Commentary

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2009;61:305–308
DOI: 10.1159/000245159

Published online: October 10, 2009

Should Jitter Be Measured by Peak Picking or by Waveform Matching?

Paul Boersma

University of Amsterdam, Amsterdam, The Netherlands

In their article 'Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat', Maryn et al. [1] make a comparison between the jitter measurements in Praat and the Multi-Dimensional Voice Program (MDVP), and conclude that the two programs give *different* results. However, the readers of this journal might like to know as well which of the two programs gives the *best* result. After all, jitter is defined (according to Deliyski's [2] MDVP manual) as the 'period-to-period variability of the pitch period', a definition that suggests that speech sounds possess an underlying 'true' jitter that analysis programs could aim to discover.

As for which of the two programs provides the better jitter measurements, the authors only give indirect clues. On page 225 they acknowledge (following Boersma [3]) that the difference between Praat and MDVP is to be ascribed to the methods with which the programs try to determine the time locations of the glottal pulses: Praat's standard method is 'waveform matching', and that of MDVP is 'peak picking'. As for the quality of the two methods, the authors cite (on page 218) Titze and Liang [4] for finding that the waveform-matching method outperforms the peak-picking method for signals with a jitter below 6% (above 6%, both methods are poor). From this, the reader can indirectly infer that Praat's method is to be preferred over that of MDVP, but no further explanation is given. The present paper aims at providing the information lacking in the article by Maryn et al. by explaining the exact cause of the difference so that the reader can make up his or her own mind. I will discuss, then, the circumstances under which the two methods yield identical or different results.

Consider first the sound in figure 1. This waveform represents a computer-generated [a], created from a perfectly sampled pulse train with a frequency of 117 Hz, filtered with formants at 820, 1,300, 2,300 and 6 higher frequencies. This sound is meant to be representative of what patients are asked to produce in clinical jitter measurement procedures. The short vertical dashed lines indicate the time locations of the underlying pulse train.

The tick marks along the bottom of figure 1 indicate the 'glottal pulses' as measured by the waveform-matching method. This technique tries to find out at what time distance two consecutive waveshapes look maximally similar. The tick marks along the top of figure 1 indicate the glottal pulses as measured by the peak-picking method, which looks for time locations where the waveform is at its maximum. In the case of the perfectly periodic sound of figure 1, the two methods give identical results. We can see this because the dotted lines that go up from the tick marks at the bottom exactly touch the tick marks at the top. Also, the figure illustrates that both methods correctly find that all periods, measured as the time distances between consecutive tick marks, are 0.008547 s.

Things change when an amount of jitter is applied to the underlying pulse train. Figure 2 shows a sound that is identical to the one in figure 1, except that the underlying pulse train has an average 'local jitter' of 1%. This means that two consecutive underlying periods are on average different by 1%. For instance, the first underlying period (the time distance between the first and second dashed

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7762/09/0615–0305$26.00/0

Accessible online at:
www.karger.com/fpl

Paul Boersma
Phonetic Sciences, University of Amsterdam
Spuistraat 210
1012VT Amsterdam (The Netherlands)
Tel. +31 20 525 2385, Fax +31 20 525 3021, E-Mail paul.boersma@uva.nl

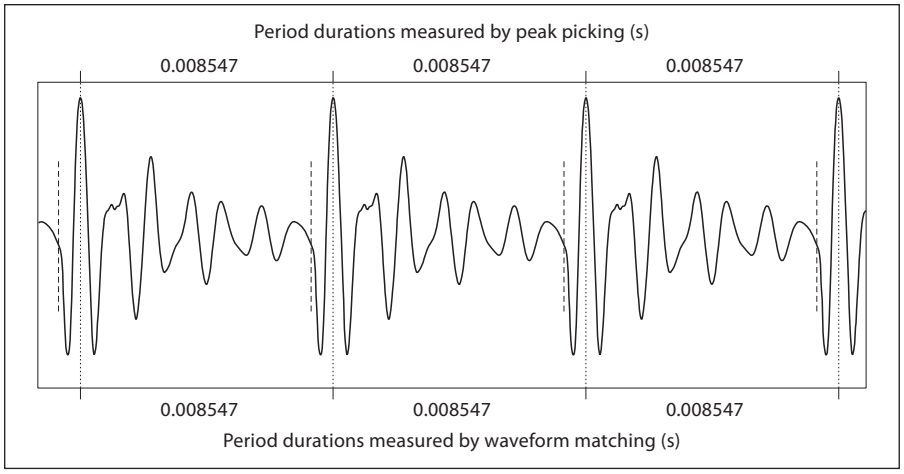

Fig. 1. A perfectly periodic sound.

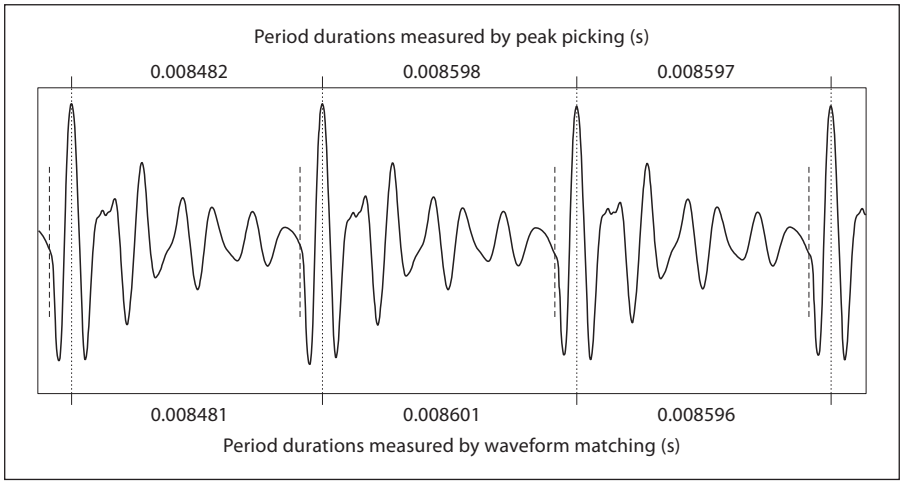

Fig. 2. A sound with 1% jitter.

lines in figure 2) is 0.008472 s, whereas the second underlying period (the time distance between the second and third dashed lines) is 0.008619 s. The difference between these periods is therefore 0.000147 s, which is 1.72% of the average of the 2 periods (0.0085455 s). Likewise, the third underlying period is 0.008596 s so that the difference between the second and third underlying periods is 0.27% of the average of these periods. Averaging these percentages over all underlying periods in a time stretch of 2 s, we arrive at an average local jitter for this sound of 1.004%.

As we can see from the tick marks and the distances between them, both the waveform-matching method and the peak-picking method detect the time differences between the consecutive periods. In fact, both techniques slightly underestimate these differences, apparently because the previous resonances have not yet fully damped out when the next resonances start: waveform matching measures the jitter as 0.827%, peak picking as 0.809%.

Table 1 shows the measured jitter as a function of the underlying jitter of the pulse train, for both methods. The peak-picking method appears twice in the table, once as measured by Praat (parenthesized because it is a nonstandard measurement in Praat that requires more mouse clicks than the waveform-matching method) and once as measured by MDVP (my thanks go to Maria Cristina Jackson-Menaldi of Wayne University, who volunteered to provide the MDVP measurements of these sounds).

The table shows that both methods yield essentially identical results on all sounds with underlying jitter values from 0.001 to 20%: basically correct values for the whole range from 0.001 to 5%, a breakdown from 10% as a result of a failing pitch measurement, and a slight underestimation due to the overlap of the resonances.

Until now, the two programs give identical results. As Titze and Liang [4] observed, however, the methods can

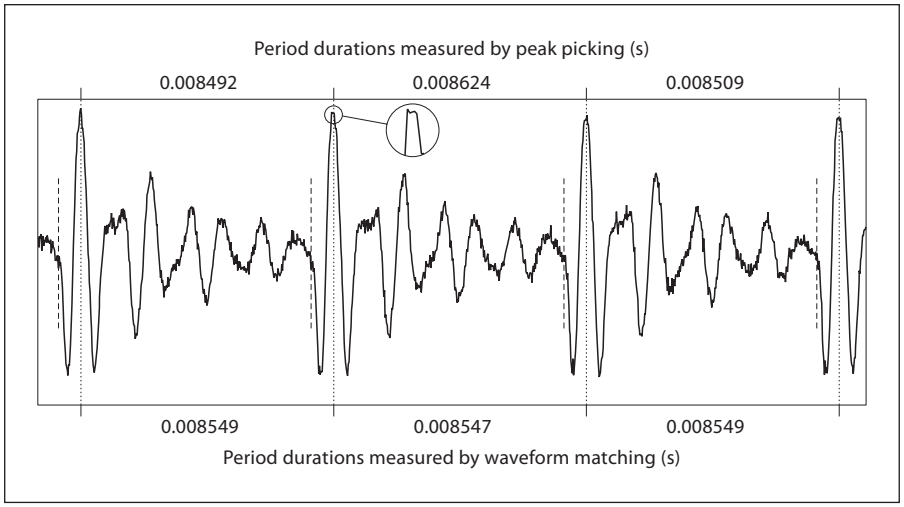

Fig. 3. A sound with 1% additive noise.

yield very different results if noise is added to the sound, and I will now explain this in detail and show that measurements done with Praat and MDVP indeed confirm Titze and Liang's observation.

Consider, then, the sound in figure 3. It is identical to the periodic sound in figure 1, except that white noise, with a power of 1% of the power of the original sound, has been added (at a sampling frequency of 44,100 Hz).

The underlying periods are still 0.008547 s, but the two methods have trouble measuring these periods accurately. The extent of this trouble, however, differs appreciably between the two methods. Waveform matching takes the whole shape of the wave into account and is therefore influenced only slightly by the very local noisy perturbations: in figure 3, the inaccuracy can be seen as 0.000002 s, and averaged over the whole 2 s the waveform-matching method measures a jitter of 0.020%. By contrast, the peak-picking technique looks at the time locations where the waveform is at its maximum, and is therefore strongly influenced by the random perturbations: in figure 3 we can see that the top of the second pulse contains 2 tiny spikes, of which the left one is the higher; as a result, the peak-picking method picks this randomly higher peak and decides that it represents the glottal pulse; in figure 3 we can therefore see that the second tick mark at the top is shifted to the left with respect to the dotted line that comes up from the second tick mark at the bottom; as a result, the peak-picking method underestimates the first period, overestimates the second one and ends up measuring an average jitter of 0.56% for the whole sound. We can conclude that the peak-picking method is 28 times more sensitive to additive white noise than the waveform-matching method, at least for the sustained [a] under consideration here.

Table 1. Jitter measurements for nonnoisy sounds (percent)

Underlying jitter	Praat		MDVP peak picking
	waveform matching	(peak picking)	
0.001	0.001	(0.002)	0.001
0.002	0.002	(0.003)	0.002
0.005	0.004	(0.005)	0.004
0.009	0.007	(0.007)	0.007
0.020	0.016	(0.016)	0.015
0.050	0.041	(0.041)	0.040
0.090	0.074	(0.076)	0.074
0.212	0.171	(0.168)	0.169
0.509	0.413	(0.404)	0.398
1.004	0.827	(0.809)	0.805
2.071	1.763	(1.723)	1.695
2.919	2.644	(2.446)	2.602
3.675	3.468	(3.576)	3.434
4.718	4.542	(4.697)	4.449
9.334	8.080	(7.501)	8.481
18.352	9.594	(8.990)	9.780

If the noisy periodic signal in figure 3 is measured as having a jitter of 0.020 or 0.56%, then one must expect that jitter is difficult to determine for noisy sounds with low underlying jitter values. This indeed turns out to be the case. Table 2 shows that the waveform-matching method can reliably measure the underlying jitter if it is 0.050% or higher and that the peak-picking method can reliably measure the underlying jitter if it is 1% or higher. This means

Table 2. Jitter measurements for sounds with 1% additive white noise (percent)

Underlying jitter	Praat		MDVP peak picking
	waveform matching	(peak picking)	
0.001	0.021	(0.566)	0.562
0.002	0.021	(0.556)	0.553
0.005	0.020	(0.631)	0.747
0.009	0.020	(0.602)	0.928
0.020	0.026	(0.586)	0.585
0.050	0.047	(0.605)	0.604
0.090	0.076	(0.519)	0.518
0.212	0.172	(0.625)	0.816
0.509	0.413	(0.642)	0.639
1.004	0.831	(0.954)	1.079
2.071	1.762	(1.754)	1.728
2.919	2.672	(2.642)	2.773
3.675	3.367	(3.614)	3.430
4.718	4.548	(4.706)	4.417
9.334	8.012	(7.888)	8.001
18.352	9.523	(9.295)	10.037

that the peak-picking method is reliable only for jitter values in pathological ranges (which, according to the MDVP manual, are above 1.03%). For this reason, Praat's standard method is waveform matching rather than peak picking.

The robustness of the jitter measure against additive noise is generally taken to be the quality criterion for jitter measurement methods [4, 5]. In line with the results of the present paper, including its comparisons between Praat and MDVP, Titze and Liang [4] remark:

> The waveform matching method meets the high-precision criterion of being able to extract a 1% frequency change (per cycle) with a 1% accuracy, as long as the signal-to-noise ratio is greater than about 40 dB and concomitant amplitude modulations are below about 5%. [...] Peak-picking and zero-crossing methods do not meet the high-precision criterion consistently, especially not when frequency perturbations are in the normal 0.1 to 1.0% range. Great care must be taken in the interpretation of jitter and shimmer with these single-event detectors because they are not noise-resistant.

Therefore, Titze and Liang conclude that

> Until more is known about the perturbation patterns to be detected in natural voice, it makes sense to use a method that gives the best results for artificially produced patterns (modulations). For these, waveform matching is the clear choice when frequency variations are below about 6% per cycle. For higher variations, no statement about accuracy can be made for any method at this point.

No information gathered in the literature on 'perturbation patterns to be detected in natural voice' since Titze and Liang's paper seems to have been able to modify this verdict.

Given that waveform matching is the method one would choose on the basis of its quality, there remains the problem that only the peak-picking method comes with an established criterion for pathology. As Maryn et al. [1] note: Deliyski's MDVP manual states that jitter values above 1.03% are pathological. Does this mean that for the waveform-matching method 1.03% is a good criterion as well? That depends on whether the criterion was determined for noiseless sounds. If it was, then 1.03% would be a good criterion for both the peak-picking method (under noiseless circumstances) and the waveform-matching method (under both noisy and noiseless circumstances). If, however, the criterion of 1.03% was measured for sounds that could include noise, the criterion has been contaminated by noise (caused by the false alarms of pathological jitter yielded by the peak-picking method) and the criterion for jitter alone (i.e. when the waveform-matching method is used) would have to be some value below 1.03%. When Praat measures jitter values above 1.03, however, we can say that the jitter in the sound is pathological a fortiori.

The reader will now know why Praat's standard method for glottal pulse detection is waveform matching rather than peak picking (as it is in MDVP): it is because I agree with Titze and Liang [4] and Parsa and Jamieson [5] that robustness against additive noise is a relevant criterion for the quality of jitter measurement methods. I also agree with Maryn et al. [1] that pathology thresholds have to be determined for the waveform-matching method. This becomes more urgent now that we know which of the two techniques is preferred on the basis of its quality.

References

1 Maryn P, Corthals P, De Brodt M, Van Cauwenberge P, Deliyski D: Perturbation Measures of Voice: a comparative study between Multi-Dimensional Voice Program and Praat. Folia Phoniatr Logop 2009;61:217–226.
2 Kay Elemetrics Corp: Multi-Dimensional Voice Program (MDVP) Model 5105: Software Instruction Manual. Lincoln Park, Kay Elemetrics, 2003.
3 Boersma P: Stemmen meten met Praat. Stem-, Spraak- en Taalpathologie 2004;12: 237–251.
4 Titze IR, Liang H: Comparison of F_0 extraction methods for high-precision voice perturbation measurements. J Speech Hear Res 1993;36:1120–1133.
5 Parsa V, Jamieson DG: A comparison of high precision F0 extraction algorithms for sustained vowels. J Speech Lang Hear Res 1999; 42:112–126.