Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2009, Article ID 784379, 7 pages
doi:10.1155/2009/784379

Research Article

Removing the Influence of Shimmer in the Calculation of Harmonics-To-Noise Ratios Using Ensemble-Averages in Voice Signals

Carlos Ferrer, Eduardo González, María E. Hernández-Díaz, Diana Torres, and Anesto del Toro

Center for Studies on Electronics and Information Technologies, Central University of Las Villas, C. Camajuaní, km 5.5, Santa Clara, CP 54830, Cuba

Correspondence should be addressed to Carlos Ferrer, cferrer@uclv.edu.cu

Received 1 November 2008; Revised 10 March 2009; Accepted 13 April 2009

Recommended by Juan I. Godino-Llorente

Harmonics-to-noise ratios (HNRs) are affected by general aperiodicity in voiced speech signals. To specifically reflect a signal-to-additive-noise ratio, the measurement should be insensitive to other periodicity perturbations, like jitter, shimmer, and waveform variability. The ensemble averaging technique is a time-domain method which has been gradually refined in terms of its sensitivity to jitter and waveform variability and required number of pulses. In this paper, shimmer is introduced in the model of the ensemble average, and a formula is derived which allows the reduction of shimmer effects in HNR calculation. The validity of the technique is evaluated using synthetically shimmered signals, and the prerequisites (glottal pulse positions and amplitudes) are obtained by means of fully automated methods. The results demonstrate the feasibility and usefulness of the correction.

1. Introduction

When the source-filter model of speech production [1] is assumed in Type 1 [2] signals (no apparent bifurcations/chaos), the sources of periodicity perturbations in voiced speech signals can be divided in four classes [3]: (a) pulse frequency perturbations, also known as jitter, (b) pulse amplitude perturbations, also known as shimmer, (c) additive noise, and (d) waveform variations, caused either by changes in the excitation (source) or in the vocal tract (filter) transfer function. Vocal quality measurements have focused mainly in the first three classes (see [4] for a comprehensive survey of methods reported in the previous century). The findings of significant interrelations among measures of jitter, shimmer, and additive noise [5] raised the question on "whether it is important to be able to assign a given acoustic measurement to a specific type of aperiodicity" (page 457). This ability of a measurement to gauge a particular signal attribute, being insensitive to other factors, has been a persistent interest in vocal quality research.

Harmonics-to-Noise-Ratios (HNRs) have been proposed as measures of the amount of additive noise in the acoustic waveform. However, an HNR measure insensitive to all the other sources of perturbation is, if feasible, still to be found. Methods in both time and frequency (or transformed) domain do always have intrinsic flaws. Schoentgen [6] described analytically the effects of the different perturbations in the Fourier spectra of source and radiated waveforms. According to the derivations from his models, it is not possible to perform separate measurements of each type of perturbation by using spectral-based methods. Time domain methods have been criticized [7, 8] for depending on the correct determination of the individual pulse boundaries, among many other method-specific factors.

Yumoto et al. introduced a time-domain method for determining HNR [9], where the energy of the harmonic (repetitive) component is equal to the variance of a pulse "template" obtained as the ensemble average of the individual pulses. The energy of the noise component is calculated

as the variance of the differences between the ensemble and the template (see (4) in Section 2).

The original ensemble-averaging technique has been criticized [10, 11] for its slow convergence with N, the number of averaged pulses. The requirement of large N facilitates the inclusion of slow waveform changes in the ensemble, which are incorrectly treated as noise by the method. The sensitivity of the method to jitter and shimmer has also been reported [5], and many approaches attempting to overcome these limitations have been proposed.

In [12] the need of averaging a large number of pulses is suppressed, by determining an expression which corrects the ensemble-average HNR.

Qi et al. used Dynamic Time Warping (DTW) [13] and later Zero Phase Transforms (ZPTs) [14] of individual pulses prior to averaging to reduce waveform variability (and jitter) influences in the template. For the same purpose the ensemble averaging technique was applied to the spectral representations of individual glottal source pulses in [3], where a pitch synchronous method allowed to account for jitter and shimmer in the glottal waveforms. However, the assumptions are valid only on glottal source signals; hence results are not applicable to vocal tract filtered signals. Functional Data Analysis (FDA) has also been used to perform the optimal time alignment of pulses prior to averaging [15].

Shimmer corrections to ensemble averages HNRs have received lesser attention than pulse duration (jitter) corrections, in spite of being a prerequisite for some of the mentioned jitter correction methods. DTW and FDA, for instance, depart from considering equal amplitude pulses to determine the required expansion/compression of the waveform duration. Besides, shimmer always increases the variability of the ensemble with respect to the template in the reported methods. A normalization of each individual pulse by its RMS value was proposed in [7] to reduce shimmer effects on HNR and was first used on a method that also accounted for jitter and offset effects in [16]. This pulse amplitude (shimmer) normalization can help in the time warping of the pulses and actually reduces the variance of the template in Yumoto's HNR formula. However, it still yields only an approximate value of HNR.

In this paper, an analysis on the original ensemble average HNR formula in the presence of shimmer is performed, which results in a general form of Ferrer's correcting formula [12] and allows the suppression of the effect of shimmer in HNR.

2. Ensemble-Averages HNR Calculation

The most widely used model for ensemble averaging assumes each pulse representation $x_i(t)$ prior to averaging as a repetitive signal $s(t)$ plus a noise term $e_i(t)$:

$$x_i(t) = s(t) + e_i(t). \tag{1}$$

This representation has been used for source [3] and radiated signals [5, 9, 14, 16] as well as for both indistinctly [12, 15]. If we denote the glottal flow waveform as $g(t)$, the vocal tract impulse response as $h(t)$, the radiation at lips as $r(t)$, and the turbulent noise generated at the glottis as $n(t)$, the components of the pulse waveform in (1) can be expressed differently for the source and radiated signals. If (1) represents the excitation signal, then $s(t) = g(t)$, and $e(t) = n(t)$, while for radiated signals $s(t) = g(t) * h(t) * r(t)$ and $e(t) = n(t) * h(t) * r(t)$ [17], with the asterisk denoting the convolution operation. Some important differences between both alternatives are [17] as follows.

(i) HNR measured in the radiated signal differs from HNR in the glottal signal.

(ii) Jitter in the glottal signal produces shimmer in the radiated signal.

(iii) Additive White Gaussian Noise (AWGN) in the glottis (a rough approximation [18] frequently assumed) yields colored noise at the lips.

In the general form of the ensemble average approach, if the noise term $e_i(t)$ is stationary and ergodic and $s(t)$ and $e_i(t)$ are zero mean signals (the typical assumptions in the minimization of the mean squared error [12, 19, 20]) with variances ${\sigma_s}^2$ and ${\sigma_e}^2$, the actual HNR for the set of N pulses is

$$\begin{aligned} \mathrm{HNR} &= \frac{E\left[\sum_{i=1}^{N} s(t)^2\right]}{E\left[\sum_{i=1}^{N} e_i(t)^2\right]} \\ &= \frac{N \times E\left[s(t)^2\right]}{\sum_{i=1}^{N} E\left[e_i(t)^2\right]} \\ &= \frac{{\sigma_s}^2}{{\sigma_e}^2} \end{aligned} \tag{2}$$

with $E[\,]$ denoting the expected value operation. The ensemble averaging method proposed by Yumoto et al. [9] is based on the use of a pulse template $\overline{x}(t)$ as an estimate of the repetitive component $s(t)$:

$$\begin{aligned} \overline{x}(t) &= \frac{\sum_{i=1}^{N} x_i(t)}{N} \\ &= s(t) + \frac{\sum_{i=1}^{N} e_i(t)}{N}. \end{aligned} \tag{3}$$

This approximation to $s(t)$ is then used to obtain an estimate of $e_i(t)$ according to (1), and both estimates are used in (2) to produce Yumoto's HNR formula:

$$\mathrm{HNR}_{\mathrm{Yum}} = \frac{N \times E\left[\overline{x}^2(t)\right]}{\sum_{i=1}^{N} E\left[\left(x_i(t) - \overline{x}(t)\right)^2\right]}. \tag{4}$$

The bias produced in $\mathrm{HNR}_{\mathrm{Yum}}$ due to the use of (3) on its calculation and the terms needed to correct it are described in [12], where it is shown that

$$\mathrm{HNR} = \frac{{\sigma_s}^2}{{\sigma_e}^2} = \frac{N-1}{N}\mathrm{HNR}_{\mathrm{Yum}} - \frac{1}{N}. \tag{5}$$

However, the model previously described neglects the effect of shimmer when the different replicas of the repetitive signal are of different amplitude.

3. Insertion of Shimmer in the Model

To account for shimmer, a variable a_i can be added to the model in (1):

$$x_i(t) = a_i s(t) + e_i(t). \tag{6}$$

For this model, the actual HNR is

$$\begin{aligned} \text{HNR} &= \frac{E\left[\sum_{i=1}^{N}(a_i s(t))^2\right]}{E\left[\sum_{i=1}^{N} e_i(t)^2\right]} \\ &= \frac{\sum_{i=1}^{N} a_i{}^2 E\left[s(t)^2\right]}{\sum_{i=1}^{N} E\left[e_i(t)^2\right]} \\ &= \frac{\sum_{i=1}^{N} a_i{}^2 \sigma_s{}^2}{N\sigma_e{}^2}. \end{aligned} \tag{7}$$

Using the original ensemble average procedure, the template yields

$$\bar{x}(t) = \frac{\sum_{i=1}^{N} x_i(t)}{N} = \frac{s(t)\sum_{i=1}^{N} a_i + \sum_{i=1}^{N} e_i(t)}{N}, \tag{8}$$

and its variance is

$$\begin{aligned} &\sigma_{\bar{x}}^2 \\ &= E[\bar{x}^2(t)] \\ &= \frac{E[(s(t)\sum_{i=1}^{N} a_i)^2 + 2s(t)\sum_{i=1}^{N} e_i(t)\sum_{k=1}^{N} a_k + \sum_{i=1}^{N} e_i(t)\sum_{k=1}^{N} e_k(t)]}{N^2}. \end{aligned} \tag{9}$$

If $e_i(t)$ is uncorrelated with $s(t)$ or any $e_k(t)$ such that $k <> i$, the second term between brackets in (9) as well as all the products in the third term where $k <> i$ can be suppressed:

$$\begin{aligned} E[\bar{x}^2(t)] &= \frac{\left(\sum_{i=1}^{N} a_i\right)^2 E\left[s(t)^2\right] + \sum_{i=1}^{N} E\left[e_i(t)^2\right]}{N^2} \\ &= \left(\sum_{i=1}^{N} a_i\right)^2 \frac{\sigma_s^2}{N^2} + \frac{\sigma_e^2}{N}. \end{aligned} \tag{10}$$

With the inclusion of shimmer in the model, the denominator in (4) is

$$\begin{aligned} \text{Den} &= \sum_{i=1}^{N} E\left[(x_i(t) - \bar{x}(t))^2\right] \\ &= \sum_{i=1}^{N} E\left[\left(a_i s(t) + e_i(t) - \sum_{j=1}^{N} \frac{a_j s(t)}{N} - \sum_{j=1}^{N} \frac{e_j(t)}{N}\right)^2\right] \\ &= \sum_{i=1}^{N} E\left[\left(a_i \frac{(N-1)}{N} s(t) - \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{a_j}{N} s(t) \right.\right. \\ &\qquad \left.\left. + e_i(t)\frac{(N-1)}{N} - \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j(t)}{N}\right)^2\right]. \end{aligned} \tag{11}$$

To simplify further derivations, the letters m, n, o, and p are used to represent the four terms summed and squared in (11):

$$\begin{aligned} m &= a_i \frac{(N-1)}{N} s(t), \qquad n = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{a_j}{N} s(t), \\ o &= e_i(t)\frac{(N-1)}{N}, \qquad p = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j(t)}{N}. \end{aligned} \tag{12}$$

Using (12), (11) can be written as

$$\begin{aligned} \text{Den} = \sum_{i=1}^{N} E[&m^2 + n^2 + o^2 + p^2 + 2mn + 2mo + 2mp \\ &+ 2no + 2np + 2op], \end{aligned} \tag{13}$$

where the last five terms between brackets can be suppressed, since $E[e_i(t)e_j(t)] = 0$ for any $i <> j$. From the first five terms, it was already shown in [12] that

$$\sum_{i=1}^{N} E[o^2 + p^2] = (N-1)\sigma_e^2. \tag{14}$$

The summations of the other nonzero expected values ($E[m^2]$, $E[n^2]$ and $E[2mn]$) are examined as follows:

$$\begin{aligned} \sum_{i=1}^{N} E[m^2] &= \sum_{i=1}^{N} E\left[a_i^2 \frac{(N-1)^2}{N^2} s^2(t)\right] \\ &= \frac{(N-1)^2 \sum_{i=1}^{N} a_i^2}{N^2} \sigma_s^2, \end{aligned} \tag{15}$$

while

$$\sum_{i=1}^{N} E[n^2] = \sum_{i=1}^{N} E\left[\frac{s^2(t)}{N^2}\sum_{\substack{j=1\\ j\neq i}}^{N} a_j \sum_{\substack{k=1\\ k\neq i}}^{N} a_k\right] = \frac{\sigma_s^2}{N^2}\sum_{i=1}^{N}\left(\sum_{\substack{j=1\\ j\neq i}}^{N} a_j \sum_{\substack{k=1\\ k\neq i}}^{N} a_k\right), \tag{16}$$

and using

$$\sum_{i=1}^{N}\left(\sum_{\substack{j=1\\ j\neq i}}^{N} a_j \sum_{\substack{k=1\\ k\neq i}}^{N} a_k\right) = \left(\sum_{i=1}^{N}(a_i)^2 + (N-2)\left(\sum_{i=1}^{N} a_i\right)^2\right) \tag{17}$$

(16) yields

$$\sum_{i=1}^{N} E[n^2] = \frac{\sigma_s^2}{N^2}\left(\sum_{i=1}^{N}(a_i)^2 + (N-2)\left(\sum_{i=1}^{N} a_i\right)^2\right). \tag{18}$$

Finally

$$\sum_{i=1}^{N} E[2mn] = \frac{-2(N-1)E[s^2(t)]}{N^2}\sum_{i=1}^{N} a_i \sum_{\substack{j=1\\ j\neq i}}^{N} a_j, \tag{19}$$

since

$$\left(\sum_{i=1}^{N} a_i\right)^2 = \sum_{i=1}^{N}(a_i)^2 + \sum_{i=1}^{N} a_i \sum_{\substack{j=1\\ j\neq i}}^{N} a_j, \tag{20}$$

then (19) results in

$$\sum_{i=1}^{N} E[2mn] = -2\sigma_s^2\frac{(N-1)}{N^2}\left(\left(\sum_{i=1}^{N} a_i\right)^2 - \sum_{i=1}^{N}(a_i)^2\right). \tag{21}$$

The sum of (15), (18), and (21) is

$$\sum_{i=1}^{N} E[m^2 + n^2 + 2mn] = \sigma_s^2\left(\sum_{i=1}^{N}(a_i^2) - \left(\sum_{i=1}^{N} a_i\right)^2\frac{1}{N}\right). \tag{22}$$

Now, substituting (14) and (22) in the denominator of (4) and (10) in the numerator gives

$$\mathrm{HNR_{Yum}} = \frac{\left(\sum_{i=1}^{N} a_i\right)^2(\sigma_s^2/N) + \sigma_e^2}{\sigma_s^2\left(\sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i\right)^2(1/N)\right) + \sigma_e^2(N-1)}. \tag{23}$$

From (23) the ratio of signal and noise variances can be determined as

$$\frac{\sigma_s^2}{\sigma_e^2} = \frac{[\mathrm{HNR_{Yum}}(N-1)-1]}{\left(\sum_{i=1}^{N} a_i\right)^2(1/N) - \mathrm{HNR_{Yum}}\left(\sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i\right)^2(1/N)\right)}, \tag{24}$$

and the actual HNR given by (7) can be rewritten as

$$\mathrm{HNR} = \frac{[\mathrm{HNR_{Yum}}(N-1)-1]\sum_{i=1}^{N} a_i^2}{\left(\sum_{i=1}^{N} a_i\right)^2 - \mathrm{HNR_{Yum}}\left(N\sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i\right)^2\right)}. \tag{25}$$

Equation (25) can be simplified by using a factor K defined as

$$K = \frac{N\sum_{i=1}^{N} a_i^2}{\left(\sum_{i=1}^{N} a_i\right)^2} \tag{26}$$

and HNR expressed as

$$\mathrm{HNR} = \frac{K[\mathrm{HNR_{Yum}}(N-1)-1]}{N(1-\mathrm{HNR_{Yum}}(K-1))}. \tag{27}$$

According to (26), K will be a positive number ranging from one (in the no-shimmer case, being all a_i equal) to N when a single pulse is a lot greater than all the others. The latter situation is not the case in voiced signals, where the largest shimmer almost never exceeds the 50% of the mean amplitude [2] in extremely pathological voices. Equation (27) is a generalization of Ferrer's correcting formula [12] expressed in (5), being equal in the no-shimmer case ($K = 1$).

4. Experiment

The calculation of (27) requires the prior determination of both pulse boundaries and amplitudes. Pulse boundaries are usually determined by means of a cycle-to-cycle pitch detection algorithm (PDA). The determination of pulse amplitudes relies on the pitch contour detected by the PDA, and a comparison of several amplitude measures can be found in [21]. In practice, the detected pulse boundaries and amplitudes differ from the real ones, causing a reduction in the theoretical usefulness of (27). An additional deterioration can be expected in the presence of correlated noise, as should be the case in radiated speech signals.

To evaluate the effects of these deteriorations, synthetic voiced signals were used with known pulse positions, noise and shimmer levels. The synthesis procedure of the speech signal $s(t)$ is described by (28):

$$s(t) = h(t) * \sum_{i=1}^{M} k_i g'(t - iT_0) + e(t), \tag{28}$$

where $h(t)$ is the vocal tract impulse response, $*$ denotes the convolution operation, k_i is the variable pulse amplitude, $g(t)$ is the glottal flow waveform, i is the pulse number, T_0 is the pitch period, and $e(t)$ is the additive noise in the signal. The effect of lip radiation has been included as the first derivative operation present in $g'(t)$. This synthesis procedure is similar to the one used in [12, 19, 21, 22], but using a more refined glottal excitation than an impulse train. In this case, a train of Rosemberg's type B polynomial model pulses [23] was chosen; this alternative is used in [3, 24].

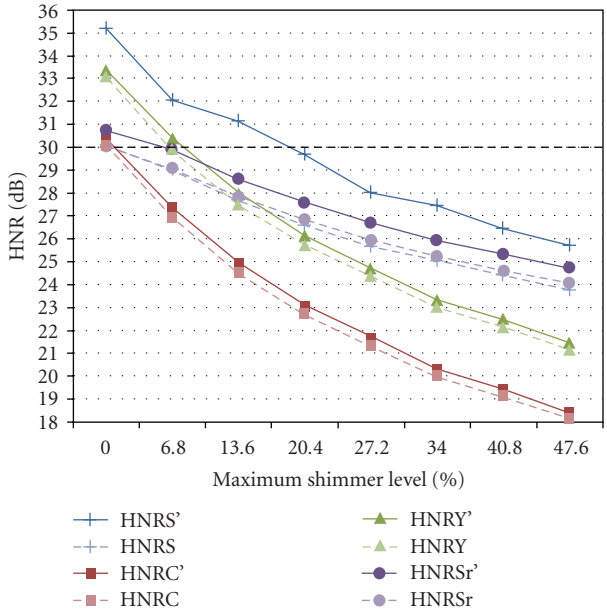

Figure 1: Results for the different HNR estimation methods. HNRY (in triangles) is the original formula in [9], HNRC (squares) the pulse number correction in [12], HNRS (plus signs) the shimmer correction proposed here (using known pulse amplitudes), and HNRSr (circles) the shimmer correction using estimated pulse amplitudes. Dashed lines represent results with AWGN; solid lines and apostrophes represent vocal tract filtered AWGN. Horizontal dashed line at 30 dB represents true HNR.

The discrete implementation of (28) was performed by setting a sampling frequency of 22050 Hz, a fundamental frequency of 150 Hz (yielding 147 samples per period), and $M = 300$, to produce an approximate of 2 seconds of synthesized voice. The $h(t)$ was obtained as the impulse response of a five formant all-pole filter, with the same parameters used in [12, 19, 21, 22]. The glottal flow was generated using a rising time of $0.33T_0$ and a falling time of $0.09T_0$; the values which resulted in the most natural-sounding synthesis in [23].

The shimmer was controlled by changing the value of each pulse amplitude k_i, obtained as $k_i = 1 + v_i$, where v_i is a random real value, uniformly distributed in the interval $\pm v_m$. Eight levels of shimmer were synthesized, using values of v_m from 0% to 47.6% in steps of 6.8%, measured in percent of the unaltered amplitude $k = 1$, the same values as in [12, 21].

The estimates of HNR calculated were the original ensemble average formula by Yumoto given in (4), the correction for any number of pulses given in (5), and the removal of shimmer effects given by (27). The three HNR estimates were calculated using first the known pulse durations and amplitudes, and then using the positions given by a well-known PDA (the superresolution approach from Medan et al. [19]), and the amplitudes were calculated with Milenkovic's formula [20] using the procedure described in [21].

A base level of noise was added to the signal, to avoid values near to zero in the denominator of HNR_{Yum} in (4). The variance of the noise added was chosen to produce an actual HNR = 1000 (30 dB). Two types of noise were added: AWGN, in conformity with the assumptions of uncorrelated noise made on deriving (27), and a vocal tract filtered version, having some level of correlation which is most likely the case in radiated signals.

The HNR estimates were found for ensembles of two consecutive pulses ($N = 2$) in the synthesized signals, and the overall HNR was found as the average of these pairwise HNR's.

5. Results and Discussion

The average value for 100 realizations of the random variables involved (noise and shimmer) was found for each HNR estimation variant on each shimmer level. It is relevant to note that the PDA detected the pulse positions without any error (not even a sample), for all realizations and all levels of shimmer. For this reason, (4) and (5) produced the same results using both the known and the detected pulse positions. Equation (27) produced different results since it involves also the calculation of the amplitude ratios among pulses, which produced results different to the values used in the synthesis.

The results for the different methods facing both noise types are shown in Figure 1, and the discussion below is first centered in the AWGN and later in the effect of the correlation present in the vocal tract filtered noise.

AWGN. For the zero-shimmer level the results are as predicted: the original approach (*HNRY*) overestimates the actual HNR (30 dB), while the corrected approaches produce adequate and equivalent results. When shimmer appears, *HNRC* begins to fall in parallel with *HNRY*, while both approaches considering shimmer, *HNRS* and *HNRSr*, show superior performance, with their values less affected by the increasing levels of shimmer.

Two relevant facts are as follows.

(i) Shimmer-corrected approaches (*HNRS* and *HNRSr*) are nevertheless deteriorated by the shimmer level.

(ii) There is a better performance of *HNRSr* in comparison with *HNRS*, in spite of using estimated values for the pulse amplitudes.

Both facts can be explained by the presence, in any pulse of the signal, of the decaying tails of previous pulses. This summation of tails adds differences to the pulses, interpreted as noise in the model and causing a reduction in the calculated HNR as the introduced shimmer increases. On the other hand, the summation of tails in one pulse is not completely uncorrelated with the summation of tails in the other. For this reason, the estimation of relative pulse amplitudes, based in the assumption of uncorrelated noise, produces amplitudes with an overestimation of the signal component, yielding a higher *HNRSr* than *HNRS*.

It is to be expected that in the presence of jitter *HNRSr* will perform worse, since pulse tails would not always be aligned with the adjacent pulse, and the correlation should

be lower. The evaluation of the influence of jitter (as well of other levels of noise and their combinations) in the performance of the PDA and *HNRSr* would require extensive tests and is out of the scope of this paper.

Vocal tract filtered AWGN. When noise is not uncorrelated as assumed in the derivation of (27), a fraction of it is regarded as signal, incrementing HNR estimates (solid lines) in all variants with respect to the results with uncorrelated noise (dashed lines). A significant fact is that this overestimation is more relevant in *HNRS* (plus signs in Figure 1) than in *HNRSr* (circles). The correlated contributions of noise and shimmered tails add to what is considered signal by the model in *HNRS*, while in *HNRSr* this effect seems to be compensated by its related consequence in estimating pulse amplitudes with the same assumptions about noise and signal correlations.

In general, shimmer corrections with estimated amplitude contours (*HNRSr*, in circles in Figure 1) produce the closest estimates to the true HNR, which for these experiments would be the flat horizontal line at 30 dB shown in Figure 1.

6. Conclusions

The performed analysis shows that shimmer effects can be reduced in HNR estimations based in the ensemble-averages technique using similar assumptions than in [3, 20]. The requirements for the calculation of (27) (detection of pulse positions and amplitudes) can be performed with satisfactory results using available methods.

More tests should be performed considering more types of perturbations (different noise and jitter values, as well as their combinations) as well as different vocal tract configurations. However, the experiments in this paper were performed using configurations reported in other works, and based on the preliminary results shown, the proposed approach appears to be an alternative for the estimation of HNR in the time domain superior to previous ensemble averages techniques.

Acknowledgments

This research was partially funded by the Canadian International Development Agency Project Tier II-394-TT02-00 and by the Flemish VLIR-UOS Program for Institutional University Cooperation (IUC).

References

[1] G. Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, The Netherlands, 1960.

[2] I. R. Titze, *Workshop on Acoustic Voice Analysis: Summary Statement*, National Center for Voice and Speech, 1994.

[3] P. J. Murphy, "Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2866–2881, 1999.

[4] E. H. Buder, "Acoustic analysis of vocal quality: a tabulation of algorithms 1902–1990," in *Voice Quality Measurement*, R. D. Kent and M. J. Ball, Eds., pp. 119–244, Singular, San Diego, Calif, USA, 2000.

[5] J. Hillenbrand, "A methodological study of perturbation and additive noise in synthetically generated voice signals," *Journal of Speech and Hearing Research*, vol. 30, no. 4, pp. 448–461, 1987.

[6] J. Schoentgen, "Spectral models of additive and modulation noise in speech and phonatory excitation signals," *Journal of the Acoustical Society of America*, vol. 113, no. 1, pp. 553–562, 2003.

[7] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.

[8] Y. Qi and R. E. Hillman, "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals," *Journal of the Acoustical Society of America*, vol. 102, no. 1, pp. 537–543, 1997.

[9] E. Yumoto, W. J. Gould, and T. Baer, "The harmonic-to-noise ratio as an index of the degree of hoarseness," *Journal of the Acoustical Society of America*, vol. 71, pp. 1544–1550, 1982.

[10] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1329–1334, 1986.

[11] J. Schoentgen, M. Bensaid, and F. Bucella, "Multivariate statistical analysis of flat vowel spectra with a view to characterizing dysphonic voices," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 6, pp. 1493–1508, 2000.

[12] C. Ferrer, E. González, and M. E. Hernández-Díaz, "Correcting the use of ensemble averages in the calculation of harmonics to noise ratios in voice signals," *Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 605–607, 2005.

[13] Y. Qi, "Time normalization in voice analysis," *Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2569–2576, 1992.

[14] Y. Qi, B. Weinberg, N. Bi, and W. J. Hess, "Minimizing the effect of period determination on the computation of amplitude perturbation in voice," *Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2525–2532, 1995.

[15] J. C. Lucero and L. L. Koenig, "Time normalization of voice signals using functional data analysis," *Journal of the Acoustical Society of America*, vol. 108, no. 4, pp. 1408–1420, 2000.

[16] N. B. Cox, M. R. Ito, and M. D. Morrison, "Data labeling and sampling effects in harmonics-to-noise ratios," *Journal of the Acoustical Society of America*, vol. 85, no. 5, pp. 2165–2178, 1989.

[17] P. J. Murphy, K. G. McGuigan, M. Walsh, and M. Colreavy, "Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals," *Journal of the Acoustical Society of America*, vol. 123, no. 3, pp. 1642–1652, 2008.

[18] R. E. Hillman, E. Oesterle, and L. L. Feth, "Characteristics of the glottal turbulent noise source," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 691–694, 1983.

[19] Y. Medan, E. Yair, and D. Chazan, "Super resolution pitch determination of speech signals," *IEEE Transactions on Signal Processing*, vol. 39, no. 1, pp. 40–48, 1991.

[20] P. Milenkovic, "Least mean square measures of voice perturbation," *Journal of Speech and Hearing Research*, vol. 30, no. 4, pp. 529–538, 1987.

[21] C. Ferrer, E. González, and M. E. Hernández-Díaz, "Using waveform matching techniques in the measurement of shimmer in voice signals," in *Proceedings of the 8th Annual Conference of the International Speech Communication Association (INTERSPEECH '07)*, pp. 1214–1217, Antwerp, Belgium, August 2007.

[22] V. Parsa and D. G. Jamieson, "A comparison of high precision Fo extraction algorithms for sustained vowels," *Journal of Speech, Language, and Hearing Research*, vol. 42, pp. 112–126, 1999.

[23] A. E. Rosemberg, "Effect of glottal pulse shape on the quality of natural vowels," *Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 583–590, 1971.

[24] I. R. Titze and H. Liang, "Comparison of Fo extraction methods for high-precision voice perturbation measurements," *Journal of Speech, Language, and Hearing Research*, vol. 36, pp. 1120–1133, 1993.