PLANTILLA OFICIAL PARA LA PRESENTACIÓN DE TRABAJOS
II CONVENCIÓN CIENTÍFICA INTERNACIONAL
"II CCI UCLV 2019"

DEL 23 AL 30 DE JUNIO DEL 2019.
CAYOS DE VILLA CLARA. CUBA.

II International Conference on Information Processing "CIPI 2019"

El jitter como perturbación de la periodicidad con mayor efecto en algoritmos de detección de período fundamental

Jitter as the main affecting factor in the performance of PDAs

Reinier Rodríguez Guillén[1,2], Carlos A. Ferrer Riesgo[2]

1-Reinier Rodríguez Guillén. Fábrica de Calderas "Jesús Menéndez", Cuba.

reinierrg@uclv.cu

2- Carlos A. Ferrer Riesgo. Universidad Central "Marta Abreu" de Las Villas, Cuba.

cferrer@uclv.edu.cu

Resumen:

La medición de jitter es una tarea frecuente en el análisis de voces patológicas y esta recae en la determinación previa del pulso glotal. El sistema software Praat se ha convertido en una herramienta ampliamente usada para obtener los límites del pulso y la medida de jitter. Sin embargo, resultados provistos por Praat pueden ser erróneos en su implementación si son ignorados procedimientos usados a la hora de obtener los límites del pulso. Este trabajo se direcciona sobre la influencia de algunas configuraciones del Praat a la hora de obtener esos marcadores. Fueron conducidas simulaciones para evaluar los efectos de diferentes tipos de perturbaciones periódicas (*jitter, shimmer*, ruido y sus combinaciones) y sus niveles en el desempeño resultante.

Palabras Clave: Praat, PDAs, perturbaciones, voces patológicas.

Abstract:

Jitter measurement is a frequent task in the acoustic analysis of pathological voices, and it relies in the previous determination of glottal pulse boundaries. The Praat software system has become a widely used tool to obtain both the pulse boundaries and the jitter measures. However, results provided by Praat can be misleading if the internal implementations of the procedures used to obtain the pulse boundaries are ignored. This paper addresses the influence of some of these settings in the ability of Praat to obtain these markers. Simulations are conducted to evaluate the effects of

different types of periodicity perturbations (jitter, shimmer, noise, and their combination) and their levels on the resulting performance.
Keywords: *jitter, shimmer, noise, Praat, PDAs, pathological voice.*

1. Introduction

Speech is arguably the most natural and effective means of human communication. We are all familiar with the use of speech, both in face-to-face conversations and in telephony. The usefulness of speech to humans is clear and therefore devices and services that employ speech in some form or another are always going to be important in society in general as well as in business and commerce. The fact that almost everyone uses spoken language all the time makes it both easy and difficult to describe the nature of it. It is easy because everyone can relate to the phenomena that occur in spoken language. But the ubiquitousness of spoken language can make it difficult to identify important aspects of it. A more careful examination reveals that some of those units are periodic and others have noise-like qualities (Gudnason, 2014; Naylor, 2014).
A pitch detector is an essential component in a variety of speech processing systems. Besides providing valuable insights into the nature of the excitation source for speech production, the pitch' contour of an utterance is useful for recognizing speakers, for speech instruction to the hearing impaired, and is required in almost all speech analysis-synthesis (vocoder) system how Praat (Lawrence R.Rabiner, 1976).

Because of the importance of pitch detection, a wide variety of algorithms for pitch detection have been proposed in the speech processing literature (Denis Jouvet, 2017; Hong Su, 2016; M Kiran Reddy, 2017; RaviShankar Prasad, 2015; verdiyev, 2015; Xu-Kui Yang, 2016; Xueliang Zhang 2016; Yoav Medan, 1991).

Small and apparently random perturbation of time period, amplitudes and shapes of consecutive periods in voiced speech waveform, respectively called shimmer, jitter and the CP (Complexity Perturbation), are the phenomena which exist in all normal human speech signal. The glottal pressure waves producing voiced speech are quasi- periodic in nature. The waveforms of consecutive glottal pulses vary randomly in terms of period, amplitude and complexity, though by a small extent. The reasons for this quasi-periodicity lie in the way the vocal cords operate. Traditionally the vocal cord oscillations were believed to be alike the vibration of a pair of rigid reeds. However, it has been seen that the voiced speech is not completely periodic, it is quasi-periodic. This means that if we examine closely two consecutive periods they are not exactly alike. These random differences (jitter, in time period; shimmer, in amplitude; and CP in complexity) though by a tiny amount are good enough to provide a feeling of naturalness.

Acoustic jitter and shimmer are measures of the cycle-to-cycle variations of fundamental frequency and amplitude, respectively, which have long been used for the description of pathological voice quality in a number of studies, while a few studies deal with emotional classification of speech using these features. Jitter and shimmer are commonly measured for long sustained vowels, and values of jitter and shimmer above a certain threshold value are said to be related to pathological voices, which are perceived by humans as breathy, rough or hoarse voices. Are reported that significant differences can occur in jitter and shimmer measurements between different speaking styles, especially in shimmer measurement. Jitter is a measure of vocal stability and for normal voices, the jitter value for normal voices are less than1% (Asoke Kumar Datta, 2019).

For all explain the objective of this paper is the analysis the jitter influence in pitch detection algorithm.

2. Methodology

In this paper we used synthetic signals with known levels of jitter, shimmer, and noise. These perturbations were evaluated in different types of PDAs.

2.1 Synthetic Signals

The synthesis follows the common All-Pole formant synthesizer scheme (Klatt, 1980; Klatt & Klatt, 1990; C Manfredi, d'Aniello, Bruscaglioni, & Ismaelli, 2000) according to the source-filter theory (Fant, 1981). The expression for the synthesized signal *s(n)* is given in (1), where *h(n)* is the vocal tract's impulse response, *g'(n)* is the derivative of the glottal flow, *e(n)* is an additive noise term, and a_i and t_i are the amplitude and time of occurrence, respectively, of the i^{th} pulse. The * sign denotes the convolution operation:

$$s(n) = h(n) * \sum_i \left(a_i \delta(n - t_i) * g'(n)\right) + e(n) \quad (1)$$

The *h(n)* was synthesized using a five formant model for the vowel 'a', with the same parameters used in (C. Ferrer, González, Hernández-Díaz, Torres, & del Toro, 2009; C. Ferrer, Hernández-Díaz, & González, 2007; C. Ferrer & Torres, 2010; C. A. Ferrer, González, & Hernández-Díaz, 2006; Parsa & Jamieson, 1999; Yoav Medan, 1991), and truncated to 300 samples (more than 99% of the total energy included) before performing the convolution. The glottal excitation has been refined here, compared to (C. Ferrer, et al., 2009; C. Ferrer, et al., 2007; C. Ferrer & Torres, 2010; C. A. Ferrer, et al., 2006; Medan, Yair, &

Chazan, 1991; Parsa & Jamieson, 1999), such that the simple impulse train was replaced by an actual *g(t)* waveform according to the polynomial model B of Rosenberg (Rosenberg, 1971) for the glottal flow:

$$g(t)=\begin{cases}3\left(\frac{t}{T_u}\right)^2-2\left(\frac{t}{T_u}\right)^3, & 0<t<T_u\\ 1-\left(\frac{t-T_u}{T_d}\right)^2, & T_u<t<T_u+T_d\\ 0, & T_u+T_d<t<T_0\end{cases} \quad (2)$$

The first two intervals in the formula represent the open (non zero) phase, a first one corresponding to the opening segment and a second one to the closing segment. The graphical depiction of the waveform and involved parameters are shown in Figure 1.

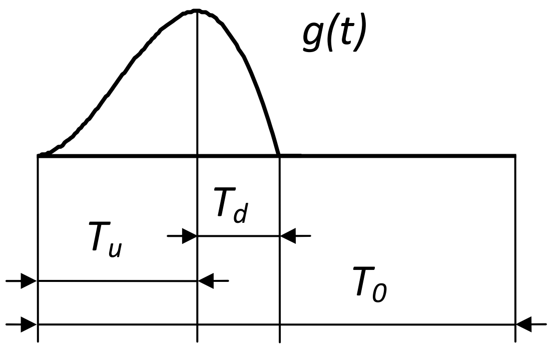

Figure 1. Glottal flow cycle waveform *g(t)* in the Rosenberg's type B polynomial model.

The synthesis was performed using an opening time of T_u=.33T_0 and a closing time of T_d=.09T_0, the values producing the most naturally-sounding voices in (Rosenberg, 1971). The sum of the two non-zero sections yield an Open Quotient (OQ) of 0.42. This synthesis procedure follows a Piecewise Warping (PW) model (C. A. Ferrer, Torres, González, Calvo, & Castillo, 2015) for glottal pulse shape change in the presence of jitter, which is shown to be the least stressing for PDAs (C. A. Ferrer, et al., 2015).

Sampling frequency Fs was set to 22050 Hz, mean fundamental frequency F0= 150 Hz, with signal duration of 2 seconds, yielding an average of 300 pulses per signal. We introduced varying levels of each single perturbation, creating three sets of signals (jittered only, shimmered only, or noisy only). An additional 4th set is created by simultaneously mixing all the perturbations as their levels are increased.

2.1.1 Noisy Signals

Additive white Gaussian noise is added to an otherwise periodic signal, so that the Signal to Noise Ratio (SNR) attains a given value. The SNR chosen for the seven levels of perturbations were (C. Manfredi, 2012; Gerratt, 2000; H. D. Huici, 2016; K. Hosokawa, 2014; O. Amir, 2009; R. M. Roark, 2006; Titze, 1995) dB.

2.1.2 Jittered Signals

The position of the i^{th} glottal pulse is obtained as:

$$t_i = t_{i-1} + T0 + u(i) \quad (3)$$

where *u(i)* is a random perturbation, whose value is uniformly distributed in the interval $\pm u_{max}$. Seven levels of perturbation were evaluated, corresponding to values of u_{max} of [3.4, 6.8, 10.2, 13.6, 17, 20.4, 23.8]*T0/100. The maximum level of perturbation follows the comment in (Medan, et al., 1991) that consecutive pulses almost never differ in more than 25% of their lengths. Since the total length of the signal consists of the sum of all these random separations, the mapping of 300 pulses into 2-second duration signals can only be approximate in the presence of jitter.

2.1.3 Shimmered Signals

Pulse amplitudes were modified by a factor:

$$a_i = 1 + k(i) \quad (4)$$

where *k(i)* is a random real value uniformly distributed in the interval ±kmax. It is mentioned in (Titze, 1995) that shimmer is usually within 50%, so the seven values of kmax were chosen as twice the values of umax to keep the ratio towards the mentioned (Medan, et al., 1991) limit value for jitter (25%).

2.1.4 Composite perturbations

In these signals, the perturbation parameters (umax, kmax, and SNR) were varied simultaneously, in ascending order of perturbation, to obtain also seven levels of composite perturbations. The values used are the same as in the synthesis of the three previously described individual perturbations. In this way, the nth level of composite perturbation will be synthesized with the nth level of jitter, the nth level of shimmer, and the nth level of noise.

2.2 Algorithm PDAs considered

The three basic PDAs considered are the Superresolution method in (Medan, et al., 1991), and the Cross Correlation and Peak Picking methods provided by Praat. A brief description is provided in order to explain the different configurations used.

2.2.1 Superresolution method

The Superresolution (SR) cycle to cycle waveform-matching PDA searches for the minimum weighted difference between two contiguous, non-overlapping segments in a given search range for the speech segment. The expression to minimize is:

$$J(\tau)=\frac{\sum_{n=1}^{\tau}\left|x(t_i+n)-a_{\tau,t_i}x(t_i+\tau+n)\right|^2}{\sum_{n=1}^{\tau}x(t_i+n)^2} \quad (5)$$

A relevant feature of this cost function is the fact that the segment sizes (summation limit) vary in equal magnitude as the displacement (time difference between both segments of *x*). There is also an optimal value of the amplitude variability *a*, which is a function of the particular pulse and of the value of displacement being evaluated.

The range of values of τ where the minimum of the function *J* is to be searched for is initially broad and covering the whole possible range of T0 accepted in the particular study. After an initial T0 is determined, the range for the search of the next position (τ) is narrowed to the vicinity of the previous T0 (T_E). It is mentioned in (Medan, et al., 1991) that the variability from one pulse to the other is generally within ±10%, and almost never exceeds ±25%, so this latter limit can be chosen, and is actually used in (C. Ferrer, et al., 2009; C. Ferrer, et al., 2007; C. Ferrer & Torres, 2010; C. A. Ferrer, et al., 2006; C. A. Ferrer, et al., 2015).

This SR method is included as a reference for comparison with Praat's cycle-to-cycle PDAs since it has shown (C. Ferrer & Torres, 2010; C. A. Ferrer, et al., 2015) superior performance compared to the least mean squares approach in (Milenkovic, 1987) , and the latter's modification in (Titze & Liang, 1993).

2.2.2 Praat's Internal PDAs: Peaks and Cross Correlation

Praat provides several implementations of both (short-term and cycle-to-cycle) types of PDAs. The functioning of Praat's cycle-to cycle PDAs depends on the previous result of a short-term PDA, providing the frame-based estimate of T0 (T_E) over which the search for the individual pulse duration is performed in the cycle to cycle PDA.

The cycle-to-cycle PDAs implemented in Praat are Peak Picking (PP) and Cross Correlation (CC).

Praat PP PDA departs from the most relevant maximum (or minimum) at the center of the voiced segment, and searches to the left and to the right for similar points in the T_E±20% vicinity. It is provided for compatibility with other commercial systems, but its use is not encouraged.

Praat CC, being a waveform-matching PDA, is more robust to noise than the previous (Boersma, 2009), evaluating the pulse as a whole and not a single point in the search for the periodicity. The cost function in the Praat CC PDA can be written as:

$$J(\tau)=\frac{\sum_{n=1}^{T_E}\left|x(t_i+n)x(t_i+\tau+n)\right|^2}{\sqrt{\sum_{n=1}^{T_E}\left|x(t_i+n)\right|^2}\sqrt{\sum_{n=1}^{T_E}\left|x(t_i+\tau+n)\right|^2}} \quad (6)$$

Where important differences with (5) are that this is a function to maximize, and the displacement is no longer equated to the window size, with the latter being now constant and equal to T_E. The search range (set of values of τ) is restricted to ±20% of T_E as in the PP PDA, being a constant in the Praat environment.

As we have mentioned, discarding highly irregular segments/frames as unvoiced produces a tendency to underestimate jitter measures. There are several configuration parameters in Praat's short term PDAs, which are explained on the software's help system, which can influence the V/U decision. When considering a frame in isolation from its neighbors, a 'Voicing Threshold' parameter is used, so that the frame is considered Voiced if the maximum value of its autocorrelation is above that threshold. The information of neighboring frames is included by in Praat in a post-processing stage that smoothes the detected contour. A "Voiced to Unvoiced cost" parameter is included that makes less likely to transition to unvoiced frames. While most published research adheres to the default values, we illustrate the influence of configuration in the results obtained by modifying these two parameters.

Disregarding the classification of a frame as voiced in the V/U decision, Praat CC will discard any detected pulse within that given frame if the maximum value of *J* in (6) falls below .3.

2.3 Considered Measures of PDAs performance

The selection of the measures of performance to be used is not trivial. When evaluating factors affecting the calculation of jitter (i.e. length contour T(n) variability) the real issue should be the assessment of the similarity of the contour detected by the PDA, Td(n), and a reference one, Tr(n), assumed to be flawless.

As with the case of an intra-contour variability like α , a similarly expressed inter-contour variability was proposed in (C. Ferrer & Torres, 2010), and denoted as β, following the idea in (Parsa & Jamieson, 1999):

$$\beta = \frac{1}{N}\sum_{n=1}^{N}\frac{|T_d(n) - T_r(n)|}{T_r(n)} * 100 \tag{7}$$

Measures of inter-contour variability are the standard comparison procedure in the case of short-term PDAs (either in the form of absolute values like β (C. Ferrer & Torres, 2010; Medan, et al., 1991; Parsa & Jamieson, 1999) or as RMS values (Bagshaw, Hiller, & Jack, 1993; Lawrence R.Rabiner, 1976; Shahnaz, Zhu, & Ahmad, 2005) , where *n* is a frame index in both *T(n)* contours (either reference or detected), and there is a fixed relation between its value and the time of occurrence of the n^{th} frame in both contours. In Cycle-to-Cycle PDAs, where *n* is a pulse index, this synchronization does not exist, and most of the time both contours are of different lengths.

In this paper we adopt this DTW procedure in order to overcome the limitations of α comparisons. However, given our stated objective of evaluating the relevance of the exclusion of highly irregular frames and even individual pulses in Praat for its jitter measures, we paid special care to measuring the amount of missing pulses.

3. Results and Discussion

In this section four types of perturbations in synthetic signals are given.

3.1 Only Noise

The results for the signals where noise was the only introduced perturbation are shown in Table 1. For space reasons, the 8 measures were paired side-by-side, with the pairs chosen by similitude of meanings: intra and inter contour variability measures α and β (the ones expressed in %), PI and PD as the Global Constraints violations from the DTW, and SL and SR as the Local Constraints corrections in the DTW. The remaining two measures (GE & CS) constitute the other pair. Non-integer values have been shown with up to two decimal digits, and cells which are actually zero valued are shown empty, to distinguish them from the ones where the 0.00 value was reached by rounding. These two conventions are kept throughout the reports for the remaining types of perturbation. The rows corresponding to the pair of measures SL and SR are not shown for this perturbation type since they were zero in all cases. Among the actually expected results is the weakness of peak-picking alternatives when facing noise (the only ones showing Gross Errors and Pitch Deletions). All the measures also show the logical worsening of all the PDA's configurations as the noise increases, departing from practically zero values in all instances for the 22 dB SNR case. The exception to the latter point is the value of CS, whose minimum value (quite constant until the higher levels of perturbation) is 2 pulses.

Table 1: **Performance Measures for each PDA configuration facing the seven levels of Noise.**

		PDA	Level of Perturbation								PDA	Level of Pertutbation						
			1	2	3	4	5	6	7			1	2	3	4	5	6	7
Aligned Contour Measures	α_c	CC	0.00	0.03	0.05	0.10	0.20	0.28	0.40	β_c	CC	0.00	0.03	0.05	0.10	0.20	0.32	0.49
		PP	0.00	0.01	0.07	0.21	0.68	1.22	2.18		PP	0.00	0.01	0.08	0.22	0.76	1.40	2.64
		SR			0.00	0.01	0.16	0.27	0.39		SR			0.00	0.01	0.17	0.33	0.50
Gross Errors Measures	GE	CC								CS	CC	2	2	2	2	2	2.05	3.90
		PP				0.02	3.57	18.93	52.24		PP	2	2	2	2	2.02	2.18	3.06
		SR									SR	2	2	2	2	2	2.07	2.29
DTW outcomes	PI	CC								PD	CC							
		PP									PP						0.01	0.36
		SR									SR							

3.2 Only Jitter

The results for the signals containing only jitter are shown in Table 2. In this case, the eight measures are shown since all produced some non-zero values instances.

A visible difference compared to the only noise signals, is the outcome for the CS measure. Here the shrinkage of the contours for the first level of perturbation are dissimilar for all PDAs, and about two pulses shorter (CS could be rounded to four pulses in all cases) than the one in the noisy subset. The main issue for the shortening is the internal storage used in the implementation of the synthesis, where the signals with the seven levels of perturbations were stored as rows in a single matrix. Since in the presence of random jitter the total length of each signal is not exactly 2 seconds, the seven jittered signals within the particular run were truncated to the length of the shortest. Adopting other alternative for a common length would have eventually forced the detectors to work on empty spaces. For this reason, on each of the 100 runs, six out of the seven signals fed to the PDAs would be slightly shorter than the one corresponding to the 300 pulses. Another visible outcome from the first level of CS values is the larger shrinkage of SR-based variants. In this regard, we must recall that by using (5), the SR method needs to work with signal samples which are located T_E+2*SrRg beyond the current pitch mark, while Praat-based alternatives work up to T_E+SrRg. This is the cause for SR variants to stop searching for the next pulse earlier than its Praats counterparts, and missing eventually some pulse by doing so. The contour produced by the default configuration in Praat (CC) is around 35% (~107 pulses) shorter than the original (~300) in the worst level of perturbation.

Table 2: Performance Measures for each PDA configuration facing the seven levels of jitter.

		PDA	Level of Perturbation (mean jitter in %)								PDA	Level of Perturbation (mean jitter in %)						
			1.24	2.38	3.50	4.66	5.79	7.01	8.14			1.24	2.38	3.50	4.66	5.79	7.01	8.14
Aligned Contour Measures (%)	α_c-α_r	CC	-.01	.85	1.93	2.01	1.4	.71	.36	β_c	CC	.02	6.59	12.42	16.94	20.34	23.41	28.26
		PP	0	.36	.74	.84	.57	-.07	-.70		PP	0	4.53	9.31	14.59	17.40	21.22	24.87
		SR	0	.04	.08	.04	-.03	-.08	-.11		SR	0	.09	.16	.16	.18	.21	.23
Gross Errors Measures	*GE*	CC		19.28	48.19	64.07	78.30	93.60	97.62	*CS*	CC	3.83	17.63	55.13	76.95	84.59	88.21	106.9
		PP		8.89	2.94	38.84	61.6	95.54	119.79		PP	3.83	9.13	15.87	23.62	26.94	31.55	35.99
		SR									SR	4.42	4.32	4.23	3.94	4.32	4.16	4.28
DTW outcomes (Pulse and Shift errors)	*PI*	CC								*PD*	CC		7.86	36.1	55.57	61.54	64.2	83.27
		PP							.06		PP		3.17	7.35	13.47	15.75	19	21.94
		SR									SR							
	SL	CC		.01		.2	.58	.84	.83	*SR*	CC		5.44	13.71	16.49	17	17.91	17.28
		PP				.11	.36	.87	1.41		PP		1.96	4.38	6.33	7.2	8.91	1.77
		SR									SR							

A common phenomenon is perceivable in for all PDAs as the level of jitter increases, in which there is a tendency to an increased overestimation of jitter values in the lower levels, then a reduction in this overestimation beginning at the middle levels, reaching the point of underestimating α_r in the highest levels. A similar knee-like progression in the measurement of jitter was already found for Praat and other systems in (DeJonckere et al., 2011; Claudia Manfredi et al., 2012). While the larger part of this marked tendency to underestimate jitter as the perturbation increases could be linked to the removal of the most distorted frames from analysis by the short-term stage of the PDAs, there seems to be also some sensitivity in the cycle-to-cycle stage, even when the latter is of a much smaller magnitude.

The results shown also confirm how misleading can be the use of αc-αr as an indicator of a PDA performance, in spite of it spread of use.

3.3 Only Shimmer

The results for shimmer are shown in Table 3. In this case, the rows corresponding to the pairs of measures PI and PD, as well as SL and SR, are not shown since they were zero in all cases. There was no deterioration in terms of contour shrinkage (CS) for any PDA, at any level, except for the already explained (in the noise section) suppression of 2 pulses for all variants.

The configurations run within Praat, the PP variants to be the only ones to report GEs and at the same time outperform the CC-based configurations in terms of β. However, the proper nature of the PP allows to report a peak corresponding to the previous pulse damping oscillations, as long as this pulse is larger enough compared to the current one. These are necessarily GEs, and occur less than 20 times in the ~30000 pulse pairs. For all other pairs, the PP reports exactly the correct position, yielding very good values on average. The CC configurations run within Praat seem to report some markers shifted, in small magnitudes not reaching the GE magnitude, more frequently.

Table 3: Results for each PDA configuration facing the seven levels of shimmer.

		PDA	Level of Perturbation								PDA	Level of Pertutbation						
			1	2	3	4	5	6	7			1	2	3	4	5	6	7
Aligned Contour Measures (%)	α_c	CC		0.00	0.00	0.02	0.04	0.04	0.04	β_c	CC		0.00	0.00	0.02	0.04	0.04	0.04
		PP			0.00	0.00	0.01	0.02	0.01		PP			0.00	0.00	0.01	0.02	0.01
		SR									SR							
Gross Errors Measures	GE	CC								CS	CC	2	2	2	2	2	2	2
		PP			0.01	0.03	0.11	0.18	0.12		PP	2	2	2	2	2	2	2
		SR									SR	2	2	2	2	2	2	2

3.4 Composite perturbations

The results for the composite perturbations are shown in Table 4. It can be generally perceived that the relative performances of the PDAs among them are reminiscent to the ones obtained for the case of jitter perturbation, in this case worsened, which reinforces the notion of jitter as the main cause of PDA failure. There is for instance, the same knee-like overestimation of α_c in all cases, not present in the shimmered or noisy signals.

The deterioration in performance is extremely marked for Praat's default configuration (CC), where average CS rises in the worst case from 106.7 pulses/run (35% of the pulses) to 293.4 pulses/run (97.8% of the pulses undetected). An in-depth check revealed that 15 out of the 100 runs in the most extreme level of composite perturbations yielded absolutely no pulse marks. Praat's failure to detect pitch marks (at least on its default configuration) in extremely distorted signals is likely producing the consequent failure to report intra-contour variability measures described in (Claudia Manfredi, et al., 2012), when the jitter introduced exceeded 23%. The results obtained here show that this effect can be manifest for significantly lower levels of mean jitter (i.e. 8.14) if it is combined with noise and shimmer, as expected in pathological voices.

Table 4: **Results for each PDA configuration facing the seven levels of composite perturbations.**

		PDA	Level of Perturbation 1	2	3	4	5	6	7		PDA	Level of Perturbation 1	2	3	4	5	6	7
Aligned Contour Measures (%)	α_c-α_r	CC	-.01	.87	1.76	2.41	2.69	4.58	-4.74	β_c	CC	.04	6.69	12.69	19.05	25.83	49.09	34.72
		PP	.00	.40	.86	.88	.50	.08	.01		PP	.00	4.74	9.99	14.66	17.64	21.91	28.12
		SR	.00	.06	.11	.06	-.01	-.08	-.15		SR	.00	.10	.18	.21	.35	.75	3.60
Gross Errors Measures	*GE*	CC		19.64	49.51	63.04	61.89	28.77	2.62	*CS*	CC	3.83	18.53	60.49	101.00	159.50	244.03	293.39
		PP		9.65	23.06	41.7	73.01	12.76	165.90		PP	3.83	9.69	17.02	23.88	26.82	33.83	41.65
		SR					.04	2.34	14.83		SR	4.43	4.32	4.25	3.96	4.35	4.24	4.44
DTW outcomes (Pulse and Shift errors)	*PI*	CC								*PD*	CC		8.70	41.06	78.69	133.84	207.60	128.75
		PP						.04	3.09		PP		3.6	8.29	14.21	16.11	19.59	28.25
		SR									SR							
	SL	CC		.01		.12	.18	.05		*SR*	CC		5.47	13.95	16.17	14.49	6.66	.73
		PP				.23	.68	1.06	.91		PP		2.08	4.61	6	7.03	1.7	12.5
		SR						.26	1.66		SR						.27	1.75

4. Conclusions

The results with synthetic signals have shown that previous concerns regarding the influence of the discarded frames in the measurement of jitter were founded. The effects increase as the signals get more distorted. The main factor contributing to the reduction of the contour to be analyzed is the classification of frames as unvoiced by the short-term PDA. Although there is some amount of configuration available within the program, it seems reasonable that for jitter measurement, which is usually performed on sustained vowels, a 100% voiced TE estimate is fetched to Praat's cycle-to-cycle estimator. First, using a V/U decision when the signal was completely uttered as voiced tends to discard the most irregular frames, and as such to underestimate the level of perturbation present. This is important both in the clinical setting (values would be altered) as well as in the research setting, where Praat's internal cycle-to-cycle approach can wrongfully "accused" of poor performance. Secondly, letting the Thrs (in the CC-WM-PDA) suppress individual pulses would also leave Praat in disadvantage when dealing with synthetic signals where there are no missing pulses. Care must be taken when considering measures obtained from the point processes reported by Praat. The removal of Unvoiced segments occurs mostly in highly perturbed regions, and measured jitter in the remaining segments will certainly be lower. The removal of U segments is not the only factor affecting the performance: Praat's CC PDA discards pulses found when crosscorr. falls below .3.

5. References

Asoke Kumar Datta, R. S., Kaushik Banerjee and Dipak Ghosh. (2019). Acoustical Analysis of the Tanpura S. a. C. Technology (Ed.)

Bagshaw, P. C., Hiller, S., & Jack, M. A. (1993). *Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching.* Paper presented at the Third European Conference on Speech Communication and Technology.

Boersma, P. (2009). Should jitter be measured by peak picking or by waveform matching. *Folia Phoniatrica et Logopaedica, 61*(5), 305-308.

C. Manfredi, A. G., J. Schoentgen, S. Fraj, L. Bocchi, and P. H. Dejonckere. (2012). "Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Biomed. Signal Process. Control, 7.*

DeJonckere, P., Schoentgen, J., Giordano, A., Fraj, S., Bocchi, L., & Manfredi, C. (2011). Validity of jitter measures in non-quasi-periodic voices. Part I: perceptual and computer performances in cycle pattern recognition. *Logopedics Phoniatrics Vocology, 36*(2), 70-77.

Denis Jouvet, Y. L. (2017). *Performance Analysis of Several Pitch Detection Algorithms on Simulated and Real Noisy Speech Data*. Paper presented at the 25th European Signal Processing Conference, Kos, Greece.

Fant, G. (1981). The source filter concept in voice production. *STL-QPSR, 1*(1981), 21-37.

Ferrer, C., González, E., Hernández-Díaz, M. E., Torres, D., & del Toro, A. (2009). Removing the influence of shimmer in the calculation of harmonics-to-noise ratios using ensemble-averages in voice signals. *EURASIP Journal on Advances in Signal Processing, 2009*(1), 784379.

Ferrer, C., Hernández-Díaz, M. E., & González, E. (2007). *Using waveform matching techniques in the measurement of shimmer in voiced signals.* Paper presented at the Eighth Annual Conference of the International Speech Communication Association.

Ferrer, C., & Torres, D. (2010). Using dynamic time warping of T0 contours in the evaluation of cycle-to-cycle pitch detection algorithms. *Pattern Recognition Letters, 31*(6), 517-522.

Ferrer, C. A., González, E., & Hernández-Díaz, M. E. (2006). *Evaluation of time and frequency domain-based methods for the estimation of harmonics-to-noise-ratios in voice signals.* Paper presented at the Iberoamerican Congress on Pattern Recognition.

Ferrer, C. A., Torres, D., González, E., Calvo, J. R., & Castillo, E. (2015). *Effect of different jitter-induced glottal pulse shape changes in periodicity perturbation measures.* Paper presented at the Sixteenth Annual Conference of the International Speech Communication Association.

Gerratt, J. K. a. B. R. (2000). "Measuring Vocal Quality". *in Voice quality measurement, R. D. Kent and M. J. Ball, Eds. San Diego: Singular*.

Gudnason, J. (2014). Speech Production Modeling and Analysis. *Signal Processing*.

H. D. Huici, H. A. K., H. Martens, G. Van Nuffelen, and M. S. De Bodt. (2016). H. D. Huici, H. A. Kairuz, H. Martens, G. Van Nuffelen, and M. S. De Bodt. *H. D. Huici, H. A. Kairuz, H. Martens, G. Van Nuffelen, and M. S. De Bodt, 27*.

Hong Su, H. Z., Xueliang Zhang, Guanglai Gao. (2016). CONVOLUTIONAL NEURAL NETWORK FOR ROBUST PITCH DETERMINATION. *IEEE transactions on signal processing*.

K. Hosokawa, M. O., M. Hashimoto, and H. Inohara. (2014). "Statistical analysis of the reliability of acoustic and electroglottographic perturbation parameters for the detection of vocal roughness,". *J. Voice,, 28*.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *the Journal of the Acoustical Society of America, 67*(3), 971-995.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *the Journal of the Acoustical Society of America, 87*(2), 820-857.

Lawrence R.Rabiner, M. J. C., Aaron E. Rosenberg, Carol A.Mcgonegal. (1976). A comparative performance study of several pitch detection algorithms. *IEEE transactions on acoustics, speech and signal processing, 4*.

M Kiran Reddy, a. K. S. R. (2017). Robust pitch extraction method for HMM-based speech synthesis system. *IEEE Signal Processing Letters*.

Manfredi, C., d'Aniello, M., Bruscaglioni, P., & Ismaelli, A. (2000). A comparative analysis of fundamental frequency estimation methods with application to pathological voices. *Medical engineering & physics, 22*(2), 135-147.

Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., & Dejonckere, P. H. (2012). Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Biomedical signal processing and control, 7*(4), 409-416.

Medan, Y., Yair, E., & Chazan, D. (1991). Super resolution pitch determination of speech signals. *IEEE transactions on signal processing, 39*(1), 40-48.

Milenkovic, P. (1987). Least mean square measures of voice perturbation. *Journal of Speech, Language, and Hearing Research, 30*(4), 529-538.

Naylor, P. A. (2014). Introduction to Speech Processing. *Signal Processing*

O. Amir, M. W., and N. Amir. (2009). "A clinical comparison between two acoustic analysis softwares: MDVP and Praat". *Biomed. Signal Process. Control, 4*.

Parsa, V., & Jamieson, D. G. (1999). A comparison of high precision Fo extraction algorithms for sustained vowels. *Journal of Speech, Language, and Hearing Research, 42*(1), 112-126.

R. M. Roark. (2006). Frequency and Voice: Perspectives in the Time Domain. *J. Voice,, 20*(3).

RaviShankar Prasad, B. Y. (2015). Robust pitch estimation in noisy speech using ZTW and group delay function. *Speech and Vision Laboratory*.

Rosenberg, A. E. (1971). Effect of glottal pulse shape on the quality of natural vowels. *the Journal of the Acoustical Society of America, 49*(2B), 583-590.

Shahnaz, C., Zhu, W.-P., & Ahmad, M. O. (2005). *Robust pitch estimation at very low SNR exploiting time and frequency domain cues.* Paper presented at the Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.

Titze, I. R. (1995). *Workshop on acoustic voice analysis: Summary statement*: National Center for Voice and Speech.

Titze, I. R., & Liang, H. (1993). Comparison of Fo extraction methods for high-precision voice perturbation measurements. *Journal of Speech, Language, and Hearing Research, 36*(6), 1120-1133.

verdiyev, L. i. S. a. a. Y. g. I. (2015). A Comparative Analysis of Pitch Detection Methods Under the Influence of Different Noise Conditions. *The Voice Foundation*.

Xu-Kui Yang, L. H., Dan Qu and Wei-Qiang Zhang. (2016). Voice activity detection algorithm based on long-term pitch information. *EURASIP Journal on Audio, Speech and Music processing*.

Xueliang Zhang , H. Z., Shuai Nie, Guanglai Gao and Wenju Liu. (2016). A Pairwise Algorithm Using the Deep Stacking Network for Speech Separation and Pitch Estimation. *IEEE transactions on audio, speech and language processing*.

Yoav Medan, E. Y., and Dan Chazan. (1991). Super Resolution Pitch Determination of Speech Signals. *IEEE transactions on signal processing, 39*.