

## Should jitter be measured by peak picking or by waveform matching?

Paul Boersma, University of Amsterdam

In their article “Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat”, published in this issue, Maryn, Corthals, De Bodt, Van Cauwenberge and Deliyski make a comparison between the jitter measurements in Praat and 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 give the *best* result. After all, jitter is defined (according to Deliyski’s 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 provide the better jitter measurements, the authors provide only indirect clues. On page xx they acknowledge (following [1]) 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 MDVP’s method is “peak picking”. As for the quality of the two methods, the authors cite (on page xx) Titze and Liang [2] 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 MDVP’s method, but no further explanation is given. The present paper aims at providing the information lacking from Maryn *et al.*’s article 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 yields 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, 1300, 2300 and six 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.

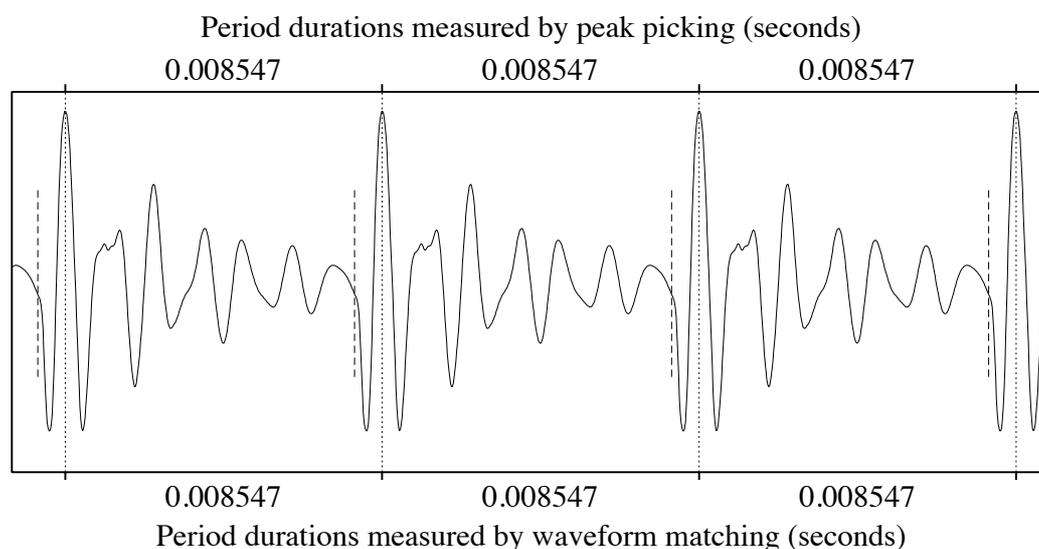


Fig. 1. A perfectly periodic sound.

The tick marks along the bottom of Figure 1 indicate the “glottal pulses” as measured by the waveform-matching method; this method 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; this method 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 seconds.

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 percent. This means that two consecutive underlying periods are on average different by 1 percent. For instance, the first underlying period (the time distance between the first and second dashed lines in Figure 2) is 0.008472 seconds, whereas the second underlying period (the time distance between the second and third dashed lines) is 0.008619 seconds. The difference between these periods is therefore 0.000147 seconds, which is 1.72 percent of the average of the two periods (0.0085455 seconds). Likewise, the third underlying period is 0.008596 seconds, so that the difference between the second and third underlying periods is 0.27 percent of the average of these periods. Averaging these percentages over all underlying periods in a time stretch of 2 seconds, we arrive at an average local jitter for this sound of 1.004 percent.

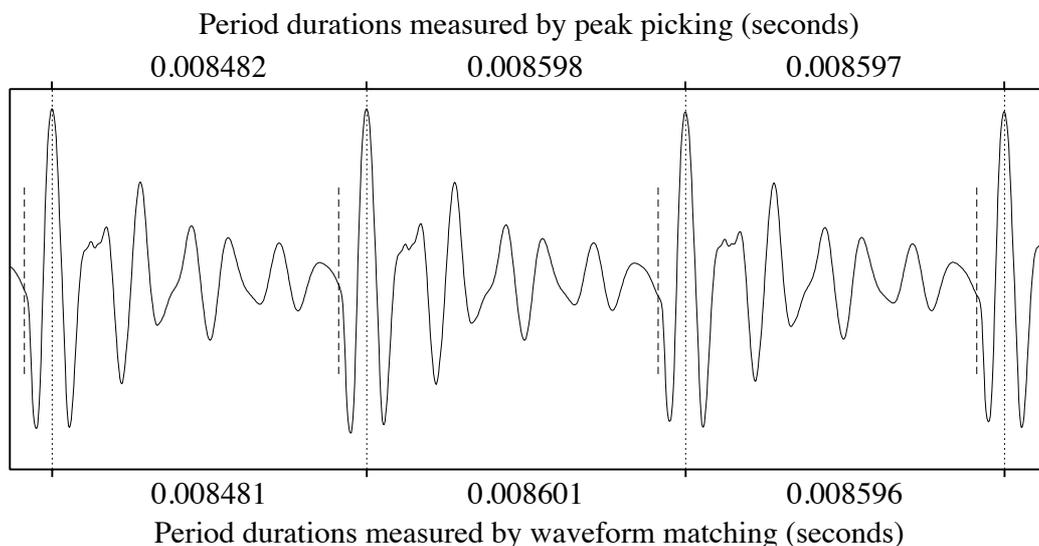


Fig. 2. A sound with 1 percent jitter.

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 methods 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 percent, peak picking as 0.809 percent.

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).

Table 1. Jitter measurements for nonnoisy sounds.

Underlying jitter	Praat waveform matching	(Praat peak picking)	MDVP 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%

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

Until now, the two programs give identical results. As Titze & Liang observed, however, the two methods can 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 & 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 percent of the power of the original sound, has been added (at a sampling frequency of 44100 Hz).

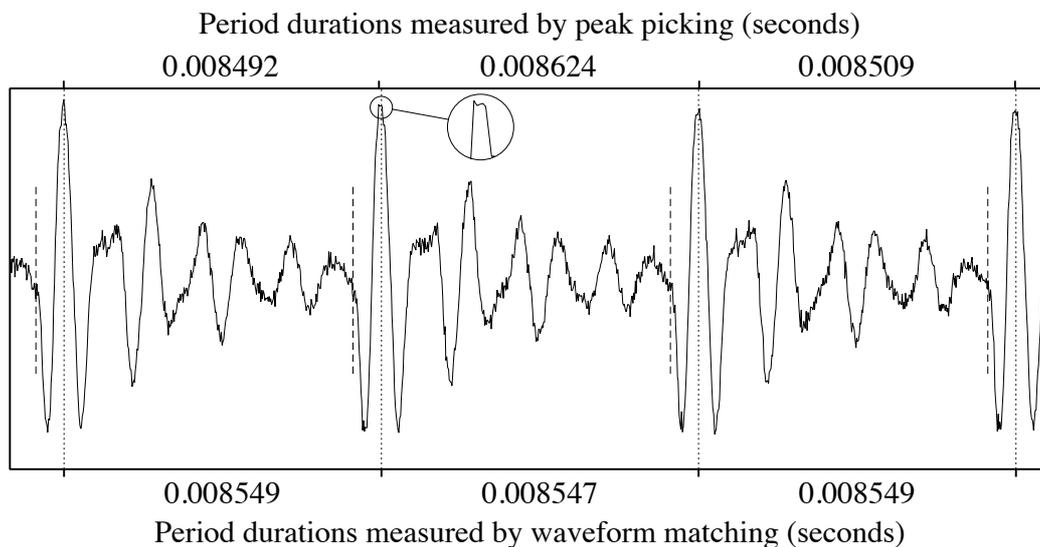


Fig. 3. A sound with 1 percent additive noise.

The underlying periods are still 0.008547 seconds, but the two methods have trouble measuring these periods accurately. The amount of this trouble, however, differs appreciably between the two methods. The waveform-matching method takes the whole shape of the wave into account, and is therefore influenced only slightly by the very local noisy perturbations: in the Figure, the inaccuracy can be seen as 0.000002 seconds, and averaged over the whole 2 seconds the waveform-matching method measures a jitter of 0.020 percent. By contrast, the peak-picking method looks at the time locations where the waveform is at its maximum, and is therefore strongly influenced by the random perturbations: in the Figure we can see that the top of the second pulse contains two 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 the Figure 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 period, and ends up measuring an average jitter of 0.56 percent 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.

If the noisy periodic signal in Figure 3 is measured as having a jitter of 0.020 or 0.56 percent, then one must expect that jitter is difficult to measure 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 percent or higher, and that the peak-picking method can reliably measure the underlying jitter if it is 1 percent or higher. This means that the peak-picking method is reliable only for jitter values in pathological ranges (which, according to the MDVP manual, are jitter values above 1.03 percent). For this reason, Praat's standard method is waveform matching rather than peak picking.

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

Underlying jitter	Praat waveform matching	(Praat peak picking)	MDVP 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%

The robustness of the jitter measure against additive noise is generally taken to be quality criterion for jitter measurement methods [2, 3]. In line with the results of the present paper, including its comparisons between Praat and MDVP, Titze & Liang [2] 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 & 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 & 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.* note: Deliyski’s MDVP manual states that jitter values above 1.03 percent are pathological. Does this mean that for the waveform-matching method 1.03 percent is a good criterion as well? That depends on whether the criterion was determined for noiseless sounds. If it was, then 1.03 percent 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 percent 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 percent. 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 & Liang [2] and Parsa & Jamieson [3] that robustness against additive noise is a relevant criterion for the quality of jitter measurements methods. I also agree with Maryn *et al.* that pathology thresholds have to be determined for the waveform-matching method. This becomes more urgent now that we know which of the two methods is preferred on the basis of its quality.

## References

- [1] Boersma, P. (2004): "Stemmen meten met Praat." *Stem-, Spraak- en Taalpathologie* **12**: 237–251.
- [2] Titze, I.R., and Liang, H. (1993): "Comparison of F<sub>0</sub> extraction methods for high-precision voice perturbation measurements." *Journal of Speech and Hearing Research* **36**: 1120–1133.
- [3] Parsa, V., and Jamieson, D.G. (1999): "A comparison of high precision F<sub>0</sub> extraction algorithms for sustained vowels." *Journal of Speech, Language and Hearing Research* **42**: 112–126.