ACOUSTIC TREMOR MEASUREMENT: COMPARING TWO SYSTEMS

M. A. E. Brückl, E. Ibragimova, S. Bögelein

Institute for Language and Communication, Technische Universität Berlin, Berlin, Germany
markus.brueckl@tu-berlin.de
elvira.ibragimova@campus.tu-berlin.de
silke.boegelein@campus.tu-berlin.de

Abstract: **A study is presented comparing two software systems that measure vocal tremor acoustically by analyzing sustained vowels. As measure for the comparison serves the criterion validity, here derived from the determination coefficients of simple linear regressions between the tremor measures and the synthetically given tremor values. For this purpose, the vowels to be analyzed were generated completely by acoustic synthesis. The two systems in comparison are a proprietary and widely, also clinically, used voice quality measurement tool and a self-developed algorithm that is based on autocorrelation of pitch and amplitude contours and implemented as a script of an open-source speech analysis program. The comparison's result is that the open-source software clearly achieves the more valid measurements.**
Keywords : **Vocal tremor, acoustic measurement, system comparison, open-source software**

I. INTRODUCTION

The acoustic measurement of vocal tremor bears a high potential to serve for early diagnosis of several, mostly neuro-degenerative diseases like Parkinson's (PD), Alzheimer's, multiple sclerosis, etc. Tremor often is defined as involuntary cyclic movement, or movement deviation, of the limbs. But, at least if it is caused by deficits of the central nervous system, it is most likely that speech production is affected too, since the production of speech involves the coordinated processing of about 1,400 motor commands per second. So, the more than 80 muscles of the vocal apparatus may all show tremor and thus vocal tremor may have many sources. But once the acoustic output is investigated, all of these organic modulation sources combine to only two types of tremor: subsonic quasi-cyclic modulations of the frequency and of the amplitude. And the acoustic signal may easily be captured.

In spite of the potential of auditive or acoustic vocal tremor assessment, its reliability and therewith its validity still provide great room for improvement. This may be a reason why e.g. simple perturbation measures are used in multi-feature PD detection systems [1, 2], whereas more specific tremor features are either not even evaluated to contribute to the system [1] or they are rather circuitously derived via frequency-domain techniques, but not directly within the time-domain [2], and are thus more error-prone.

Hence, the aim of this study is to compare two acoustic tremor measurement systems according to their criterion validity, that is here defined as goodness in measuring synthetically generated and thus known tremor.

II. METHODS

A. Acoustic synthesis of the test stimuli with known tremor properties in three steps

A completely synthetic sustained vowel is created by formant synthesis. (1) The glottal source signal (3s duration, 200Hz mean fundamental frequency ($\bar{F}_0$) is modelled according to [3] and then (2) filtered by a time-invariant 'female'-/a/-shaped filter function. This /a/-sound, which is perceived as rather natural, serves as the carrier for the frequency and amplitude modulations. (3) These modulations are done by re-synthesis according to the overlap-and-add method [4]. Both modulation types are modelled with a sinusoidal shape that is varied in frequency and amplitude, resulting in 4 synthesis arguments: the frequency tremor frequency (FTrF [Hz]), the amplitude tremor frequency (ATrF [Hz]), the relative frequency tremor intensity (FTrI [%]), and the relative amplitude tremor intensity (ATrI [%]). Each argument is varied in 4 equally spaced steps across each range of naturally occurring values. Additionally, both a frequency (decF) and an intensity decline (decA) are synthesized and varied in order to also simulate these naturally occurring effects. Thus, the synthesis of the modulations may be formulated as functions of time (t):

$$F_0M(t) = F_{0,s} + FTrI \cdot \bar{F}_0 \cdot sin(FTrF \cdot 2\pi \cdot t) - decF \cdot t \quad (1)$$

$$AM(t) = A_s + ATrI \cdot \bar{A} \cdot sin(ATrF \cdot 2\pi \cdot t) - decA \cdot t \quad (2)$$

where $F_{0,s}$ and A_s are the fundamental frequency resp. the amplitude at the sound's start that are depending on the sound's duration, on the means, and on the declines.

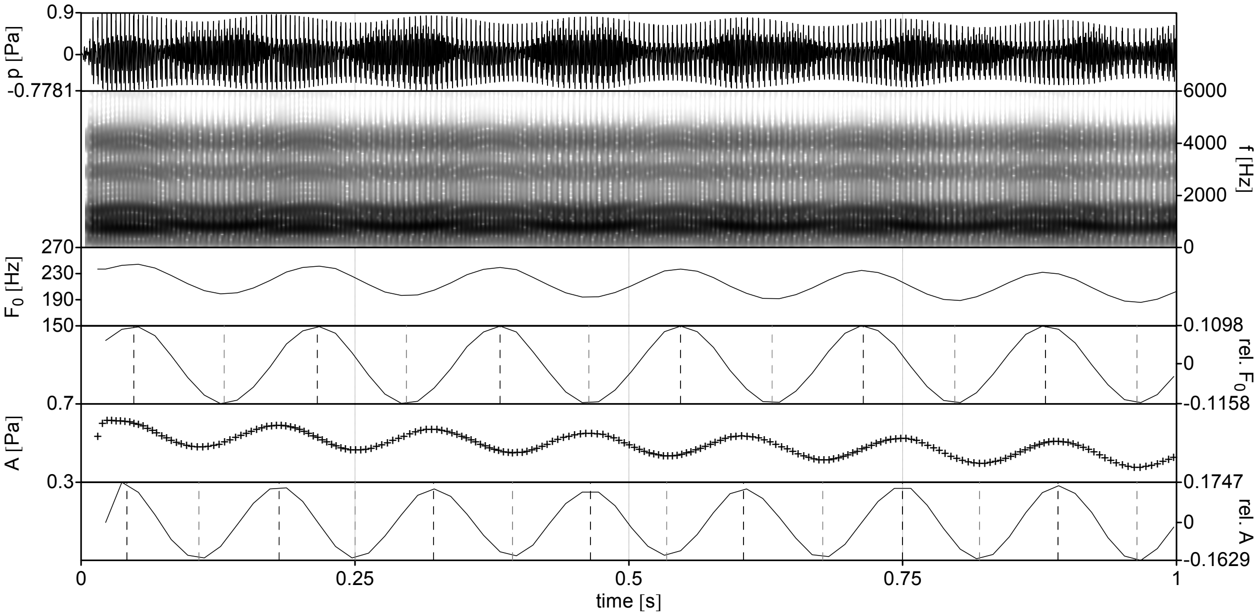

Figure 1: An exemplary synthesized sound and its tremor analysis by TREMOR.PRAAT*: The – from top to bottom – 1st subfigure shows an oscillogram of the first second of the synthesized sound with set tremor values of FTrF=6.0Hz, ATrF=7.0Hz, FTrI=11.5%, ATrI=15.5% as well as declines of decF=15Hz/s and decA=0.15Pa/s. The 2nd subfigure displays a short-time spectrogram of this sound. The contour in Subfigure 3 depicts* TREMOR.PRAAT*'s F_0-analysis. Subfigure 4 contains this F_0-contour, but de-declined and normalized. The short dashed vertical lines denote the times of minima (gray lines) and maxima (black lines) found by* TREMOR.PRAAT. *The 5th subfigure shows the sound's amplitudes per period, extracted by* PRAAT*'s To Amplitude. Subfigure 6 depicts the resampled, de-declined and normalized amplitude contour, again with found minima and maxima.*

A sound example is shown in Fig. 1. The sinusoidal shape and the decline of the amplitude envelope can be seen in particular from SubFig. 1. The frequency modulation may be recognized by the cyclic changes in the density of the glottal pulses in SubFig. 2.

In total $4^6 = 4{,}096$ test sounds result from a complete variation of the 6 synthesis arguments. All 3 synthesis steps as well as the arguments' variation are implemented as a PRAAT [5] script that is added to [6].

B. The tremor measurement systems

The two compared systems are (1) the Multi-Dimensional Voice Program (MDVP) [7] and (2) TREMOR.PRAAT, version 3.01 [6], a revised version of the algorithm presented in [8], including some newly developed tremor measures.

MDVP is a commonly known and widely used voice quality measurement tool. Its standard procedure extracts 4 tremor measures that should correspond to the above mentioned synthesis arguments (MDVP is proprietary software, thus computational details are not known): The frequency of the strongest low-frequency modulation of the fundamental frequency (Fftr [Hz]) or respectively of the amplitude (Fatr [Hz]), and the mean magnitude of the strongest low-frequency modulation of the fundamental frequency (FTRI [%]) or respectively of the amplitude (ATRI [%]).

TREMOR.PRAAT is open-source software and implemented as a PRAAT script. It extracts 14 tremor measures. 4 out of these 14 meet the definitions of the above named MDVP measures, i.e. they also correspond theoretically to the synthesis arguments and are named like them. TREMOR.PRAAT determines the tremor frequencies (FTrF and ATrF) by autocorrelating the F_0-contour, see SubFig. 3 of Fig. 1, and the amplitude contour, see its SubFig. 5. But before the contours get autocorrelated, the linear declines are removed by subtracting the linear regression estimates. Also, the amplitude contour must be resampled at a constant time step, since PRAAT's *To Amplitude* function extracts amplitudes per time-varying periods.

For the computation of the intensity indices (FTrI and ATrI), the contours are normalized, i.e. the deviations about the means ($\bar{F}_0$ or $\bar{A}$) are expressed relative to these means in the analyzed sound – just like in the MDVP:

$$rel.F_0(t) = \frac{F_0(t) - \bar{F}_0}{\bar{F}_0}\ ;\ rel.A(t) = \frac{A(t) - \bar{A}}{\bar{A}} \qquad (3)$$

This normalization is needed, since *tremor intensity* shall denote the magnitude of a cyclic deviation, and

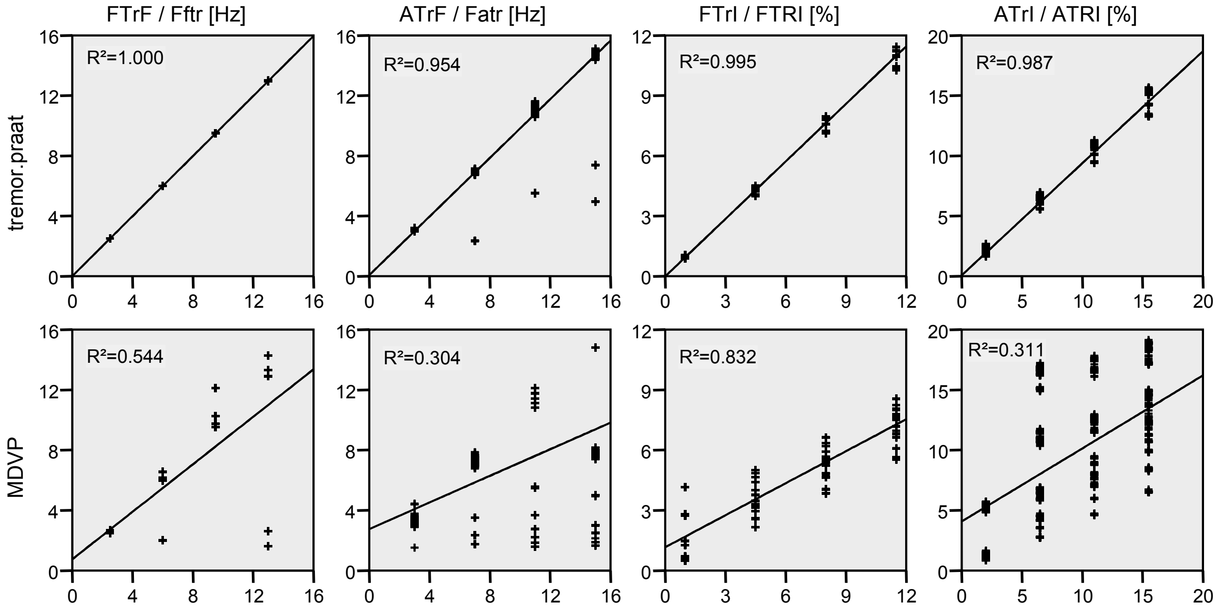

Figure 2: Scatterplots showing the measured values (ordinates) as a function of the values that were set by synthesis (abscissae). The lines are the linear regression models.

thus it should be expressed relative to its mean. The points in time at which this deviation magnitude is largest and that additionally fit to the already determined tremor frequency are found by PRAAT's function *To PointProcess (peaks)*. These steps are visualized in Sub-Fig. 4 and 6 of Fig. 1: The vertical lines mark the times of found extrema. The ordinates of each contour at these times are the searched tremor magnitudes (max, min). Finally, these magnitudes get averaged to the tremor intensity indices:

$$(F,A)TrI = \left(\frac{\sum_{i=1}^{m} |max_i|}{m} + \frac{\sum_{j=1}^{n} |min_j|}{n}\right) \div 2 \qquad (4)$$

where n and m denote the numbers of the found minima resp. maxima.

The default settings of the *search ranges for tremor frequencies* were expanded in both programs to 1.5Hz – 16 Hz. The *amplitude tremor octave cost* was raised to 0.2 in TREMOR.PRAAT in order to compensate for the unnaturally high cyclicality of the synthetically generated tremor contours that induces – together with the rather large analysis window and the sinusoidal shapes – sub-octave errors in determining ATrF, see Discussion.

C. Statistical methods

In order to assess the dependence of the 8 measured values on the values that are set by synthesis, 8 simple linear regressions are computed. Their determination coefficients (R^2) denote the proportion of variance in the measured values that can be explained by the set values' variance, thus they may serve as coefficients of validity of the measurement instrument. 99.99% confidence intervals (CIs) around these coefficients are calculated in order to indicate if the populations of corresponding coefficients differ from another.

III. RESULTS

The results of the regression analyses are shown in Fig. 2: MDVP fails to extract amplitude tremor measures in 513 cases and frequency tremor measures in 256 cases. Although TREMOR.PRAAT achieves to extract all measures from all sounds, its errors are highly significantly smaller, i.e. its measures are highly significantly more valid than those of the MDVP. In order to illustrate this significant superiority, Fig. 3 shows that the best estimates of R^2 do not fall within the CIs of corresponding measures of the other system, and that TREMOR.PRAAT's coefficients always denote higher validities.

TREMOR.PRAAT's measurement of FTrF is (nearly) totally valid: The regression line fits all data points and equals the coordinate system's angle bisector. Also, the other TREMOR.PRAAT measures can be considered excellent. In contrast the MDVP's extractions exhibit considerably more and greater measurement errors.

The MDVP is not built to be able to cope with naturally occurring declines, neither of the amplitudes nor of the frequency. In order to adjust for this, a further statistical analysis was executed that was reduced to the 4^4 = 256 sounds without any decline. But the highly significant differences between the two measurement systems remain – again with a confidence greater than 99.99%, just like in the analysis that comprises all 4.096 sounds.

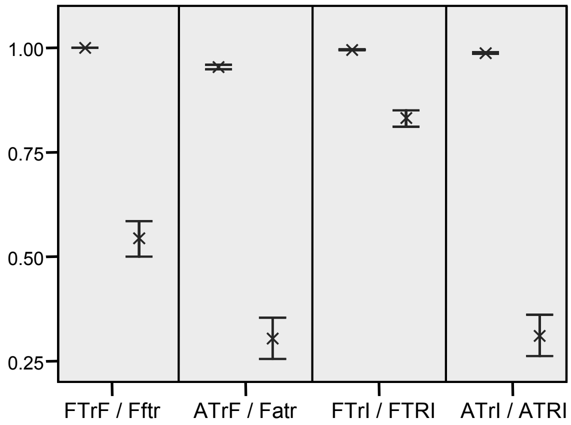

Figure 3: The best estimates (x) and the 99.99% CIs (double-T-bars) of the regressions' determination coefficients (R^2): TREMOR.PRAAT*'s measures are highly significantly more valid than those of the MDVP.*

IV. Discussion

All errors in TREMOR.PRAAT's measurements may be reduced by shortening the *analysis time step* (default value: 0.015s), at the cost of an exponentially increasing computational load.

TREMOR.PRAAT's tremor intensity measures (FTrI and ATrI) exhibit greater underestimations at greater synthetically set values. These errors are due to the combination of the sinusoidal shapes of the modulations with the averaging of these shapes within analysis windows: Sinusoids reach extreme values only punctually, whereas analysis windows mandatorily span a duration.

If ATrF gets extracted deficiently, then exactly one or two octaves too low, cp. Fig. 2. These octave errors result from correctly detecting sub-harmonics of the modulation frequencies that – again – are artificially induced by sampling the synthetically exactly sinusoidal contours at a rather low rate. Additionally to reducing these errors by shortening the *analysis time step*, they can be avoided by further raising the *tremor octave cost* argument. Apart from that, these errors will hardly occur when analyzing natural sounds, since natural tremor modulations are far less cyclic, wherefore a "rough" sampling seldom will construe sub-harmonics.

Errors in the MDVP's extractions seem to be far less systematic. Their sources must remain unrevealed, since the MDVP's algorithm is proprietary and thus unknown.

Besides, TREMOR.PRAAT still is developed to comprising more indices that in their totality are perceptually and biologically more valid for the concept of *tremor* than those alone that are already known and implemented: The newly developed indices FTrP and ATrP, for example, combine tremor frequency and intensity. As reported in [9] they seem to better picture the medical concept of *tremor severity* than the known intensity indices and thereby indicate PD – provided that the speakers' age and sex is considered. Furthermore, the concept of *cyclicality* is highly likely to contribute to a holistic concept of *tremor strength* or *severity*, just as well as to consider the fact that often there is not just one, the strongest, tremor frequency in a voice. Consequently, the most recent inventions in TREMOR.PRAAT [6] are indices that integrate tremors at multiple frequencies, whereat considering each cyclicality and intensity.

V. Conclusion

Although TREMOR.PRAAT is still under development, it has been shown that it is already far more valid in measuring vocal tremor than the standard program MDVP. Thus, it can only be advised to use TREMOR.PRAAT for acoustic tremor measurement. Furthermore, formerly gained results that were based on the MDVP's tremor measures are very likely to improve in precision and variety if they were re-measured with TREMOR.PRAAT. Also, the PD detection rates of the approaches described in [1] and [2] are likely to improve, if the measures of TREMOR.PRAAT were added to the feature sets.

References

[1] A. Tsanas et al., "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease", *IEEE Transactions on Biomedical Engineering*, vol. 59 no. 5, pp. 1264–1271, 2012.
[2] D. Hemmerling et al., "Automatic detection of Parkinson's disease based on modulated vowels", *INTERSPEECH–2016*, San Francisco, pp. 1190–1194, 2016.
[3] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels", *Journal of the Acoustical Society of America*, vol. 49, pp. 583–590, 1971.
[4] E. Moulines, F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Speech Communication*, vol. 9, pp. 453–467, 1990.
[5] P. Boersma, D. Weenink, *Praat: doing phonetics by computer (Version 6.0.29)* [Computer program], University of Amsterdam. http://www.praat.org/, accessed on 25 June 2017.
[6] M. A. E. Brückl, *tremor.praat (Version 3.01)* [Computer program], Technische Universität Berlin. http://brYkl.de/tremor3.01.zip, 2017.
[7] Kay Elemetrics Corp. / PENTAX Medical, *Multi-Dimensional Voice Program (MDVP), Model 5105 (Version 2.6.2)* [Computer program], 1993/2003.
[8] M. A. E. Brückl, "Vocal tremor measurement based on autocorrelation of contours", *INTERSPEECH–2012*, Portland (OR), 715–718, 2012.
[9] M. A. E. Brückl, A. Ghio, F. Viallet, "Measurement of tremor in the voices of speakers with Parkinson's disease", *1st International Conference on Natural Language and Speech Processing*, Algiers, 44–48, 2015.