

# TIME-FREQUENCY-MODULATION REPRESENTATION OF STOCHASTIC SIGNALS

Philip J.B. Jackson

Centre for Vision, Speech & Signal Processing (CVSSP), Univ. of Surrey, Guildford, UK.

## ABSTRACT

When a noise process is modulated by a deterministic signal, it is often useful to determine the signal's parameters. A method of estimating the modulation index  $m$  is presented for noise whose amplitude is modulated by a periodic signal, using the magnitude modulation spectrum (MMS). The method is developed for application to real discrete signals with time-varying parameters, and extended to a 3D time-frequency-modulation representation. In contrast to squared-signal approaches, MMS behaves linearly with the modulating function allowing separate estimation of  $m$  for each harmonic. Simulations evaluate performance on synthetic signals, compared with theory, favouring a first-order MMS estimator.

*Index Terms*— time-frequency, modulation, estimation

## 1. TIME-VARYING NOISE ANALYSIS

A broadband stochastic signal modulated by a periodic signal differs substantially from a typical telecommunications or active sensing system in which a known or deterministic carrier signal is modulated by a band-limited message signal or signature. Nonetheless, noise-like signals can become modulated in many practical situations. Any stochastic process subject to periodic disturbances may provide signals with these characteristics, such as biological processes affected by heartbeat or breathing, chemical plants by daily temperature variations, and radiological emissions or electrical noise influenced by rotating machinery. In audio, these include passive underwater sonar, turbine condition monitoring, or road noise in a car. The author's interest is voiced fricative speech sounds (e.g., /z/) in which frication noise (the /s/-like part) is modulated by the quasi-periodic voicing source [1]. In order to describe quantitatively the observed modulation, we wish to extract an efficient estimate of modulation index from recorded signals.

In most amplitude modulation (AM) systems, the carrier signal is a sinusoid of much higher frequency than those in the modulating signal. Detection and demodulation of such signals is comprehensively covered in signal processing texts (e.g., [2, 3, 4]). But, we have here an unpredictable wideband

noise-like carrier, modulated by a narrowband, periodic signal. Treatment of narrowband Gaussian signals in Gaussian white noise provides a framework for detection and parameter estimation [5] and optimal realizable receiver design [6]. Although Gaussian noise with periodic amplitude modulation is not specifically covered, this approximation to our case falls into the class of signals discussed in [6], chapter 3. An important feature of optimal demodulation is that it include a matched filter (aka. inverse kernel), and envelope extraction (usually by squaring and smoothing). The present method uses the *magnitude modulation spectrum* (MMS) to match the modulating frequency. The squaring operation normally used within the optimal receiver, being non-linear, introduces harmonic distortion. This kind of distortion also occurs, for example, with the Teager-Kaiser energy operator, as discussed in [7]. Instead, the magnitude of complex coefficients is taken here, across a narrow band of frequencies. So, the proposed technique also contains the main features of an optimal filter.

Our approach takes the magnitude of the signal, then performs frequency analysis to find and quantify the dominant modulation on the noise envelope. It can also deal with other harmonic or inharmonic modulation components. Two modulation-index estimators are examined, based on first and second moments of the MMS. Validation is obtained from simulation results, rather than mathematical proof, although statistics and a form of likelihood maximization are derived. Simulations to evaluate the estimation accuracy provide empirical evidence of the performance and robustness. The magnitude operation can be applied not just to the time series for determining the noise envelope but to any short-term Fourier transform (STFT) representation. Thus, similar to an acoustic and modulation frequency representation [8], the MMS is extended to time-frequency elements (or atoms) without harmonic distortion artefacts. This provides a third dimension for visualization or parameterization: time, frequency and modulation frequency. The trade off in resolution between these can be optimized according to the application.

## 2. MMS THEORY AND REPRESENTATION

### 2.1. Magnitude modulation spectrum (MMS)

We first consider estimating the modulation index for noise modulated by a single fundamental frequency  $f$ , under sinu-

---

Email: p.jackson@surrey.ac.uk. Thanks to George Szajnowski, Josef Kittler and Jonathan Pincas for helpful comments on an earlier draft.

	mean	variance
$x_n$	0	$\sigma^2 (1 + m \cos \omega_1 n)^2$
$y_n$	$\sqrt{\frac{2}{\pi}} \sigma (1 + m \cos \omega_1 n)$	$\sigma^2 (1 + m \cos \omega_1 n)^2 (1 - \frac{2}{\pi})$
$z_n$	$\sigma^2 (1 + m \cos \omega_1 n)^2$	$2\sigma^4 (1 + m \cos \omega_1 n)^4$

**Table 1.** Time-domain expectation mean and variance for modulated signal  $x_n$ , its magnitude  $y_n = |x_n|$ , and power  $z_n = x_n^2$ .

mod. num.	0	$\pm k_1$	$\pm 2k_1$	otherwise
$X_k$ mean	0	0	0	0
var.	$N\sigma^2 M$	$N\sigma^2 M$	$N\sigma^2 M$	$N\sigma^2 M$
$Y_k$ mean	$\sqrt{\frac{2}{\pi}} N\sigma$	$\frac{m}{\sqrt{2\pi}} N\sigma$	0	0
var.	$N\theta_y^2$	$N\theta_y^2$	$N\theta_y^2$	$N\theta_y^2$
$Z_k$ mean	$N\sigma^2 M$	$N\sigma^2 m$	$N\sigma^2 \frac{m^2}{4}$	0
var.	$2N\sigma^4 M^2$	$2N\sigma^4 M^2$	$2N\sigma^4 M^2$	$2N\sigma^4 M^2$

**Table 2.** Frequency-domain expectation mean and variance for spectra of the modulated signal  $X_k$ , its magnitude  $Y_k$  and power signal  $Z_k$ . We define  $M = (1 + m^2/2)$ , and so the noise floors for  $W_k$  and  $Y_k$  are related  $\theta_y^2 = M\theta_w^2$ .

soidal amplitude modulation (SAM). Taking finite discrete measurements, a window function is used to accommodate effects of time-varying signal parameters. Hence an expression is obtained for calculating practical estimates of the modulation index, including at harmonics of  $f$ .

In AM, the carrier signal,  $w_n$ , is multiplied by a gain,  $a_n$ , to produce an amplitude-modulated signal,  $x_n = w_n a_n$ . For the purposes of modeling broadband noise, let us assume that  $w_n$  is Gaussian white noise with variance  $\sigma^2$ . In the SAM case, the modulating function  $a_n$  is a dc gain with a pure tone:  $x_n = w_n (1 + m \cos(\omega_0 n + \phi))$ , where  $m$  is the modulation index (or modulation depth),  $\phi$  is an arbitrary phase offset, and  $\omega = 2\pi f / f_S$  is the normalized angular frequency in terms of  $f$  and sampling frequency  $f_S$ .

More generally,  $a_n$  can contain multiple frequencies  $f_h$ :

$$x_n = w_n \left( 1 + \sum_{h=1}^H m_h \cos(\omega_h n + \phi_h) \right), \quad (1)$$

where  $m_h$  and  $\phi_h$  are now modulation indices and phase offsets for each one. The derivation below describes the SAM case, but the result for combinations of modulating frequencies follows trivially, for  $|a_n| \geq 0$ .

Taking  $N$  points of modulated noise  $x_n$ , then its spectrum is represented as  $X_k = \mathcal{F}\{x_n\}$  in the frequency domain, and defining  $y_n = |x_n|$  we get  $Y_k = \mathcal{F}\{y_n\}$ . Angular frequency is related to modulation number  $k_h = N f_h / f_S = N \omega_h / 2\pi$ , the spectral frequency bin where the modulation frequency  $f_h$  resides. This finite frame of samples allows signal parameters ( $\sigma$ ,  $m_h$ ,  $f_h$  and  $\phi_h$ ) to be updated for each frame.

**Time-domain statistics.** To derive the modulation estimate, let us consider the statistical properties of  $x_n$  and  $y_n$

for SAM noise. The first and second moments are  $E[x_n] = 0$  and  $E[x_n^2] = \sigma^2 (1 + m \cos(\omega_1 n + \phi))^2$  for the input signal,  $E[y_n] = \sqrt{\frac{2}{\pi}} \sigma (1 + m \cos(\omega_1 n + \phi))$  and  $E[y_n^2] = \sigma^2 (1 + m \cos(\omega_1 n + \phi))^2$  for its magnitude, where  $E[\cdot]$  denotes the expectation. Their time averages (assuming that  $N$  captures whole periods) are  $\langle E[x_n] \rangle = 0$ ,  $\langle E[y_n] \rangle = \sqrt{\frac{2}{\pi}} \sigma$ , and  $\langle E[x_n^2] \rangle = \langle E[y_n^2] \rangle = \sigma^2 (1 + \frac{m^2}{2})$ . However, the variance of  $y_n^2$  is only  $(1 - \frac{2}{\pi}) \sigma^2 (1 + \frac{m^2}{2})$ . Note that  $\sqrt{2/\pi} \approx 0.798$  and  $(1 - 2/\pi) \approx 0.363$ .

**Frequency-domain statistics.** The conventional DFT pair,  $X_k = \mathcal{F}\{x_n\} = \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi n k}{N}}$ ;  $x_n = \mathcal{F}^{-1}\{X_k\} = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j \frac{2\pi n k}{N}}$ , is used here. For all values of  $X_k$  in the frequency domain, this gives  $E[X_k] = 0$  and  $E[X_k^* X_k] = N\sigma^2 (1 + \frac{m^2}{2})$  for the original signal spectrum (\* denotes complex conjugate). So, averages across all frequencies are identical. The magnitude-spectrum,  $Y_k$ , defined as the MMS, has expected mean value with a spike at dc plus a pair of spikes from the modulation:

$$E[Y_k] = \begin{cases} \sqrt{\frac{2}{\pi}} N\sigma & \text{for } k = 0, \\ \frac{m}{\sqrt{2\pi}} N\sigma e^{j\phi \operatorname{sgn} k} & \text{for } k = \pm k_1, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The MMS's second moment reveals a noise floor  $\theta^2$ :

$$E[Y_k^* Y_k] = \begin{cases} \frac{2}{\pi} N^2 \sigma^2 + \theta^2 & \text{for } k = 0, \\ \frac{m^2}{2\pi} N^2 \sigma^2 + \theta^2 & \text{for } k = \pm k_1, \\ \theta^2 & \text{otherwise.} \end{cases} \quad (3)$$

The frequency averages are  $\langle E[Y_k] \rangle = \sqrt{\frac{2}{\pi}} \sigma (1 + m \cos \phi)$  and  $\langle E[Y_k^* Y_k] \rangle = \frac{2}{\pi} N\sigma^2 (1 + \frac{m^2}{2}) + \theta^2$ , which Parseval's theorem relates to the time averages  $\langle E[Y_k^* Y_k] \rangle = N \langle E[y_n^2] \rangle$ . Thus, we can calculate  $\theta^2 = (1 - \frac{2}{\pi}) N\sigma^2 (1 + \frac{m^2}{2})$ , where  $N\sigma^2 (1 + \frac{m^2}{2})$  is obtained from  $\sum_n E[x_n^2]$ ,  $\sum_n E[y_n^2]$  or  $\langle E[X_k^* X_k] \rangle$ . The time- and frequency-domain statistics with  $\phi_1 = 0$  are summarized in Tables 1 and 2, respectively.

**Estimators.** The first-order statistic in eq. 2 leads to an estimate of the modulation index that is biased by noise,

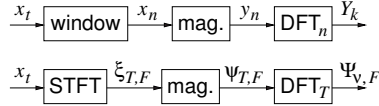
$$\hat{m}_1 = 2 |Y_{k_1}| / |Y_0| = 2(Y_{k_1}^* Y_{k_1} / Y_0^* Y_0)^{1/2}; \quad (4)$$

whereas the second-order statistic from eq. 3 provides a potentially more accurate result (particularly for low  $m$ ),

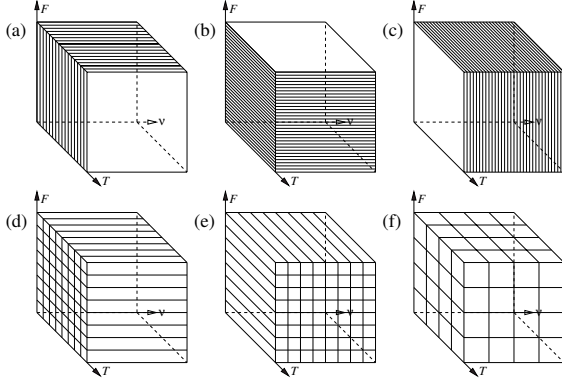
$$\hat{m}_1 = 2 \left( \frac{Y_{k_1}^* Y_{k_1} - N\theta^2}{Y_0^* Y_0 - N\theta^2} \right)^{\frac{1}{2}}. \quad (5)$$

For a maximum likelihood estimate (MLE), we maximize the likelihoods of the measurements  $p(|Y_0|)$  and  $p(|Y_{k_1}|)$ :

$$\Lambda = \frac{1}{\pi^{\frac{3}{2}} \theta^3} \exp - \frac{1}{2\theta^2} \left[ \frac{(\Re\{Y_0\} - \kappa\sigma)^2}{2} + (\Re\{Y_{k_1}\}) \right]$$



**Fig. 1.** Processes: (top) magnitude modulation spectrum  $Y_k(T)$  and (bottom) time-frequency-modulation representation  $\Psi_{\nu,F}(T)$ .



**Fig. 2.** Time-frequency-modulation representations: (a) time series, (b) frequency spectrum, (c) modulation spectrum, (d) wide-band spectrogram, (e) frequency and modulation [8], and (f) our transformation combining STFT with MMS.

$$-\frac{m_1}{2} \kappa \sigma \cos \phi)^2 + \left( \Im \{Y_{k_1}\} - \frac{m_1}{2} \kappa \sigma \sin(\phi \operatorname{sgn} k) \right)^2 \Big]$$

where  $\kappa$  is a constant. Taking the logarithm, removing constants, and setting the exponent's derivative to zero,

$$\frac{\partial}{\partial m_1} \left[ \frac{1}{2} (Y_0 - \kappa \sigma)^2 + |Y_{k_1}|^2 - m_1 |Y_{k_1}| \kappa \sigma + \frac{m_1^2}{4} (\kappa \sigma)^2 \right] = -|Y_{k_1}| \kappa \sigma + \frac{m_1}{2} (\kappa \sigma)^2, \quad (6)$$

yields a 1st-order MLE of modulation index:  $\hat{m}_1 = 2|Y_{k_1}|/|Y_0|$ , as in eq. 4, since  $|Y_0| = (\kappa \sigma)^2$ .

**Effect of windowing.** Windowing convolves the modulation spectrum with the window's Fourier transform. So, for a Hann window, the mean is halved and second-order statistics (moment and variance) are three eighths. Yet, provided the window contains sufficient periods, the peak at  $f_1$  can be resolved from that at dc. The presence of multiple modulating frequencies requires they be sufficiently separated too, to be resolved. Thus, if  $m_1$  can be estimated, so can indices  $m_h$  for harmonics  $f_h$ , unlike the squared-signal spectrum which induces distortion at harmonics of the modulation frequency.

Although windowing reduces the effects of changing signal parameters on the analysis and allows different values between analysis frames, it does not deal with variation in  $f_h$  within the analysis window. As already mentioned, if  $N$  does not contain a whole number of modulation periods,  $k_h$

falls inconveniently between spectral bins. Frequency variation further broadens the peak around  $k_h$  which we address by summing over a wider range. The latter problem is resolved by calculating the MMS at more frequencies, which zero padding the analysis frame achieves. Thus, numbers of coefficients are increased in both numerator and denominator of eqs. 4 and 5 (otherwise unchanged). In practice, ranges  $\tilde{k}_h$  and  $\tilde{0}$  are set to include all bins above the noise floor from the spectral peak. An overview of the process for computing the MMS from an input signal is in Figure 1(top).

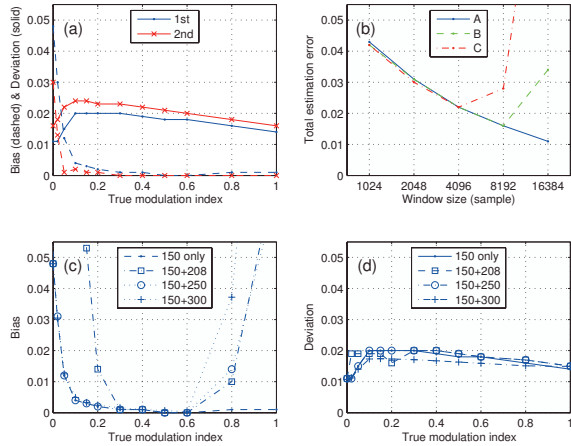
## 2.2. Time-frequency-modulation representations

In [8], the square and Fourier transform was applied to time-frequency elements  $\xi_{T,F}$  (e.g., from STFT or Gabor representation) to produce a mixed representation of broad acoustic frequency bands versus narrow modulation frequency information. Here, the magnitude operation on  $x_n$  that is used in the MMS can be applied to  $\xi_{T,F}$ , as in Fig. 1(bottom). Figure 2 depicts division of time ( $T$ ), frequency ( $F$ ) and modulation ( $f$ ) space, with 1D and 2D representations for comparison. For example, taking a 16 kHz signal, we first compute a wideband spectrogram using  $N = 32$  bins. Then, grouping magnitude spectra for 64 frames, we compute the modulation index  $\Psi_{\nu,F}$  within each band. Thus, we obtain a measure of modulation frequencies with 8 Hz resolution and  $\Psi_{\nu,F}$  for each 0.5 kHz band; time resolution of  $\Psi_{\nu,F}(T)$  is 128 ms.

## 3. SIMULATIONS TO EVALUATE ESTIMATES

**Method and results.** To see how well the two estimators  $\hat{m}$  and  $\hat{m}_1$  work, simulation tests were conducted on synthetic signals with known modulation levels. A hundred files of Gaussian white noise of 1 s duration were generated at a standard audio sample rate ( $f_S = 44.1$  kHz). Twelve files were produced from each with modulation indices ranging 0–1. The default modulation frequency was  $\bar{f}_1 = 150$  Hz, but descending  $f_1$  glides and jitter were included as typical perturbations [9]. For samples in each period  $i$ , beginning at time  $0 \leq t_i \leq 1$  s, phase was incremented by  $(f_{\max} - f_{\min} - f_{\min}) t_i \times (1 + r_i \sigma_T \sqrt{\pi}) / 200 f_S$ , where  $f_1$  went from  $f_{\max}$  to  $f_{\min}$ , the jitter percentage was  $\sigma_T$ , and  $r_i \sim \mathcal{N}(0, 1)$  was a zero-mean, unit-variance Gaussian random variable.

A battery of SAM estimators, including the first and second order MMS estimators  $\hat{m}_1$  and  $\hat{m}_1$ , and other variants based on the power spectrum, were evaluated with various window sizes  $N$  and three test conditions ( $f_{\max} - f_{\min}, \sigma_T$ ): A 0 Hz, 0%; B 20 Hz, 0.5%; C 60 Hz, 1.5%. Fig. 3(a) shows the set A results for both  $\hat{m}_1$  and  $\hat{m}_1$  ( $N = 4096$ ) in terms of the imposed modulation index  $m_1$ . The bias errors are much smaller than the standard deviation of the estimates, except at very low values of  $m_1 < 0.1$ , where estimates remain non-negative despite their variability, leading to larger bias and approximately half the deviation. We see that, al-



**Fig. 3.** Simulation results. (a)  $\hat{m}_1$  and  $\tilde{m}_1$  estimation errors vs. modulation index as average bias and standard deviation over 100 files:  $f_0 = 150$  Hz, 0% jitter,  $N = 4096$ ,  $8 \times$  zero padding. (b) Bias and deviation of estimation errors, averaged over all frames and  $m_1$  (1200 files), vs. window size  $N$  with  $8 \times$  zero padding: 1024 (23 ms), 2048 (46 ms), 4096 (93 ms), 8192 (186 ms), 16384 (372 ms). Settings for  $f_1$  and jitter: A 150 Hz, 0.0%; B 160–140 Hz, 0.5%; C 180–120 Hz, 1.5%. (c,d) Bias and deviation as in (a) for  $\hat{m}_1$  with two modulating frequencies  $f_1 = 150$  Hz and  $f_2 = \{208, 250, 300\}$  Hz.

though  $\tilde{m}$  gives negligible bias, its variance is higher, which results in poorer estimates overall. This trend was consistent across all our tests. Equally, none of the other estimators performed better than  $\hat{m}_1$  and  $\tilde{m}_1$ . Fig. 3(b) gives results, averaged across the 12  $m_1$  values, for  $\hat{m}_1$  with sets A, B and C at various  $N$ . As expected, the overall estimation error increases when fewer periods are in the analysis frame. When the modulating frequency is perfectly periodic (A), the performance improves exponentially with increasing  $N$ ; however, the decreasing tolerance to any frequency perturbation (B, and especially C) produces problems with big  $N$ , at least with our peak summation over  $k_1$ . Figs. 3(c,d) plot the bias and deviation with two modulating frequencies  $f_1 = 150$  Hz (as before) and  $f_2 \in \{208, 250, 300\}$  Hz with  $m_2 = 0.3$ . The single frequency result for  $\hat{m}_1$  is re-plotted for reference. In general, the variances are not substantially altered, though  $f_2 = 300$  Hz exhibits a possible improvement, but there are two situations that produce large bias errors: (i) the small separation between  $f_1$  and  $f_2 = 208$  Hz, (ii) the constraint  $|a_n| \geq 0$  is broken (i.e., for  $m_1 \geq 0.7$ ). Otherwise, the results clearly indicate the independence of  $\hat{m}_1$  and  $\hat{m}_2$ , even when  $f_1$  and  $f_2$  are harmonically related.

**Discussion.** A significant drawback with power-based approaches is that they use a squaring operation, which is definitively non-linear, distorting the modulating function  $a_n$ . For example, SAM at  $f_1$  induces spectral components of power

$z_n$  at  $2f_1$ , making it hard to separate effects of higher-harmonic components in  $a_n$ . Therefore, an important advantage of the MMS is its linearity in respect of  $a_n$ . From a computational perspective, however, the power modulation spectrum  $Z_k$  is preferable to  $Y_k$ , which requires a square root. Yet there are other issues: the MLE formulation from  $Z_k$  requires numerical solution of a cubic for one harmonic, and it becomes intractable for multiple harmonics due to overlapping spikes. The MLE from  $Y_k$  gives a first-order modulation estimator that is biased; the second-order estimator reduces the effect of bias but at the cost of a wider variance. Windowing and zero-padding make these MMS estimates robust to variations over time and to non-integer modulation periods within the analysis frame.

#### 4. SUMMARY AND CONCLUSIONS

The MMS was proposed for estimating the modulation index for noise with AM from a periodic signal. The method was developed for application to real signals with time-varying parameters (noise variance, frequency of periodicity, modulation depth). Using synthetic test signals, we evaluated two estimators and showed that, in contrast to power-based estimators, MMS enabled separate estimation of modulation index for each modulating frequency. Simulation results gave low error on modulation index estimates (standard deviation of 0.04 for a  $2^{10}/23$  ms window, 0.02 for  $2^{12}/93$  ms). Further work will investigate theoretical bounds of  $\hat{m}$ , and its application to real stochastic signals.

#### 5. REFERENCES

- [1] P. J. B. Jackson and C. H. Shadle, "Frication noise modulated by voicing, as revealed by pitch-scaled decomposition," *JASA*, vol. 108, no. 4, pp. 1421–1434, 2000.
- [2] J. A. Betts, *Signal processing, modulation and noise*, Sci. & tech. series. English Univ. Press, London, 1970.
- [3] M. Schwartz, *Information transmission, modulation, and noise*, McGraw-Hill, London, 4th ed., 1990.
- [4] M. D. Srinath et al., *An introduction to statistical signal processing with applications*, Info. sys. sci. series. Prentice Hall, Englewood Cliffs, 1996.
- [5] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*, Academic Press, London, 2nd ed., 1995.
- [6] H. L. Van Trees, *Detection, estimation, and modulation theory*, vol. 3: Radar-sonar signal processing and Gaussian signals in noise, Wiley, New York, 2001.
- [7] A.C. Bovik et al., "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE T-Sig. Proc.*, vol. 41, no. 12, pp. 3245–3265, 1993.
- [8] L. Atlas and S. A. Shamma, "Joint acoustic and modulation frequency," *EURASIP J. Appl. Sig. Proc.*, vol. 7, pp. 668–675, 2003.
- [9] D. Michaelis et al., "Glottal-to-noise excitation ratio — a new measure for describing pathological voices," *Acta Acustica*, vol. 81, pp. 700–706, 1995.