormality in the closing phase of the EGG can be an early indication of subglottal growths which may be difficult to detect by indirect laryngoscopy.

In a previous study [2], a method to estimate acoustic jitter in sustained /a/ vowels has been described. Waveforms of these vowels produced in "comfortable levels of pitch and loudness" often presented a stable and clean zero crossing pattern in either the up-going or down-going direction (Fig. 1). In the referenced study, jitter estimates based on these zero crossing patterns were compared with EGG-derived jitter of 15 patients suffering from various laryngeal pathologies, normalised absolute differences falling within 23.81% (average = 10.95%, standard deviation = 6.44%). High differences resulted for /i/ and /u/ vowels by using this method and others, based on a wavematching technique and variations of peak-picking and zero crossings.

In the study reported in the present paper, the comparative assessment of acoustic and electroglottographic measures of jitter in /a/ vowels was extended to 45 patients, the main outcome being a method to detect minor abnormalities in the movements of one arytenoid cartilage. These abnormalities can cause a difference in longitudinal tension between the left and right vocal folds leading to irregular vibrations.

2. METHODS

2.1 Recording Procedures and Subjects

Simultaneous acoustic and EGG recordings during the maximum sustained phonation of /a/ vowels were taken from the database described in [3], which also includes other acoustic and electroglottographic recordings, as well as aetiological, aerodynamic, and videostroboscopic images. The files are stored in CD-ROMs and integrated by a MS-Access database.

The patients sat in an isolated booth and the signals, recorded with a DAT equipment, were subsequently transferred into

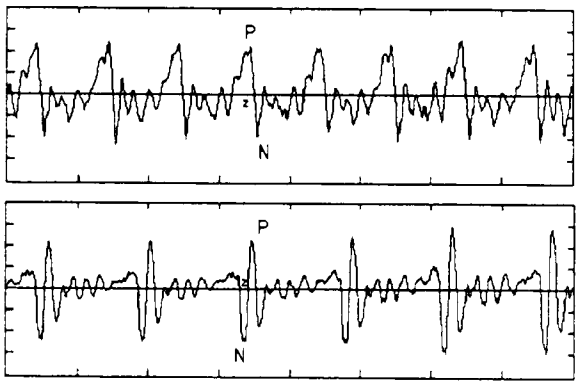

Figure 1:. Zero crossing patterns. Top: down-going (PZN) zero crossings; bottom: up-going (NZP) zero crossings.

~ On leave from Departamento de Física/ICEx/UFMG, CP 702, CEP 30161-970, Belo Horizonte/MG, Brazil.

† Supported by CNPq/Brazil (grant 200068/93-8).

".wav" files (22,050 samples per second, 16 bits per sample). Data from 52 cases (22 males, 30 females) were initially selected from 71 consecutive new patients attending the Voice Clinic of the Royal Infirmary of Edinburgh over a 6 month period. In the lack of an accepted objective criterion for EGG quality assessment, the selection was based on a visual inspection of the waveforms, which excluded recordings presenting noisy closing phases that would cause excessive errors in the automatic analysis.

2.1 Acoustic Jitter Estimation

Jitter was automatically measured from the acoustic signal by combining the robustness of a peak-picking method [4] and the accuracy of zero crossings. The F_0 detection algorithm searched for zero crossings in short intervals (1.25 ms) after and before the *significant peaks* "P" and "N" of each glottal cycle (Fig. 1).

Jitter was determined as described below, where the "10" subscript in the expressions is an allusion to a restriction which confined EGG and acoustic jitter estimates to the same range (10%), since values above that limit could not be reliably estimated from the acoustic signals due to excessive voicing-detection errors:

1. Candidate jitter values, $PF1_{10}()$, were initially obtained for up-going (NZP) and down-going (PZN) zero crossing patterns based on zero crossings before ($\bar{P}$) or after ($\tilde{P}$) positive peaks (and, similarly, $\bar{N}$ and $\tilde{N}$, for negative peaks):

$$PF1_{10}(PZN) = \tfrac{1}{2}[PF1_{10}(\bar{P}) + PF1_{10}(\tilde{N})] \quad (1a)$$

$$PF1_{10}(NZP) = \tfrac{1}{2}[PF1_{10}(\bar{N}) + PF1_{10}(\tilde{P})] \quad (1b)$$

where $PF1_{10}$ is defined for the different types of glottal cycle boundaries (i.e., $\bar{P}$, $\tilde{P}$, $\bar{N}$, or $\tilde{N}$) as:

$$PF1_{10}(\cdot) = \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \frac{|F_0(i+1) - F_0(i)|}{\frac{1}{2}\cdot[F_0(i+1) + F_0(i)]} \times 100 \; . \quad (1c)$$

In this expression, $F_0(i)$ is the instantaneous fundamental frequency, defined as the inverse of the elapsed time between the ith and (i-1)th glottal cycle, and N_{10} is the number of glottal cycles automatically detected for each utterance. Ideally, measures from $\bar{P}$ and $\tilde{N}$ (and, similarly, from $\bar{N}$ and $\tilde{P}$) would be the same.

2. The *ratio* parameter, intended to measure the deviation from this ideal situation, was calculated as:

ratio (PZN) =

$$\min[PF1_{10}(\bar{P}) / PF1_{10}(\tilde{N}),\ PF1_{10}(\bar{N}) / PF1_{10}(\tilde{P})] \quad (2a)$$

ratio (NZP) =

$$\min[PF1_{10}(\bar{N}) / PF1_{10}(\tilde{P}),\ PF1_{10}(\bar{P}) / PF1_{10}(\tilde{N})] \quad (2b)$$

3. Finally, the estimated jitter, $PF1_{10}()$, was the candidate value (Eq. 1) with the corresponding higher ratio (Eq 2), if ratio > 80%; otherwise, no value could be reliably estimated. A high ratio was normally observed either in the up-going or down-going direction, and a low ratio in both directions was associated with extremely breathy or harsh voices.

2.2 Electroglottographic Jitter Estimation

Initially, EGG signals were bandpass filtered (60-5000 Hz) with zero phase shift [5] to reduce the effects of baseline fluctuation [6]. Jitter, based on linearly interpolated zero crossings in the closing phase, was calculated with Eq. 1c, an upper 10% limit being also introduced, as justified before.

Acoustic and EGG jitter estimates were compared by means of normalised differences,

$$\Delta = \frac{PF1_{10}(\text{acoustic}) - PF1_{10}(\text{EGG})}{PF1_{10}(\text{EGG})} \times 100\% \; , \quad (3)$$

for the cases where an acoustic jitter estimate was available, i.e., ratio > 80%.

3. RESULTS

An overview of the pathologies affecting the 45 patients who provided a reliable acoustic jitter according to the 80% criterion is presented in Fig. 2. From the 71 original cases, 19 patients were excluded due to poor EGG recordings, and 7 patients due to unreliable acoustic jitter extraction.

The distribution of differences (Eq. 3) for these 45 patients is plotted in Fig. 3; the corresponding average value was 2.66%, the standard deviation 30.79%, and the mean absolute value 20.25%. In 30 cases (66.67%), EGG and acoustic jitter agreed to within ±20%, while Δ fell below -20% in 9 cases, and above +20% in 6

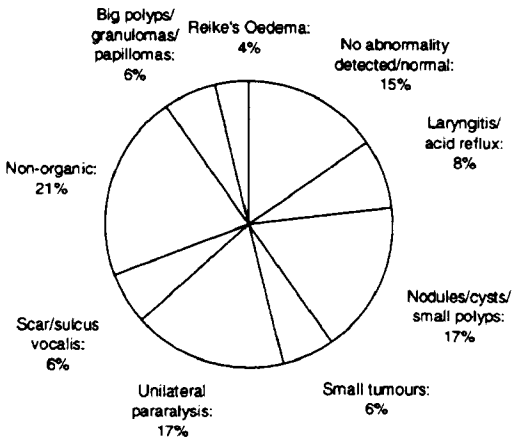

Figure 2: Distribution of pathologies (45 patients).

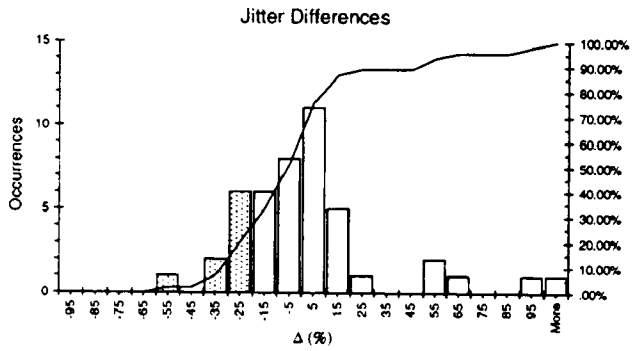

Figure 3: Distribution of differences (Eq. 3). The shaded bins indicate patients with possible abnormality in the cricoarytenoid joint.

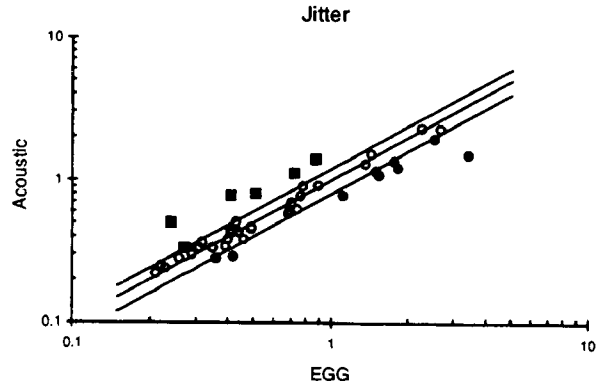

Figure 4: Scatter plotter of acoustic and EGG jitter. Squares indicate acoustic jitter more than 20% above EGG jitter and filled circles are acoustic values more than 20% below the corresponding EGG measures.

cases. The ±20% interval was adopted because it accommodated most differences observed in a previous study [2].

A scatter plot of the measures is presented in Fig. 4, where the continuous lines indicate the locus of (a) *equal-valued* measures (central line), (b) acoustic measures 20% *above* EGG measures (top line), and (c) acoustic measures 20% *below* EGG measures (bottom line). The squares and filled circles indicate the cases where acoustic measures fell above or below the 20% threshold, respectively. Considering that different signals were being analysed, there was a fair agreement over a wide range of jitter values. As mentioned before, this level of agreement was not achieved when acoustic and EGG jitter extraction methods were compared in /i/ and /u/ vowels.

In order to find possible explanations for the differences outside the ±20% interval, acoustic and EGG waveforms, plots of jitter time series, and videoendoscopic images of these 15 patients were inspected. The findings are described next.

4. DISCUSSION AND CONCLUSIONS

4.1 Differences above +20%

In one case in this group, EGG jitter was underestimated due to modal/falsetto breaks which passed undetected in the initial inspection, causing excessive unvoiced decisions in the EGG automatic analysis. In another case, the acoustic waveform did not present the expected "dominant" PZN or NZP zero crossing, and acoustic jitter was affected by noisy crossings, being therefore overestimated. The remaining 4 cases presented no common endoscopic feature, but the acoustic signals were characterised by shimmer and/or long term amplitude perturbations (tremor) which might have inflated jitter measurements.

A future study, including more objective measures and patients, will be carried out in order to obtain a better understanding of the causes of differences (Δ) above +20%.

4.2 Differences below −20%

The analysis of the videoendoscopic images revealed that 8 out of 9 patients in the group with Δ values below −20% had *different types of slight asymmetries in the movement of one arytenoid cartilage.* In regard to the remaining patient ("false positive"), acoustic jitter was underestimated due to excessive voicing detection errors. The videoendoscopic recordings of all 45 patients were re-examined and only one similar asymmetry ("false negative") was found, the corresponding Δ being -8.16%.

An inspection of the waveforms and associated jitter time series indicated that in the 8 cases with abnormal movements, EGG jitter increased due to the increase in EGG shimmer (possibly caused by vibrations of the arytenoids) with no corresponding changes in the acoustic signal. Among these subjects, 4 patients also had other laryngeal problems (1 patient with myopathy, 2 with polyps, and 1 suffering from intubation granuloma).

The lack of electromyographic (EMG) recordings is a serious limitation to the interpretation of the videoendoscopic images presenting the sort of minor unilateral paralysis observed in these patients. However, simultaneous EMG recordings from all muscles, including the left and right sides, can be difficult to obtain, the posterior cricoarytenoid being a particularly difficult muscle to reach [7]. The use of (low speed) VHS video recording is another difficulty for the analysis of the fast movements of the cricoarytenoid joint [8]. Further difficulties can come from apparent asymmetries in the posterior part of the larynx caused by image distortions due to an incorrect position (angle) of the endoscope. This artifact, though, did not affect any of the 8 mentioned cases. With these limitations in mind, some interpretation of the images will be given below.

One patient had an (anaesthetic) intubation that caused the dislocation of the left arytenoid, which did not move properly. Two patients had either a well compensated paralysis of the left

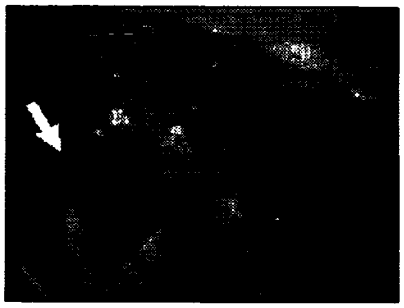

Figure 5: Possible abnormality in the right PCA (*left* side of the image), suggested by the position of the arytenoid (arrow).

inferior (recurrent) laryngeal nerve or a fixation of the left arytenoid. Images from two other cases suggested a unilateral paralysis of the superior laryngeal nerve (affecting the cricothyroid muscle); as explained in [9, pp. 218-221], the arytenoids were slightly rotated towards the paralysed side and the vocal fold in the affected side was at a lower level.

In the three remaining cases, the arytenoid cartilage tended to rotate forward, possibly due to the missing or reduced action of the posterior cricoarytenoid (PCA) muscle, which is antagonist to the pulling forces of the thyroarytenoid (TA) and the cricothyroid (CT) muscles. This sort of unilateral paralysis would allow the paralysed arytenoid to sag and would impair locking with the other arytenoid [10, p. 79].

The PCA muscles are the only laryngeal abductors, being especially active for the glottal opening during inspiration. The role of these muscles in the production of unvoiced sounds, and their high activity in the beginning of utterances has been studied in [11]. The absence of this initial activity that (in normal phonation) is slowly suppressed during the onset of voiced sounds, may be associated to the hard glottal attacks subjectively perceived in the acoustic recordings of these patients, who also had difficulties in alternating a low- and high-pitched vowel.

Hirano [12, p. 161] observed that the PCA "is usually inactive during phonation, with the exception of high tones in the modal register (...). This activity of PCA may be required in order to brace the arytenoid cartilage against the strong anterior pull of the CT." We have examined further recordings (not included in this study), where possible abnormalities in the PCA seem to be particularly harmful to singers. Videostroboscopic images showed asymmetrical longitudinal tension and different modes of vibration of the vocal folds in specific notes, which tend to become harsh in these conditions.

Minor asymmetries in the cricoarytenoid joints are possibly common in the general population [14], being caused e.g. [9] by viral neuropathy, trauma, arthritic fixation, or ageing [13]. A method has been described to detect these abnormalities based on normalised differences (Δ) between EGG and acoustic jitter below –20%. This non-invasive approach, which appears to be robust enough for clinical use, has shown the importance of integrating the various available clinical techniques.

5. REFERENCES

1. Hirano, M., and Bless, D.M. "Videostroboscopic examination of the larynx." Singular Publishing Group, San Diego, California, 1993.

2. Vieira, M.N, McInnes, F.R., and Jack, M.A, "Comparative assessment of electroglottographic and acoustic measures of jitter in pathological voices." *Submitted for publication.*

3. Vieira, M.N., McInnes, F., Jack, M., Maran, A., Watson, C., and Little, M. "Methodological aspects in a multimedia database of vocal fold pathologies." *Proc. Eurospeech'96*: 1867-1870, 1995.

4. Schäfer-Vincent, K. "Pitch period detection and chaining: Method and evaluation." *Phonetica,* 40: 177-202, 1983.

5. Kormylo, J.J. and Jain, V.K., "Two-pass recursive digital filter with zero phase shift." *IEEE Trans. Acoust. Speech Signal Process.* 22: 384-388, 1974.

6. Vieira, M. N., McInnes, F., Jack, M. "Analysis of the effects of EGG baseline fluctuations on the F_0 estimation in pathological voices." *J. Acoust. Soc. Amer.,* In press.

7. Fujita, M., Ludlow, C., Woodson, G. E, and Nauton, F.A. "A new surface electrode for recording from the posterior cricoarytenoid muscle." *Laryngoscope* 99: 316-320, 1989.

8. von Leden, H. and Moore, P. "The mechanics of the cricoarytenoid joint." *Arch. Otolaryngol.* 73: 541-550, 1961.

9. Luchsinger, R and Arnold, G. "Voice-Speech-Language, Clinical Communicology: Its Physiology and Pathology." Wadsworth Publishing Company, Beltmont, 1965.

10. Fink, R. B. and Demarest, R. "Laryngeal Biomechanics." Harvard University Press, 1978.

11. Hirose, H., and Gay, T. "The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study." *Phonetica* 25: 140-164, 1972.

12. Hirano, M. "The function of the intrinsic laryngeal muscles in singing." In: Stevens, K. and Hirano, M. (Eds.), *Vocal Fold Physiology,* pp. 155-170. University of Tokyo Press, 1981.

13. Kahane, J.C., "Age related changes in the human cricoarytenoid joint." In: Fujimura, O. (Ed.), *Vocal Physiology: Voice Production Mechanisms and Functions,* pp. 145-157. Raven Press, 1988.

14. Sellars, I. E. and Keen, E.N. "Anatomy and movements of the cricoarytenoid joint." *Laryngoscope* 88: 667-674, 1978.