# HILBERT SPECTRAL ANALYSIS OF VOWELS USING INTRINSIC MODE FUNCTIONS

*Steven Sandoval*

Arizona State University
School of Elect., Comp.
and Energy Eng.
Tempe, AZ, U.S.A.
spsandov@asu.edu

*Phillip L. De Leon*

New Mexico State University
Klipsch School of Elect.
and Comp. Eng.
Las Cruces, NM, U.S.A.
pdeleon@nmsu.edu

*Julie M. Liss*

Arizona State University
Department of Speech
and Hearing
Tempe, AZ, U.S.A.
jmliss@asu.edu

## ABSTRACT

In recent work, we presented mathematical theory and algorithms for time-frequency analysis of non-stationary signals. In that work, we generalized the definition of the Hilbert spectrum by using a superposition of complex AM–FM components parameterized by the Instantaneous Amplitude (IA) and Instantaneous Frequency (IF). Using our Hilbert Spectral Analysis (HSA) approach, the IA and IF estimates can be far more accurate at revealing underlying signal structure than prior approaches to time-frequency analysis. In this paper, we have applied HSA to speech and compared to both narrowband and wideband spectrograms. We demonstrate how the AM–FM components, assumed to be intrinsic mode functions, align well with the energy concentrations of the spectrograms and highlight fine structure present in the Hilbert spectrum. As an example, we show never before seen intra-glottal pulse phenomena that are not readily apparent in other analyses. Such fine-scale analyses may have application in speech-based medical diagnosis and automatic speech recognition (ASR) for pathological speakers.

***Index Terms*—**
Hilbert Space, Signal Analysis, Speech Analysis

## 1. INTRODUCTION

The short-time speech spectrum is the de facto analysis tool used in nearly all areas of speech analysis and applications [1, 2]. The spectrogram is a visualization of the energy structure of a signal in the coordinates of time and frequency obtained from the Short-Time Fourier Transform (STFT) [3]. The spectrogram can display a great deal of information about the properties of the speech utterance, including fundamental and formant frequencies [4].

We have recently proposed Hilbert Spectral Analysis (HSA) as a generalized time-frequency analysis framework consisting of a superposition of complex AM–FM components [5]. We have also proposed a novel 3-D visualization of the Hilbert spectrum, and a numerical method for performing HSA based on a modified version of Empirical Mode Decomposition (EMD) that utilizes Intrinsic Mode Functions (IMFs). By using HSA, we gain a degree of freedom in our analysis that may be more useful in describing the underlying physical phenomena. Although the STFT has been widely successful for many speech applications such as automatic speech recognition (ASR), coding, and speaker recognition (SR), other applications such as speech-based medical diagnosis and ASR for pathological speakers may require a more sensitive analysis, such as HSA, before finding practical use.

The contributions of the paper are as follows. First, we compare and contrast the Hilbert speech spectrum to both narrowband and wideband spectrograms for an example vowel in order to illustrate advantages of HSA. HSA components often align well with the energy concentrations in the wideband spectrogram but are not constrained by the inherent structural assumptions in the STFT. Utilizing the Instantaneous Amplitude (IA)/Instantaneous Frequency (IF) parameterization of the AM–FM components, we propose a novel method for formant estimation. Second, we illustrate the fine structure in the intra-glottal pulse revealed by the Hilbert spectrum that does not appear in spectrograms. Third, we argue that this fine structure obtainable in HSA can provide new insights in speech production models. For example, in both HSA and STFT we can compute the average fundamental frequency $f_0$, but with HSA we may quantify variations in $f_0$ more accurately.

This paper is organized as follows. In Section 2, we briefly review traditional speech analysis based on the spectrogram. In Section 3, we provide a summary of HSA theory and the HSA–IMF algorithm to numerically compute the IA/IF parameterization of the Hilbert spectrum. In Section 4, we describe the 2-D and 3-D visualizations of the Hilbert speech spectrum. Using the Hilbert spectrum, we propose a novel formant estimation technique and discuss the fine spectral structure that is present. Finally, in Section 5 we provide conclusions and future research directions for this work.

## 2. SPECTROGRAPHIC ANALYSIS OF SPEECH

The spectrogram is by far the most widely-used speech analysis tool and presents the structure of a signal's energy in time and frequency [6]. One of the parameters in the spectrogram is the window length, which controls the frequency band structure and leads to a well-known tradeoff between narrowband and wideband spectrograms. Each of these spectrogram types has its uses in speech analysis. In wideband spectrograms $f_0$ can be determined from the spectrogram by counting the number of individual vertical lines per unit time. Also, the frequencies and relative strengths of the first two formants, $F_1$ and $F_2$, are visible as dark, blurry concentrations of energy. The wide bandwidth in this type of analysis allows for excellent time resolution—the energy peaks from each individual vibration of the vocal folds are visible in the spectrogram. However, poor frequency resolution limits the ability to pick out individual harmonics. The narrowband spectrogram is the complement to the wideband spectrogram where one is able to pick out individual harmonics. However, time resolution may not be good enough to isolate each individual cycle of vibration, and the formant structure is not rendered as clearly as with a wideband analysis [7].

We first note that throughout this paper, we utilize a perceptually-motivated colormap in order to improve interpretation over other colormaps [8, 9]. For the narrowband spectrogram, we used a length 2048 Hamming window and for the wideband spectrogram we used a length 256 Hamming window; for both spectrograms we advanced the window by one sample in order to provide the most comparable representation to the Hilbert spectrum, despite the redundancy of such a large window overlap. Figures 1(a) and (b) show the narrowband and wideband spectrograms[1], respectively of the vowel /ɝ/ in an /hVd/ context, spoken by the first author of this paper, zoomed in on the vowel portion.

With a long window, the spectral harmonicity is better captured, and results in harmonic amplitudes that better reflect the underlying vocal tract spectral envelope [10]. Thus from the narrowband spectrogram in Figure 1(a), we visually estimate $f_0 = 135$ Hz by noting the frequency of the first harmonic. The formants are estimated as $F_1 = 385$ Hz and $F_2 = 1275$ Hz by noting the frequency associated with the strongest harmonic amplitudes.

With a short window, the spectral harmonicity is blurred and the harmonic amplitudes are degraded, but changes in the harmonicity and the spectral envelope are better captured [10]. Thus from the wideband spectrogram in Figure 1(b), we estimate $f_0 = 126$ Hz by noting 11 glottal cycles over a 87 ms timespan. The formants are visually estimated as $F_1 = 470$ Hz and $F_2 = 1400$ Hz by noting the center of the energy concentrations.

[1]In a strict sense the spectrogram plots the magnitude-squared of the STFT. In this paper, we plot the STFT magnitude in order to facilitate comparisons to the Hilbert spectrum.

Figure 2(a) shows the vowel waveform $x(t)$ of the example vowel /ɝ/ and Figure 2(b) shows the ten dominant Simple Harmonic Components (SHCs)[2] resulting from Fourier analysis of the waveform.

## 3. HSA THEORY AND HSA–IMF ALGORITHM

In this section, we summarize the key points of HSA and for additional details, encourage the reader to see [5]. We assume a real observation $x(t)$ of a complex "latent signal" $z(t)$ which are related by

$$x(t) = \Re\{z(t)\}. \tag{1}$$

In HSA, we decompose the latent signal into complex AM–FM components,

$$z(t) \equiv \sum_{k=0}^{K-1} \psi_k\left(t; a_k(t), \omega_k(t), \phi_k\right) \tag{2}$$

and the AM–FM component is defined as

$$\psi_k(t; a_k(t), \omega_k(t), \phi_k) \equiv a_k(t) \exp\left\{ j \left[ \int_{-\infty}^{t} \omega_k(\tau) d\tau + \phi_k \right] \right\} \tag{3a}$$

$$= a_k(t) e^{j\theta_k(t)} \tag{3b}$$

$$= s_k(t) + j\sigma_k(t) \tag{3c}$$

parameterized by the IA $a_k(t)$, IF $\omega_k(t)$, and phase reference $\phi_k$. The component can also be represented in terms of phase $\theta(t)$ as in (3b) or the real part $s_k(t)$ and imaginary part $\sigma_k(t)$ as in (3c).

As a note to the reader, HSA as developed in [5] relaxes the overly-constrictive assumption of *harmonic correspondence* resulting in a completely new formulation of AM–FM modeling unlike previous AM–FM models. Previous AM–FM models for signal analysis/synthesis usually fall into one of three main groups: 1) Hilbert Transform (HT) [13, 14, 15, 16], 2) peak tracking/sinusoidal modeling [17, 18, 19, 20], and 3) Teager energy operator [21, 22, 23, 24, 25, 26]. However, some models exist that do not fall into any of these groups [27, 23]. A historical summary of AM–FM modeling is presented by Gianfelici [28].

In [29], Huang proposed the original EMD algorithm that sequentially determines a set of IMFs, which are in fact AM–FM components, via an iterative sifting algorithm. The Ensemble Empirical Mode Decomposition (EEMD) [30] and tone masking [31] introduced ensemble averaging in order to address the mode mixing problem. The complete EEMD was proposed to address some of the undesirable features of

[2]The term SHC refers to the complex exponential with fixed amplitude and frequency that is a solution to the differential equation for simple harmonic motion [11, 12].

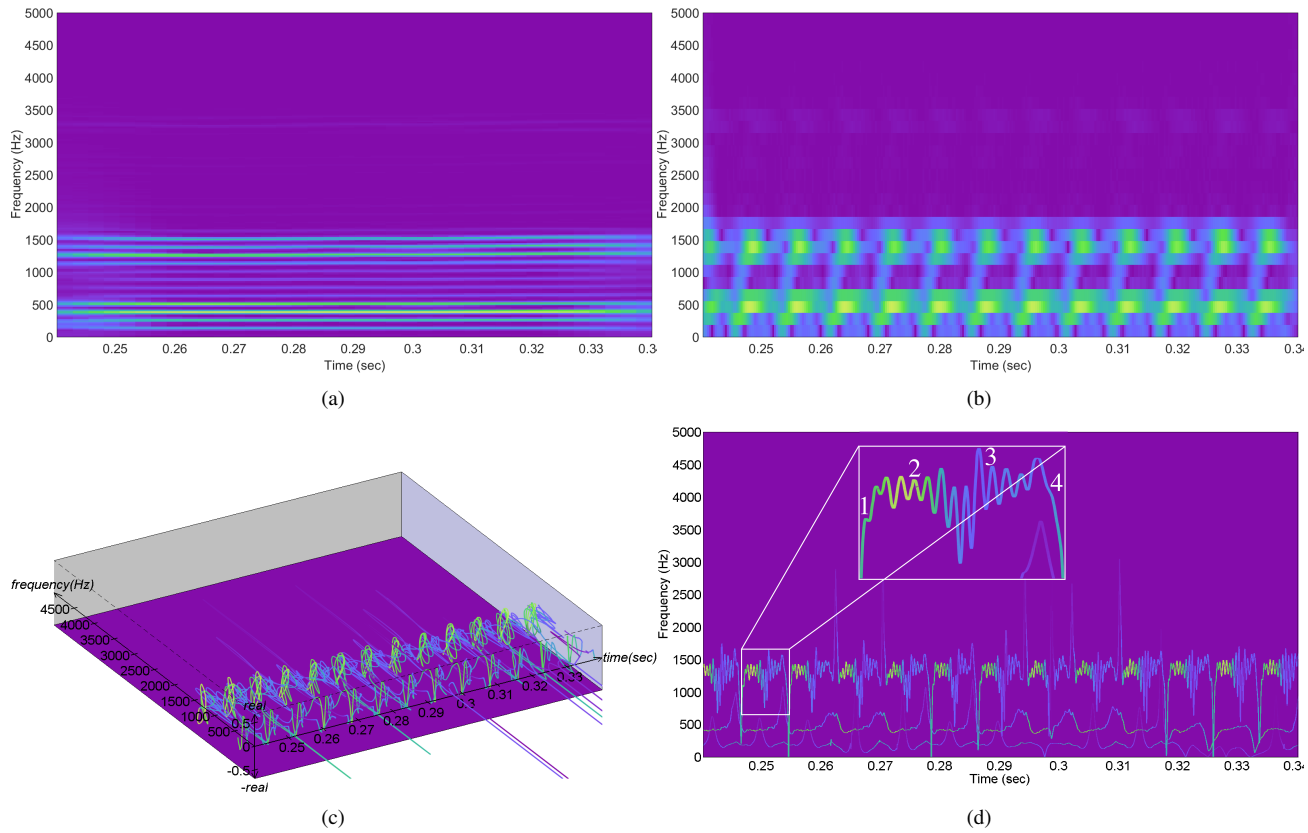


**Fig. 1**. Spectral analysis of the vowel /ɝ/ taken at the midpoint of 'herd': (a) Narrowband spectrogram, (b) wideband spectrogram, (c) 3-D Hilbert spectrum (real part of component vs. frequency vs. time), and (d) orthographic projection of the 3-D Hilbert spectrum onto 2-D (frequency vs. time). Plot line color indicates short-time magnitude in the spectrograms and instantaneous amplitude in the Hilbert spectra. The 2-D Hilbert spectrum shows fine spectral structure not available in the Fourier spectra. Note that the spectrograms are shown with linear color scaling, rather than logarithmic color scaling typically used in speech analysis, to better facilitate comparison to the Hilbert spectrum.

EEMD by averaging at the component-level as each component is estimated rather than averaging at the conclusion of EMD [32]. Several improvements to the sifting algorithm have also been proposed including those by Rato [33].

In [5], we presented a numerical algorithm, by combining the most desirable features of complete EEMD, tone masking, and Rato's improvements to the sifting algorithm, for computing the Hilbert spectrum under the assumption that the AM–FM components are IMFs [29]. Unlike previous studies, close attention is paid to the assumptions made in the definition of the IMF which are carried forward to the demodulation step, where the IA and IF parameters are estimated. In [5], we proposed a mathematically equivalent method to obtain the IF that is more numerically stable than Huang's [34] and leverages Rato's IA estimation technique [33]. We incorporate the proposed demodulation and our numerical algorithm into a single HSA–IMF algorithm which gives very good estimates for the IA and IF parameters of the AM–FM model. Finally, for the interested reader, we have posted online MATLAB scripts for HSA–IMF and HSA visualization at [35].

The effects of sampling in the context of EMD have been considered by Rilling and it is generally recommended to oversample but not resample before application of EMD, so that EMD effectively behaves like a continuous operator [36]. For this reason, the speech recordings used in this work were made in a sound booth using a high-quality microphone and a sampling rate of 44.1 kHz.

In prior work, we used filtered white Gaussian noise as the masking signal [5]. While this provides a simple method for masking signal design given no other information about the latent signal, it may not be optimal once we know the latent signal consists of speech. We have found that for speech, the use of a high-frequency, high-amplitude tone in the first two iterations of sifting can result in more stable performance than using noise. Other parameters used in HSA–IMF in this paper include: scale factor for mean envelope removal $\alpha = 0.95$, stopping threshold of 27 dB for the sifting algorithm, number of sifting iterations $I = 15$, stopping threshold of 8 dB for

HSA–IMF termination, scale factor for the additive masking signal $\beta = 0.5$, and range parameter $L = 3$ used in demodulation.

As a final note, we point out that the assumption with traditional Fourier analysis is an infinite superposition of harmonics which is almost certainly not representative of the underlying physics in speech production. On the other hand, even though IMFs may also not represent the true underlying components for speech, they can prove useful for may problems just as with the Fourier spectrum.

## 4. SPEECH ANALYSIS USING THE HILBERT SPECTRUM

### 4.1. Visualization of the Hilbert Spectrum

By plotting $\omega_k(t)$ vs. $s_k(t)$ vs. $t$ as a line in a 3-D space and coloring the line with respect to $|a_k(t)|$ for each component, the simultaneous visualization of multiple parameters for each component is possible. Further, orthographic projections yield common plots: the time-real plane (the real signal waveform), the time-frequency plane (2-D Hilbert spectrum), and the real-frequency plane (analogous to the Fourier magnitude spectrum).

### 4.2. Hilbert Spectrum of Vowel /ɝ/

Figure 1(c) shows the 3-D visualization of the Hilbert spectrum for the vowel /ɝ/ and Figure 1(d) shows the orthographic projection onto the time-frequency plane. Color variation in the plot line indicates the IA of the component at that time, i.e. the magnitude of the component. The value of the plot line along the frequency axis indicates the IF of the component at that time, i.e. the instantaneous angular velocity of the component. In the 3-D plot, displacement along the vertical axis shows the real part of the components $s_k(t)$. The superposition of $s_k(t)$ yields the speech signal $x(t)$ which can easily be seen by substituting (3c) into (2) and the result into (1).

The IA/IF parameterizaton of the components provides an alternate and very simple method of estimating a formant frequency $F$, via an IA-weighted average of the IF [37]

$$F = \frac{\int \omega_k(t) a_k(t) dt}{\int a_k(t) dt}. \tag{4}$$

For the example given, this method yields $F_1 = 431$ Hz and $F_2 = 1314$ Hz. With the spectrogram a weighted average technique for formant estimation is in theory possible, though it is not nearly as convenient or simple as (4). Thus HSA of speech provides a unique method for automatic formant estimation.

Figure 2(c) shows the real part of three AM-FM components resulting from the HSA. The components in red, green, and blue are associated with the voice bar, $F_1$, and $F_2$, respectively. The superposition of the components yields the original waveform shown in Figure 2(a)

$$x(t) = \Re \left\{ \sum_k s_k(t) \right\}. \tag{5}$$

### 4.3. Spectral Fine Structure

We believe the real advantage of HSA of speech signals lies in the ability to analyze and quantify fine spectral structure that exists in speech. In our example, this fine structure is most apparent in the upper component or $F_2$ where this detail is lost in the spectrogram regardless of the window length chosen. For the upper component, four regions in a single glottal cycle are labeled in the call out shown in Figure 1(d). In region 1, the component's IF rapidly approaches the weighted average IF with the IA approaching peak intensity for the cycle. Region 2 corresponds to the area of the glottal pulse with strongest energy concentration. In this region, the IF deviates about 100 Hz from the weighted average IF. Region 3 is described by rapid energy decay while IF deviation increases to $\sim$ 650 Hz deviation. Finally, region 4 exhibits a very large IF deviation with increasing IA prior to the start of the next glottal pulse.

### 4.4. Example Hilbert Spectra for Other Vowels

We have performed HSA for the following twelve vowels and three diphthongs in /hVd/ context: heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd, hud, herd, hoyed, hide, and how'd [1, 2]. This analysis includes the /hVd/ utterances from a female speaker and two male speakers. The resulting Hilbert spectral plots and spectrograms are collected into contact sheets to facilitate comparison and can be found online at [38]. In the online Hilbert spectral plots, we have used a Savitzky-Golay filter to smooth the IF while preserving the fine structure necessary for speech analysis [39, 40, 41]. We used one of two Savitzky-Golay filters depending on the level of smoothing desired. The filter parameters are order $k = 1$ and frame length $f = 255$ for aggressive smoothing and $k = 9$ and $f = 65$ for reserved smoothing.

## 5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have computed and visualized the Hilbert spectrum of speech using our recently proposed HSA–IMF algorithm. We compare the Hilbert spectrum of an example vowel to that of the narrowband and wideband spectrograms to illustrate the advantages of using HSA. One of the advantages is revealing spectral fine structure on small time-scales such as within a single glottal pulse, which may not be apparent in the spectrogram. We also leveraged the IA/IF parameterization of the AM–FM components to provide a simple formula to compute formant frequencies. Although the

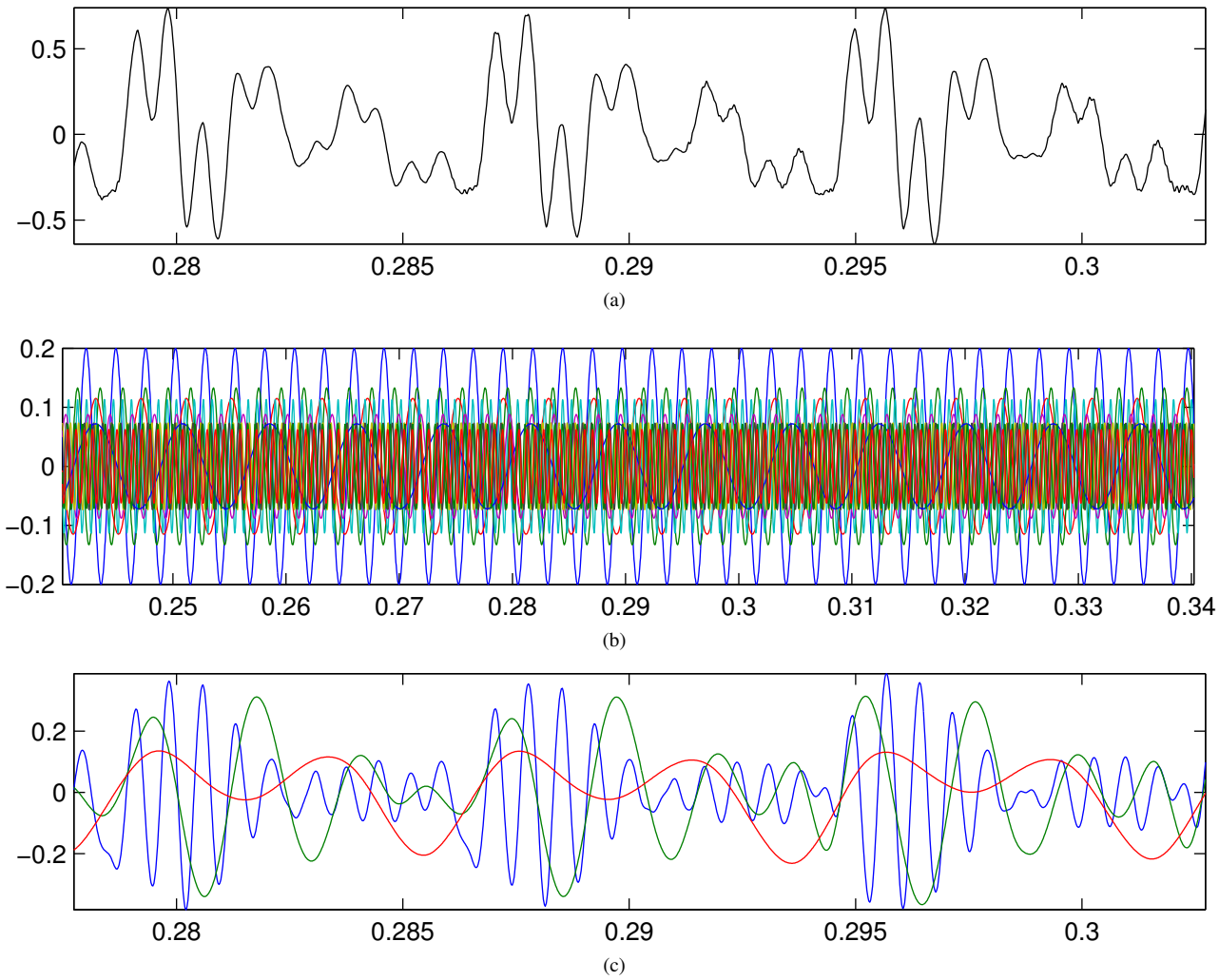


**Fig. 2**. (a) The waveform $x(t)$ associated with the vowel /ɝ/ at the midpoint of 'herd', (b) the ten dominant harmonics from the Fourier transform of $x(t)$, and (c) the real part of the three AM-FM components $s_k(t)$ comprising $x(t)$. The components in red, green, and blue are associated with the voice bar, $F_1$, and $F_2$, respectively.

HSA–IMF algorithm is iterative and requires more computation than the FFT used for spectrographic analysis, Hilbert spectra of speech sounds may be computed in a few seconds on an ordinary PC.

We believe there is potential in utilizing the spectral fine structure obtained through HSA for evaluating aspects of speech that have traditionally been difficult such as evaluation of vocal quality. For example, measures similar to jitter and shimmer, which have have proven useful in the detection of vocal tremor and vocal flutter, may be accessible from the fine-grained analysis obtainable though HSA. Finally, we are currently investigating the efficacy of features extracted from the Hilbert spectrum for classification of dysarthic speech with the goal of providing new methods for speech-based medical diagnosis and monitoring.

## 6. REFERENCES


[1] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, Mar. 1952.

[2] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3099–3111, 1995.

[3] J. B. Allen and L. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proc. IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977.

[4] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.

[5] S. Sandoval and P. L. De Leon, "Theory of the hilbert spectrum," *arXiv*, Apr. 2015, math.cv/1504.07554.

[6] D. O'Shaughnessy, *Speech Communications: Human and Machine*, Addison-Wesley, 1987.

[7] "National Center for Voice & Speech," `http://www.ncvs.org/ncvs/tutorials/voiceprod/tutorial/spectral.html`, 2015.

[8] D. Borland and R. M. Taylor II, "Rainbow color map (still) considered harmful.," *IEEE Trans. Visual. Comput. Graphics*, vol. 27, no. 2, pp. 14–17, Mar. 2007.

[9] M. Niccoli and S. Lynch, "A more perceptual color palette for structure maps," in *Proc. GeoConvention*, May 2012.

[10] T. .F. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2002.

[11] R. Shankar, *Fundamentals of Physics: Mechanics, Relativity and Thermodynamics*, Yale University Press, 2014.

[12] L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics*, Wiley Publishing, 3 edition, 1982.

[13] M. Feldman, "Non-linear system vibration analysis using Hilbert transform–I. free vibration analysis method 'FREEVIB'," *Mechanical Syst. and Signal Processing*, vol. 8, no. 2, pp. 119–127, Mar. 1994.

[14] A. Rao and R. Kumaresan, "On decomposing speech into modulated components," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 240–254, May 2000.

[15] F. Gianfelici, G. Biagetti, P. Crippa, and C. Turchetti, "Multicomponent AM–FM representations: an asymptotically exact approach," *IEEE Trans. Audio Speech Lang. Processing*, vol. 15, no. 3, pp. 823–837, Mar. 2007.

[16] M. Feldman, *Hilbert Transform Applications in Mechanical Vibration*, Wiley, 2011.

[17] R. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[18] P. Rao and F. J. Taylor, "Estimation of instantaneous frequency using the discrete Wigner distribution," *Electron. Lett.*, vol. 26, no. 4, pp. 246–248, Feb. 1990.

[19] Y. Pantazis, O. Rosec, and Y. Stylianou, "Adaptive AM–FM signal decomposition with application to speech analysis," *IEEE Trans. Audio Speech Lang. Processing*, vol. 19, no. 2, pp. 290–300, Feb. 2011.

[20] B. Boashash, G. Azemi, and J. O'Toole, "Time-frequency processing of nonstationary signals: Advanced TFD design to aid diagnosis with highlights from medical applications," *IEEE Signal Processing Mag.*, vol. 30, no. 6, pp. 108–119, Nov. 2013.

[21] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *IEEE Trans. Signal Processing*, vol. 41, no. 10, pp. 3024–3051, Oct. 1993.

[22] A. C. Bovik, P. Maragos, and T. F. Quatieri, "AM–FM energy detection and separation in noise using multiband energy operators," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3245–3265, Dec. 1993.

[23] L. B. Fertig and J. H. McClellan, "Instantaneous frequency estimation using linear prediction with comparisons to the DESAs," *IEEE Signal Processing Lett.*, vol. 3, no. 2, pp. 54–56, Feb. 1996.

[24] A. Potamianos and P. Maragos, "Speech analysis and synthesis using an AM–FM modulation model," *Speech Commun.*, vol. 28, no. 3, pp. 195–209, Jul. 1999.

[25] A.-O. Boudraa, J.-C. Cexus, F. Salzenstein, and L. Guillon, "If estimation using empirical mode decomposition and nonlinear teager energy operator," in *Proc. IEEE Int. Symp. Control, Commun. and Signal Processing*, 2004, pp. 45–48.

[26] A.-O. Boudraa, "Instantaneous frequency estimation of fm signals by $\psi$b-energy operator," *Electron. Lett.*, vol. 47, no. 10, pp. 623–624, 2011.

[27] T. F. Quatieri, T. E. Hanna, and G. C. O'Leary, "AM–FM separation using auditory-motivated filters," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 5, pp. 465–480, Sep 1997.

[28] F. Gianfelici, C. Turchetti, and P. Crippa, "Multicomponent AM–FM demodulation: the state of the art after the development of the iterated Hilbert transform," in *Proc. Int. Conf. Signal Processing and Commun.*, Nov. 2007, pp. 1471–1474.

[29] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London Ser. A*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[30] Zhaohua Wu and Norden E Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[31] R. Deering and J. F. Kaiser, "The use of a masking signal to improve empirical mode decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2005, pp. 485–488.

[32] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2011, pp. 4144–4147.

[33] R. T. Rato, M. D. Ortigueira, and A. G. Batista, "On the HHT, its problems and some solutions," *Mechanical Syst. and Signal Processing*, vol. 22, no. 6, pp. 1374–1394, 2008.

[34] N. E. Huang, Z. Wu, S. R. Long, K. C. Arnold, X. Chen, and K. Blank, "On instantaneous frequency," *Advances in Adaptive Data Analysis*, vol. 1, no. 02, pp. 177–229, 2009.

[35] "Hilbert Spectral Analysis," `http://www.HilbertSpectrum.com`, 2015.

[36] G. Rilling and P. Flandrin, "On the influence of sampling on the empirical mode decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2006, pp. 444–447.

[37] T. Strom, "On amplitude-weighted instantaneous frequencies," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 4, pp. 351–353, 1977.

[38] "Hilbert Spectral Analysis of Speech," `http://asru2015.HilbertSpectrum.com`, 2015.

[39] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[40] S. J. Orfanidis, *Introduction to Signal Processing*, Prentice-Hall, 1995.

[41] R. W. Schafer, "What is a Savitzky-Golay filter?," *IEEE Signal Processing Mag.*, vol. 28, no. 4, pp. 111–117, 2011.