ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Vocal Tremor Measurement Based on Autocorrelation of Contours

Markus Brückl

Fachgebiet Kommunikationswissenschaft, Technische Universität Berlin, Berlin, Germany

markus.brueckl@tu-berlin.de

Abstract

An algorithm to measure vocal tremor is presented, validated, and applied. The expected input is a sound file that captures a sustained phonation. The 6 output values are the frequency of frequency and amplitude tremor, intensity indices of frequency and amplitude tremor, and power indices of frequency and amplitude tremor. Basic principles of the algorithm are (1) autocorrelations of pitch and amplitude contours that are based on an autocorrelation of the input signal, (2) the correction for declination of (natural) contours as well as (3) a contour peak-picking and -averaging method for the determination of tremor intensities. The tremor power indices are new measures that weight tremor intensities by tremor frequency in order to receive bio- and psychologically more significant measures of tremor magnitude. The algorithm is implemented as a script of an open-source speech analysis program providing an most accurate pitch-detection (autocorrelation) method.

Index Terms: acoustic voice analysis, vocal tremor, vibrato, modulation

10.21437/Interspeech.2012-223

1. Introduction

Tremor (commonly shiver, tremble) is an unintentional muscular control deficit that results in cyclic movement deviations. It can be caused by a broad spectrum of phenomena like coldness, too much coffee, advancing age, and/or diseases that emerge preferentially in elder people like dementia, Alzheimer's or Parkinson's. Minor tremor is found in every muscular action of every human. Functionally speaking, all tremor causing phenomena can be seen as disturbances of or (minor) latencies in (the feedback of) the (neuronal) regulation system of a muscular process, e.g. the production of speech.

Vocal tremor is generally defined as an unintentional low-frequency modulation of the vocal fold vibration. If intentionally used in singing, such modulations are known as vibrato. But an acoustic speech signal may also show further "tremulous" components that are e.g. due to articulatorily motivated jaw movements. Thus, for a reliable measurement of vocal *tremor* in natural voices a vowel (e.g. /a/) phonation is to be preferred. This phonation should be sustained for several seconds in most comfortable pitch and loudness but as constant as possible.

However, unlike other tremors the acoustic representation of vocal tremor channels into two components: a frequency and an amplitude tremor. Speech (voice) production is a highly complex process and all its regulation disturbances/latencies are interweaved in both tremor types. That entails the fact that diagnosing vocal tremor does not allow to refer unambiguously to any underlying cause. But on the other hand this comprises also the power of vocal tremor analysis as an additional tool for the determination respectively diagnosis of a wide variety of phenomena and diseases, see also [1].

But despite the well known and broad spectrum of applicability of acoustical tremor measurement in speech research, biological, and psychological domains, only few systems are available (e.g. [2], [3], [4]) which extract vocal tremor from the acoustic signal. Therefore this paper aims to provide an open-source – and thus inspectable, discussable, adaptable, and free of charge – algorithm to automatically measuring vocal tremor.

2. The algorithm

The algorithm is implemented in the script language of the speech-processing program PRAAT [5]. This script can be found as the file named tremor.praat in the multimedia-folder attached to this paper.

2.1. Definitions

The technical tremor parameter definitions are adopted from the likely most commonly used tremor measuring instrument MDVP [3]. Accordingly the frequency tremor frequency (FTrF, in MDVP: Fftr) is the frequency of the strongest low-frequency modulation of the fundamental frequency (F_0), amplitude tremor frequency (ATrF, in MDVP: Fatr) is the frequency of the strongest low-frequency modulation of the amplitude (intensity).

The frequency tremor intensity index (FTrI, in MDVP: FTRI) is the magnitude of the strongest low-frequency modulation of F_0, the amplitude tremor intensity index (ATrI, in MDVP: ATRI) is the magnitude of the strongest low-frequency modulation of amplitude (intensity). By definition these magnitudes are expressed relative to the mean F_0 ($\overline{F_0}$), respectively the mean intensity ($\overline{A}$), in the analyzed sound; thus they are without physical unit and given out in %:

$$FTRI = FTrI = 100 \cdot \frac{absFTrI - \overline{F_0}}{\overline{F_0}} \quad (1)$$

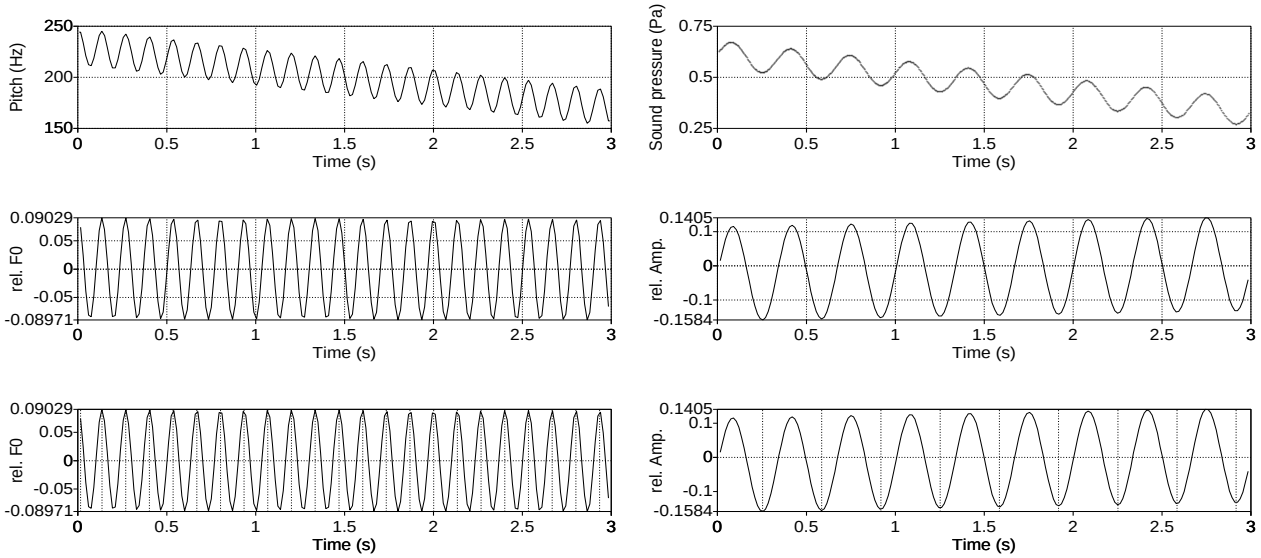
$$ATRI = ATrI = 100 \cdot \frac{absATrI - \overline{A}}{\overline{A}} \quad (2)$$

where $absFTrI$ and $absATrI$ denote absolute tremor intensities in physical measurement units (e.g. Hz and Pa).

Additionally to these four common definitions of measures of vocal tremor, here two new measures are introduced: the indices of tremor power (FTrP and ATrP). These measures result from weighting the intensity indices (FTrI and ATrI) with a factor depending on tremor frequencies (FTrF and ATrF). This factor is smaller for lower frequencies and therefore lower power indices emerge if the same tremor intensity is found at lower tremor rates.

$$FTrP = FTrI \cdot \frac{FTrF}{FTrF + 1} \quad (3)$$

$$ATrP = ATrI \cdot \frac{ATrF}{ATrF + 1} \quad (4)$$



(a) frequency tremor (b) amplitude tremor

Figure 1: *Steps of tremor analysis. Figure (a) shows the frequency and figure (b) the amplitude tremor analysis of a synthesized sinusoidal with FTrF=7.5Hz, AtrF=3Hz, FTrI=9%, ATrI=14%, decf=20Hz/s, deca=0.1Pa/s, amf=200Hz, and ama=0.5Pa.*

Power indices are thought to be biologically and psychologically more significant for the concept "tremor level" or "degree of tremor" than the known intensity indices.

2.2. General settings

In order to use the algorithm properly, a few general settings have to be made: In tremor.praat the F_0 range is currently optimized for female voices. To analyze male voices the minimal and the maximal pitch have to be adjusted. The analysis time step is set to facilitate fast but still sufficiently accurate measurement. You may wish to reduce it to receive more accurate measurements.

The three further settings pertain to the question what's to be considered a "tremor": the minimal and the maximal tremor frequency determine the frequency range of modulations that are considered. The tremor threshold ranges (potentially) from 0 to 1 and denotes the lower limit of the autocorrelation coefficient (of the contours) to assure sufficient modulation strength (to consider something a low-frequency *modulation* as opposed to irregular fluctuations). So if the (highest) autocorrelation coefficient that can be detected in the contour signal is smaller than this value (here 0.15) it is assumed that there is no tremor and therefore no tremor frequency nor intensity nor power. Raising this value will generally make missing result values become more likely. This autocorrelation threshold is currently set to a value considered to be minimal to ensure the concept of modulation.

2.3. Basic steps of the algorithm

The basic principles common to both frequency and amplitude tremor analysis are subsequently outlined:

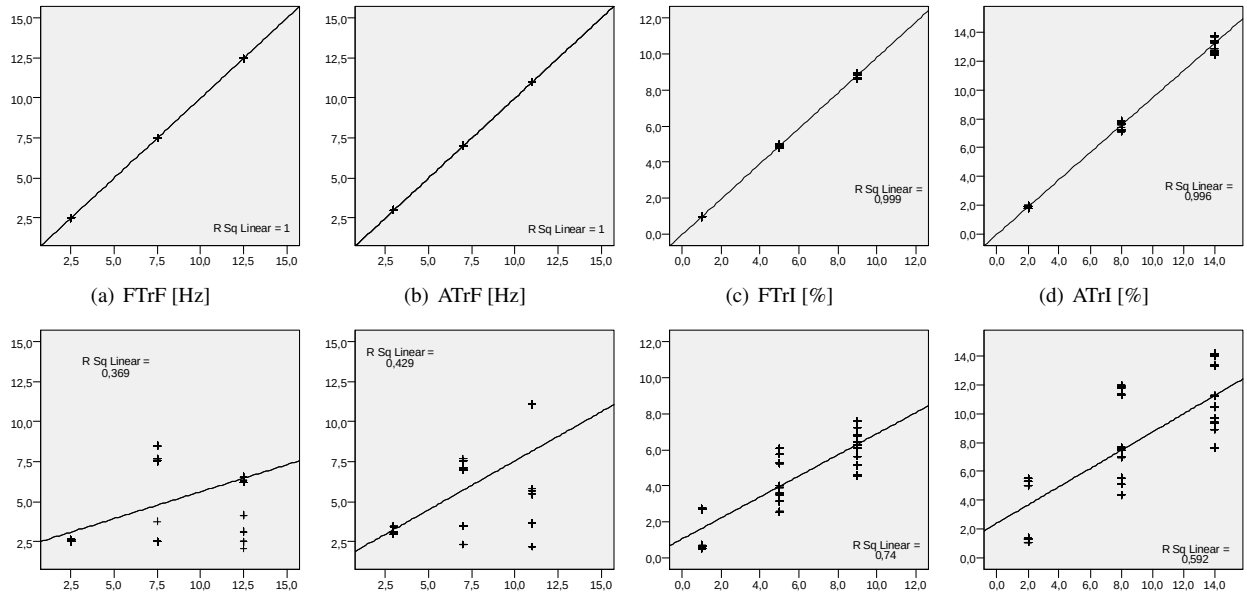
The first step is the calculation of the frequency and amplitude contours of the input signal. This is done with the (autocorrelation, see [6]) function "To Pitch (cc)" and the function "To AmplitudeTier (period)". Thus single amplitude values (per pitch period) do not represent the maximum value within this period but the integral. Examples of resulting contours are shown in the graphs right on top of Figures 1 (a) and (b). The initial pitch-synchronous calculation of the amplitude per wave cycle avoids gros artificial "modulations" (that would occur if intensities were directly read at a constant analysis time step and would depend on the relation of the analysis time step to F_0). For autocorrelation purposes these pitch synchronous amplitudes must be re-sampled at a constant rate. This is accomplished by using the same analysis time step and midpoints as in the pitch object and by determining the mean of the amplitude curve (the integral) in time-step-width around these midpoints.

Step two is the removal of linear components (declinations, overall de- or increases) in the contours which is realized by the function "Subtract linear fit". After the normalization of the contours by their means according to Equations 1 and 2 to meet the definitions, the analysis reaches the phase shown in the middle graphs of Figure 1.

The third step comprises the autocorrelation (see [6]) of amplitude and frequency contours to determine the tremor frequency. The analysis window here extends to the whole sound duration to capture the one most likely respectively strongest tremor frequency inherent in the whole contour. This frequency is the searched tremor frequency (FTrF resp. ATrF). Once the frequencies of the strongest modulations are known, (relative) minima and maxima of the the contours are picked at times depending on these frequencies. This step is done by means of the function "To PointProcess (peaks)" and visualized in the bottom graphs of Figure 1: The dotted vertical lines mark the times of found extrema (maxima in sub-figure (a) and minima in sub-figure (b)). The searched intensity values are the assigned ordinates. The new tremor intensity indices FTrI and ATrI are operationally derived from these intensity values by means of the following averaging method:

$$(F,A)TrI = \left(\frac{\sum_{i=1}^{m} |max_i|}{m} + \frac{\sum_{j=1}^{n} |min_j|}{n} \right) \div 2 \quad (5)$$

where n and m denote the number of minima resp. maxima.



(a) FTrF [Hz] (b) ATrF [Hz] (c) FTrI [%] (d) ATrI [%]

(e) MDVP Fftr [Hz] (f) MDVP Fatr [Hz] (g) MDVP FTRI [%] (h) MDVP ATRI [%]

Figure 2: *Scatter plots with regression lines to visualize the validation of tremor measurements on 729 synthesized sinusoidals: measured values [Hz] (ordinates) as (linear) functions of the synthetically given values [Hz] (abscissae). Sub-figures (a)-(d) show data obtained by applying the new algorithm, sub-figures (e)-(h) depict MDVP data. All shown regressions are highly significant ($p < 0,5\%$).*

3. Validation

The above described algorithm shall now be validated by measuring tremor in sounds with known properties. Therefore sounds were generated according to the following formula (using the script gen_dectrem.praat):

$$s(t) = [ba - deca \cdot t + atri \cdot ama \cdot \sin(2\pi t \cdot atrf)] \cdot \sin\left[2\pi t \cdot \frac{bf - decf \cdot t}{2} + \frac{ftri \cdot amf}{ftrf} \cdot \sin(2\pi t \cdot ftrf)\right] \quad (6)$$

where

t:	*a certain point in time [s]*
s(t):	*(intensity [Pa] of the) signal at time t*
ama:	*mean intensity [Pa]*
amf:	*mean fundamental frequency [Hz]*
ba:	*intensity [Pa] at t=0*
bf:	*fundamental frequency [Hz] at t=0*
deca:	*(linear) declination of intensity [Pa/s]*
decf:	*(linear) declination of fundamental frequency [Hz/s]*
ftrf:	*(synth.) frequency tremor frequency [Hz]*
atrf:	*(synth.) amplitude tremor frequency [Hz]*
ftri:	*(synth.) frequency tremor intensity index*
atri:	*(synth.) amplitude tremor intensity index*

Hence gen_dectrem.praat synthesizes sinusoidal sounds with linear frequency and intensity declinations as well as (low frequency) sinusoidal modulations of fundamental frequency and of intensity. gen_dectrem.praat also realizes the adjustment of six parameters (ftrf, ftri, atrf, atri, decf and deca) of Formula 6, each in three steps. Thus $3^6 = 729$ sounds for testing the validity are resulting. All of them have the same mean F_0 (= 200 Hz) and the same mean intensity (= 0.5 Pa). Since the first steps of declines (deca and decf) are set to zero, 81 of these sounds do not exhibit any declination. The three synthesis steps of tremor frequencies were chosen to roughly covering the whole tremor frequency range. The two tremor intensity parameters (ftri, atri) are varied so that the smallest value denotes tremor that could be found in healthy speakers, the middle one represents clearly audible tremor and the upper severe tremor (intensities) – as far as such normative values are known.

The results of measurements on these input signals are visualized in Figure 2 as scatterplots showing the measured values as functions of the synthetically given ones. An examination of the results on the new measures reveals that tremor frequency is detected exactly – each time (see Figures 2 (a) and (b)). The intensity indices (see Figures 2 (c) and (d)) correlate very highly with the synthetically given values, but are always smaller. This is caused by averaging intensities within time windows in combination with the sinusoidal form of the synthesized modulations: There is only one local maximum (resp. minimum) point and the mean in its surrounding window is mandatorily lower (higher). The tremor intensity (and power) measurements become more accurate if the analysis time step is reduced.

In contrast the MDVP parameters (Figures 2 (e)-(h)) show considerably greater measurement errors, especially in the frequency extraction. The measured tremor frequency and tremor intensity values are generally too low. The error in MDVP intensity indices is considerably reduced if only sounds without declination are analyzed (FTRI: R^2=0.948; ATRI: R^2=0.967), but the accuracy of the new measures is not reached and, moreover, the tremor frequency extraction stays poor (Fftr: R^2=0.621; Fatr: R^2=0.322). It can be concluded that declinations impair MDVP tremor intensity measurement.

Another validation of the new measures via correlation with MDVP-measures yields nearly the same results as shown in Figures 2 (e)-(h), since the values of the new measures are nearly identical to those given by synthesis (FTrF∼Fftr: R^2=0.369; ATrF∼Fatr: R^2=0.429; FTrI∼FTRI: R^2=0.746; ATrI∼ATRI: R^2=0.601).

4. An application: indicating speaker age

The new algorithm as well as the MDVP tremor procedures are also applied to natural voices from 88 women comprising all adult ages (chronological age (CA): AM=50.42a SD=17.64a). They sustained the vowel /a/ as long as possible. Three 2.2s lasting segments were extracted: a start segment including the vocal onset, a quasi-stationary (middle) segment and the end including the offset. These were rated by approximately 30 Listeners in order to estimate the speakers' age. These values [a] were corrected for listener bias and averaged to perceptive ages (PA) per vowel-segment. For details on these data please refer to [7].

The results of correlation analyses between the tremor measures and these age scales are reported in Table 1:

Table 1: *Pearson's r denoting (linear) correlations between chronological age (CA) or perceptive age (PA) and tremor intensity and power indices. Highly significant ($p < 1\%$) values are set in boldface, significant ($p < 5\%$) values are in italics.*

	/a/ start		**q.-s. /a/**		**/a/ end**	
	CA	**PA**	**CA**	**PA**	**CA**	**PA**
FTRI	**.511**	**.476**	**.410**	**.377**	*.250*	**.393**
ln(FTRI)	**.484**	**.419**	**.461**	**.315**	**.308**	**.455**
ATRI	-.055	.011	.137	.209	*.289*	*.305*
ln(ATRI)	.053	.049	.197	.147	.176	.189
FTrI	**.340**	**.359**	*.216*	**.450**	.103	**.290**
ln(FTrI)	**.422**	**.468**	**.389**	**.622**	**.290**	**.513**
ATrI	**.544**	**.504**	**.364**	**.505**	.119	**.438**
ln(ATrI)	**.588**	**.509**	**.526**	**.554**	*.228*	**.517**
FTrP	**.326**	**.359**	*.236*	**.475**	.135	**.325**
ln(FTrP)	**.404**	**.454**	**.403**	**.660**	**.315**	**.544**
ATrP	**.550**	**.530**	**.424**	**.558**	*.236*	**.492**
ln(ATrP)	**.573**	**.520**	**.577**	**.601**	**.318**	**.555**

MDVP FTRI and its natural logarithm correlate quite good with both age scales. But MDVP amplitude tremor measures do hardly show any relation to age at all. (See lines 1-4 of Table 1)

In comparison to MDVP measures the new frequency tremor (intensity and power) measures generally correlate slightly less with *age*. (Please note that this is not in contradiction to the new measures being more valid in measuring *vocal tremor*: Probably the MDVP tremor measures are more sensitive to other constructs like perturbations or additive noise that are co-varying with age as well.) However the new amplitude tremor measures indicate age better than all frequency tremor measures. The fact that amplitude tremor measures correlate better than frequency measures is also consistent to the finding of amplitude perturbations performing better as indicators of chronological and perceptive age than frequency perturbations, see [7].

If only the new measures are focused, it can be seen that the logarithmic measures yield higher correlation coefficients. Furthermore the newly invented power indices correlate rather higher to age than the known intensity indices. The most prominent exception is the logarithm of the amplitude tremor intensity index (ln(ATrI)) extracted from start segments, which alone explains 34,5% (r=0,588) of the chronological age variance. The highest correlation (r=0,660; R^2=0,435) however is found between perceptive age and the logarithm of the frequency tremor power index (ln(FtrP)) measured in quasi-stationary parts.

5. Conclusions

An algorithm to measure vocal tremor has been presented. It is considerably more valid for measuring vocal tremor than the compared MDVP algorithm, especially in evaluating amplitude tremor and tremor frequencies. The new measures also serve as very good indicators of female speakers' age extracted from sustained vowel input.

The notion of the need to consider pitch and intensity declinations in tremor measurement proved to be awarding as well as the invention of tremor power indices. Both innovations can be considered as steps towards reliably measuring glottal tremor in connected speech.

Future work will focus on a variety of topics: First, considering not only the strongest modulations seems relevant, since it is obvious that there can be more than one modulation (frequency) in natural voices that are caused by/due to possibly different sources and that these additional modulations contribute to the perception of more "shakyness". Second, the applied linear model of a declination might be not optimal for modeling an underlying non-periodic alteration of contours – maybe a quadratic or semi-elliptic model will prove more adequate. Third, since the logarithmic measures proved special (biological, age) relevance, an alteration of the algorithm to measure pitch and amplitude in Mel and Bark rather than in Hertz and Pascal is to be testet. Finally, the tremulous sinusoidals generated for the validation sound reasonably natural, although they exhibit no high-frequency components nor noise nor irregularity. Thus, it seems rewarding to incorporate a tremor generator in (the voice modules of) speech synthesizers in order to make them sound more natural.

6. Acknowledgements

I would like to thank Paul Boersma and David Weenink for inventing and constantly developing PRAAT, and the TU Berlin (especially Walter Sendlmeier) as well as the German Research Foundation (DFG) for benefits that made the presented work possible.

7. References

[1] Gillivan-Murphy, P. and Miller, N., "Voice tremor: what we know and what we do not know", Current Opinion in Otolaryngology & Head & Neck Surgery, 19(3), 155-159, 2011.

[2] Winholtz, W.S. and Ramig, L., "Vocal Tremor Analysis With the Vocal Demodulator", Journal of Speech and Hearing Research, 35, 562–573, 1992.

[3] Kay Elemetrics Corp., "Multi-Dimensional Voice Program (MDVP), Model 5105" (Version 2.6.2), [Computer program], 1993/2003.

[4] Cnockaert, L., Grenez, F. and Schoentgen, J., "Fundamental Frequency Estimation and Vocal Tremor Analysis by means of Morlet Wavelet Transforms", Proc. of the IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, 393–396, 2005.

[5] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer" (Version 5.3.04), [Computer program], University of Amsterdam. Online: http://www.praat.org/, accessed on 19 Feb 2012.

[6] Boersma, P., "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", Proc. of the Institute of Phonetic Sciences, Amsterdam, 17:97–110, 1993.

[7] Brückl, M., "Altersbedingte Veränderungen der Stimme und Sprechweise von Frauen", W. Sendlmeier [Ed], Mündliche Kommunikation, Vol. 7, Logos Verlag, Berlin, 2011.