A novel signal processing approach to auditory phantom perception

I-Hui Hsieh¹ · Jia-Wei Liu¹

Published online: 31 July 2018
© Psychonomic Society, Inc. 2018

Abstract

The ear and brain interact in an orchestrated manner to create sensations of phantom tones that are audible to listeners despite lacking physical presence in original sounds. The relative contribution of peripheral sensory cell activity and cortical mechanisms to phantom hearing remains elusive. The current study addressed the question of whether non-linear components of a complex signal exist that are not captured by the linear combination of cosines in a series. To this end, we investigated the source and spectro-temporal dynamics of non-linear components within two-tone complexes related to phantom acoustic perception. The empirical mode decomposition, a method for non-linear and non-stationary processes, was applied to extract the extra-aural existence of an oscillatory component within the original signal associated with the phantom sound. This travelling wave (phantom) has never before been observed in the sound's linear spectrum. We showed that the wave travels at a velocity that accurately maps onto the perceived phantom tone frequency. Phase coherence of oscillatory mode dynamics predicted discrimination sensitivity to phantom sounds by listeners. Perceived incidences of phantom tones correlated with magnitude of the Hilbert power spectra of the extra-aural component. Findings suggest a possible origin of phantom sounds that exists within the original signal, with potential implications for current models of non-linear cochlear mechanics and cortical dynamics in generating phantom percepts.

Keywords Empirical mode decomposition · Phantom tone · Non-linear analysis · Cubic difference tone · Intrinsic mode functions

Introduction

Phantom perception is thought to arise from the intricate interplay between the auditory system and brain. Phenomena such as missing fundamental or combination tones all produce pitch percepts that are associated with phantom hearing. These sounds are characterized by lacking an apparent source or component of the perceived frequency in the original signal based on Fourier analysis showing an absence of spectral energy at the perceived pitch. For instance, cubic difference tone (CDT), one of the most prominent combination tones (Plomp, 1965; Smoorenburg, 1972), is produced by two sinusoidal signals with frequencies f_1 and f_2 ($f_1 < f_2$) such that when played simultaneously, these tones produce the perception of a third tone with a frequency of $2f_1 - f_2$ that is not physically

present in the stimulus (Capranica & Moffat, 1980; Moore, 2003). Nevertheless, listeners report hearing a faint sound corresponding to the “phantom” frequency even at relatively low intensity levels (Barral & Martin, 2012; Goldstein, 1967). Phantom tones are interesting auditory phenomena, as the acoustical properties of these tones display sharp dependence on the amplitude and frequency interaction of the two components of the stimulus, despite the absence of the tone within the stimulus itself (Barral & Martin, 2012; Rilling & Flandrin, 2008). The phenomenon has been applied in technological music synthesis (Kendall, Haworth, & Cádiz, 2014) and diagnosis of hearing problems (Johnsen, Bagi, & Elberling, 1983; Kemp, 1978; Kendall et al., 2014), and provides insightful information to the non-linear mechanics of pitch theory development (Barral & Martin, 2012; Capranica & Moffat, 1980). This perplexing psychophysical occurrence has brought about the development of many combination-tone theories, with most incorporating some form of inner ear non-linear transducers, sensory hair-cell activities (Barral & Martin, 2012; Egüíluz, Ospeck, Choe, Hudspeth, & Magnasco, 2000; van Dijk & Manley, 2001), and/or central cortical mechanisms (Ziębakowski, 2012). However, none of these theories has

✉ I-Hui Hsieh
ihuihsieh@gmail.com

¹ Institute of Cognitive Neuroscience, National Central University,
No.300 Zhongda Rd., Zhongli County, Taoyuan City 32001, Taiwan

reached satisfying conclusions (Ashmore, 2008; Cariani & Delgutte, 1996). One problem is that these theories do not explain how globally represented combination tones can be generated via locally vibrating inner ear components (Ziebakowski, 2012). Studies investigating otoacoustic emissions have shown sound emission in external ear canals of animals corresponding to CDT frequency; however, human ears differ significantly from these animals (Michalski, Bochnia, & Dziewiszek, 2011). Further, the origins of these non-linear cochlear mechanics remain mysterious as some have argued that the spectra corresponding to non-linear distortions of the inner ear are only observable in situations where the central auditory nervous system is involved (Robles et al., 1997). Thus, whether CDTs are the product of central mechanisms responsible for pitch extraction or the result of peripheral non-linearities remains elusive.

The theory of mechanical non-linear oscillations, in which outer hair-cell activities combine with cortical processing mechanisms, seems to offer a partial picture of why phantom tones can be artificially perceived, despite lacking physical presence in the acoustic signal. However, the fact that this perception has been termed a “phantom” tone has been typically attributed to lacking physical presence on the sound’s linear-based spectrum. Linear-based analysis, such as Fourier transform or wavelet transform, is a decomposition of a time-domain signal into a frequency domain by conversion into a series of sinusoidal orthogonal functions. While these methods are generally efficient for locally linear and stationary signals, they are not satisfactory when the signals contain fast-varying temporal fluctuations, which are often apparent in many naturally occurring sounds (Huang, Long, & Shen, 1996; Huang & Shen, 2014; Huang et al., 1998; Wu et al., 2008). The Fourier spectrum does not preserve the spectral evolution and transient temporal information found within the signals, and even with short-time Fourier transform or wavelet transform attempting to overcome this problem, the revised method is still limited by the time window of activation. As most acoustical signals in speech and music are non-stationary in nature (Sheng, 2010), using linear structures for analysis could, in fact, result in a loss of critical information to listeners (Huang & Shen, 2014; Huang et al., 1998; Sheng, 2010). Therefore, the current study raises the question of whether it remains a possibility that the “phantom” tone component is simply invisible to linearly-based analysis methods.

The present study used the non-linear adaptive signal analysis approach, known as the Hilbert-Huang Transform (HHT), to reveal the extra-aural presence of a spatial-temporal travelling wave not apparent in the original periodic sound signal. The hallmark of HHT is that it is a purely data-driven, empirically-based algorithm. The underlying mechanism is based on an iterative Empirical Mode Decomposition (EMD) method, operating through a spline-sifting process on the signal’s temporal envelope, to separate the signal into

a finite number of components called Intrinsic Mode Functions (IMFs) (Huang et al., 1998). Each resultant IMF represents oscillations of a data-driven, band-pass filtered version of the original signal envelope, which can experience both amplitude and frequency fluctuations (Huang et al., 1998). In this paper the term “travelling wave” denotes the oscillatory component of the IMF component extracted by the EMD method, and not the cochlear travelling wave.

The implementation of IMFs has provided the foundation of many real-world HHT applications (Huang & Shen, 2014), ranging from image processing (Chen, Su, Zhang, Tian, & Yang, 2008; Wu, Huang, & Chen, 2009) and atmospheric variations (Huang et al., 2009; Ruzmaikin & Feynman, 2009; Wu et al., 2008) to medical and epidemiological applications (Chen, 2009; Cummings et al., 2004; Liang, Bressler, Desimone, & Fries, 2005; Nunes, Guyot, & Deléchelle, 2005), all with great success (Huang et al., 1996; Huang & Shen, 2014; Huang et al., 1998). In the area of auditory processing, IMFs have been used to track the fundamental frequency of speech sound and to enhance speech quality (Dong, 2017; H. Huang & Pan, 2006; Huang, Chen, Lo, & Wu, 2011; Khaldi, Boudraa, & Turki, 2016; Khaldi, Turki-Hadj Alouane, & Boudraa, 2010; Tilsen & Arvaniti, 2013; Wu & Huang, 2009). Yet, IMFs have not been used to characterize human psychoacoustical performance in response to sound attributes. Since phantom tone percepts have largely been attributed to non-linear underlying mechanisms, we expect that using the EMD method may uncover non-linear components of a signal not captured by linear time series analysis.

The motivation of the current study was, first, to examine whether non-linear components of a complex signal exist that are not captured by the linear combination of cosines in a series using the EMD method. In other words, we address the fundamental question of what is the true nature of a complex acoustic signal from the time-frequency perspective, extending from the use of linear time series. Secondly, if such a non-linear frequency component exists within two-tone signals and can be identified, could we then use its spectro-temporal characteristics to predict human psychoacoustic response to this component? We observed a surprising extra-aural occurrence of travelling waves emanating from the original input signal using the EMD method. We showed that the spectro-temporal dynamics of the IMFs accurately predicted psychophysical response patterns in perceiving the tone percept corresponding to the IMFs.

Methods

Participants

Six normal-hearing adults (five females) with a mean age of 26.1 ($SD = 2.4$) years participated in this experiment. All

participants reported no history of audiological or neurological disease. No participants had prior experience in psychophysical experiments. Each participant practiced the various conditions of the experiment for half an hour prior to the beginning of data collection. Participants were paid for their participation. Written informed consent was obtained from each participant in accordance with the Institutional Review Board of National Taiwan University and National Central University, Taiwan. All experiment protocols were approved and performed in accordance with the guidelines and regulations of the Institutional Review Board of National Taiwan University and National Central University, Taiwan.

Signal analysis

We applied the HHT analysis method to analyze the time-frequency distribution of two-tone signals. The HHT analysis includes the application of EMD/EEMD method for signal decomposition and the Hilbert transform (HT) application to analyze the time-frequency distribution of decomposed components. The Hilbert-Huang Transformation Toolbox for MATLAB (Huang, 2001; Huang et al., 2011) was used for the EEMD analysis. The EEMD is the noise-assisted data analysis version of the EMD developed to solve the mode mixing problem (Wu & Huang, 2005, 2009) through the addition of white noise time series to the original signal before decomposition. First, the EEMD method was applied to the two-tone signals to extract a series of IMFs through a spline sifting process. To perform the EEMD analysis, two parameters were supplied, which were the ensemble size and the standard deviation of the white noise time series. Based on initial pilot tests to find the optimal parameters, the two parameters were chosen to be iteration number $E = 30$ and $\sigma_{\text{noise}} = 0.2$, where σ represents the ratio of the standard deviation of the noise to the signal. Note that for each of the 30 iterations, a different series of white noise was added. Second, we executed the IMF extraction iteration through a spline sifting process. An IMF was selected when the following two criteria were satisfied: (1) the number of extrema and zero crossings differ at most by unity, and (2) the mean of the local envelope defined by the maxima and minima must be zero at any point. The sifting process then continued until the IMF stopping criteria were met (Wu & Huang, 2005, 2009). The decomposition usually resulted in 8–10 IMFs for the two-tone signals used in the current experiment. The residue component was considered appropriate as all the IMFs passed the significance test against white noise. Then, HT was executed to display the time-frequency spectrum of the IMFs extracted from the two-tone signals using EEMD. Lastly, we computed the average instantaneous frequency (IF) for each IMF using HT with the number of data points selected to avoid EEMD end effect. A flowchart diagram of the EEMD signal analysis procedure is shown in Fig. 1.

Psychophysical stimulus and procedure

Stimuli were generated using MATLAB software (The MathWorks, Inc. v. 2015) on an ASUS PC (Dimension 8400), presented at a sampling rate of 44.1 kHz through 16-bit digital-to-analog converters (Creative Sound Blaster Audigy 2ZS), and through Sennheiser headphones (HD 380Pro), in a double-walled, steel, acoustically isolated chamber (interior dimensions 2.5 m (L) $\times$ 2.5 m (W) $\times$ 2.0 m (H); Industrial Acoustics Company) located at National Central University, Taiwan.

The stimuli for each trial comprised two-tone complexes, separated by a 500-ms interstimulus interval. Each two-tone complex was composed of the addition of two sine waves with a duration of 500 ms, with linear rise-decay ramps of 10 ms. The stimuli can be represented as follows:

$$X = a_1 \cos(2\pi f_1 t) + a_2 \cos(2\pi f_2 t) \quad (1)$$

where a_1 and a_2 represent the amplitude of the first and second sine-wave component, respectively; f_1 and f_2 correspond to the frequency of the first and second sine wave, and t represents time in seconds.

Beat-probe paradigm

We used a beat-probe paradigm to examine the strength of the CDT percept. A quiet probe tone, with a small frequency deviation from the frequency of the expected CDT percept (i.e., < 6 Hz), was added to the two-tone complex to induce a beat sensation (due to the small frequency difference between the two tones). Such a procedure has been suggested to test the perceptibility of CDT, which can be difficult to detect directly (Fastl & Zwicker, 2013). The rationale was as follows: if the subject could detect the beat effect generated from the addition of the probe tone that invokes a “beat” percept with the CDT frequency, then the CDT was detected. Otherwise, if the beat was not perceived, this indicated that the subject did not perceive the expected CDT tone that resulted from the interaction of the two-tone complex. The percentage of beat effect perceived was used as an index of the strength of expected CDT percept. Note that, here, the beat-probe was presented diotically via headphones to subjects. We did not utilize the dichotic binaural beat phenomenon, because in our pilot study, the presentation of a probe-tone to one ear and the two-tone complex to the other ear resulted in a less salient beat effect compared to diotic presentation. This was consistent with previous reports that binaural perception of combination tones was not observed psychophysically (Ziębakowski, 2012).

We first obtained the optimal beat-probe parameters to induce a beat effect. Six groups of frequency combinations of two-tone complex were chosen, based on a literature review,

Fig. 1 Flowchart diagram of the EEMD algorithm. The parameters were selected based on the present two-tone data

which suggested that these frequency combinations produced clearer CDT percepts (Bhagat & Champlin, 2004; Cooper & Rhode, 1997). The six groups of frequency combinations of two-tone complexes and their expected CDT frequency are listed in Table 1.

Table 1 Frequency combination of two-tone complex stimuli and their expected CDT component

Two-tone pairs	Frequency (Hz)			
	f_1	f_2	CDT	f_2/f_1
1	423	500	346	1.18
2	510	693.6	326.4	1.36
3	596	720	472	1.21
4	750	890	610	1.19
5	840	950	730	1.13
6	1636	2002	1270	1.22

The amplitudes of f_1 and f_2 were both fixed at 1 for this part of the study. Two beat-probe stimulus parameters were investigated: (1) the frequency difference between the beat-probe and the expected CDT tone (range 2–10 Hz, step size = 2 Hz), and (2) amplitude of the beat-probe (range 0.1–0.9, step size = 0.2). The ability to discriminate whether the standard or comparison two-tone complex contained a percept of “beat phenomenon” was examined in a two-interval, forced choice (2IFC) block design. The total number of conditions for the beat-probe condition was 6 frequency combinations $\times$ 5 beat-probe frequencies $\times$ 5 beat amplitude conditions = 150 conditions. In addition, we included 30 “catch-trials” (i.e., trials used to determine if subjects were actually perceiving a beat, or whether they were merely pressing buttons at random). The frequency of the catch-trial probes was set at ± 100 Hz deviation from the expected CDT frequency. Catch-trial probe tones were not expected to generate a perceptual beat with the CDT component, due to the large frequency difference between the probe and CDT. Participants were able to

accurately differentiate catch-trial probes from probes that generated beat sensations (Fig. 2a)

On each trial, a randomly chosen beat-probe was added to one of the two-tone complex intervals. Tones were presented diotically to participants through headphones. The participant's task was to identify which of two intervals contained the perception of a “beat” phenomenon by pressing the number “1” key if the first interval contained a beat or the number “2” key if the second interval contained a beat. Otherwise, participants were asked to press the number “3” key if they determined that neither interval contained a beat effect. No feedback was provided at any point in the experiment. Prior to the beginning of each block, 25 trials were randomly presented as practice. Each block contained 150 beat-probes and 30 catch-trial probes given a total of 180 trials. Each participant completed a total four of blocks, resulting in a total of 720 trials per participant (4 blocks $\times$ 180 trials/block = 720 trials). Mean percentage of perceived beat was averaged over a total of 3,600 beat-probe trials and 720 catch trials. Each probe amplitude $\times$ probe frequency data point was based on 144 samples (Fig. 2b).

The effect of parametrically varying the frequency and amplitude ratio of the two-tone complex on the perceived magnitude of the CDT component and spectro-temporal pattern of the travelling wave was also examined. The same six subjects who completed part 1 of the experiment participated in this part of the study. We selected three pairs of two-tone complexes, with frequency combinations of $f_1/f_2 = 1,363/2,002$, $1,401/2,002$, and $1,636/2,002$ Hz. The amplitude ratio of the lower-frequency to higher-frequency components of the two-tone complex was varied at 0.1, 0.3, 0.5, 0.7, and 0.9, equivalent to stimulus level differences between f_1 and f_2 of -20, -10,

-6, -3, and -0.9 dB SPL. A 10-ms linear ramp was applied diotically to both the onset and the offset portions of the two-tone complex envelope. We used a beat-probe with a frequency deviation from the expected CDT tone at 6 Hz (i.e., if the expected CDT frequency was 724 Hz, the probe tone frequency could be either $724 \pm 6 = 730$ or 718 Hz). The amplitude of the probe tone was fixed at 0.3, based on part 1 of the study suggesting that these were optimal probe parameters to induce salient beat effects (Fig. 2b). Catch-trial probes exhibited a frequency difference of ± 100 Hz deviation from the expected CDT frequency and a fixed amplitude of 0.3, to remain consistent with the actual beat-probe.

On each trial, the amplitude ratio of the two-tone complex was randomly selected from the set [0.1, 0.3, 0.5] for interval 1 and the set [0.5, 0.7, 0.9] for interval 2. The presentation order of the two intervals in a 2IFC procedure was randomized across trials. Note that a beat-probe was added to both intervals for this part of the study. The participant's task was to determine which interval contained the “stronger” beat sensation by pressing the number “1” or “2” keys to record their responses. Otherwise, participants were asked to press the number “3” key if the participant determined that neither intervals contained a beat effect. A schematic diagram of the beat-probe procedure is illustrated in Fig. 3. The total number of amplitude combinations was $3 \times 3 = 9$ groups with 40 trials per group (i.e., $9 \times 40 = 360$ trials). We further included 40 catch-trial probes. Therefore, each participant completed 400 trials per each of the three frequency combinations (including 360 beat-probe trials and 40 catch-trial probes) giving a total of 1,200 trials. Each participant completed two blocks of 1,200 trials each for a total of 2,400 trials. The experiment took approximately 2.5 h to complete.

Fig. 2 Perceived beat as a function of beat-probe amplitude and frequency combinations. **(a)** Percentage of perceived occurrence of beat phenomenon for beat-probe and catch-trial probe conditions. Mean perceived beat was averaged over 3,600 beat-probe trials and 720 catch

trials. Error bar represents ± 1 standard deviation. **(b)** Perceived beat percentage as a function of frequency difference between the beat-probe and CDT. Each probe amplitude $\times$ frequency data point was based on 144 samples. Different symbol lines represent different probe amplitudes

Fig. 3 Schematic diagram of the beat-probe procedure for phantom discrimination tasks. Each interval contained a 1,000-ms tone complex separated by a 500-ms interstimulus interval (ISI)

Results

The occurrence of a spatial-temporal travelling wave, corresponding to an audible phantom tone, was revealed for raw two-tone complex signals using EEMD (Fig. 4a). The occurrence indicated a travelling wave emanating from the original sound source devoid of the phantom tone components. In comparison, this travelling wave was not observed in conventional linear STFT spectrum (Fig. 4b) and Morlet wavelet (Fig. 4c) of the two-tone complex. Fourier spectrum was based on short-time Fourier Transform (STFT) with a 20-ms Hamming window with 50% overlap between segments. The frequency resolution was compatible between the different analysis methods.¹

The IMFs obtained by EEMD for one of the two-tone complexes are shown in Fig. 5. The oscillation pattern of the third most energetic IMF generated the average IF that corresponded to the perceived phantom-tone frequency (Fig. 5b,c). This travelling wave corresponding to the phantom component (Fig. 5c; red curve) was not revealed in Fourier

frequency spectra (Fig. 5c; blue curve). Statistical comparison of the phantom-tone travelling wave against white noise suggested a physically meaningful signal difference from noise above the 99% confidence level (Fig. 5d).

To examine the role of IMFs in how well it characterizes auditory perception of phantom tones, we compared listeners' perceived phantom saliency to IMF spectro-temporal patterns for different amplitude and frequency combinations of the original two-tone sound source. Two-tone complexes that varied in frequency and amplitude ratio resulted in concurrent changes in the oscillation mode and magnitude of the travelling wave (Fig. 6a,b). A larger amplitude ratio of the sinusoidal components that comprised the original sound source produced more energetic travelling waves (Fig. 6a-c).

The spectro-temporal pattern of this travelling wave provided a robust index of the perceptual salience of phantom tones perceived by the human ear (Fig. 7a). A beat-probe paradigm provided a measure of listeners' perceived occurrence of phantom percepts. Listeners heard increased incidences of beat phenomenon (indexing the phantom tone) when the amplitude ratio of the raw two-tone complex signal was increased, which paralleled an increased power magnitude of the travelling wave (Fig. 7a). Larger power spectra of this travelling wave corresponded to higher incidences of perceivable phantom tones, and smaller power resulted in a lower percentage occurrence of perceived phantom tones, across different two-tone frequency combination conditions (Fig. 7a,b). The spectro-temporal pattern of the IMF was robust in describing amplitude-related modulation in acoustical perception to phantom tones and can reflect listeners' perceptual

¹ The frequency resolution here (Δf), defined as the frequency different between two adjacent data points along the frequency axis on the spectrogram. For the FFT spectrum, the computation was equal to the sampling rate divided by the total number of samples, $\Delta f = \frac{f_s}{N_{fft}} = \frac{f_s}{N_{fft}} = 10.76$ Hz. For the HHT spectrum, the computation was equal to the frequency range divided by the number of frequency bins, $\Delta f = \frac{f_{range}}{N_{freqbin}} = \frac{f_{range}}{N_{freqbin}} = 6.25$ Hz. For the Morlet wavelet spectrum, the computation was equal to the frequency range divided by the number of analyzed samples in log, $\Delta f = \log_{10} \frac{f_{range}}{N} = \log_{10} \frac{f_{range}}{N} = 0.42$ Hz.

Fig. 4 Comparison of Hilbert-Huang transform (HHT), short-time Fourier transform (STFT) and Morlet wavelet analyses on two-tone complex. **(a)** HHT power spectrum (left) and time-frequency spectrogram (right) of an example two-tone complex used in psychophysical procedure at $f_1 = 1303$ Hz and $f_2 = 2,002$ Hz. Empirical mode decomposition showed an additional phantom travelling wave

(CDT) at the frequency $2f_1 - f_2 = 724$ Hz. **(b)** STFT frequency spectrum (left) and time-frequency spectrogram (right) of the same two-tone complex. **(c)** Morlet wavelet power spectrum (left) and time-frequency spectrogram (right) of the same two-tone complex. No phantom component was revealed using STFT and Morlet wavelet transform methods

Fig. 5 EEMD analysis of phantom tone component. **(a)** Time series of the nine highest frequency IMFs based on EEMD separation of two-tone complex. Oscillation pattern of the third IMF corresponds to phantom-tone frequency. **(b)** HHT spectrogram of the third IMF. **(c)** Hilbert marginal spectra (red) shows frequency matches exactly onto the

frequency of phantom-tone component at 724 Hz. Fourier frequency spectra shown for comparison (blue). **(d)** Significance test of the nine IMF components against white noise. The third IMF indicated by a red asterisk (other IMFs in black) is located above the 99% confidence level marked by magenta line (blue line indicates 95%)

Fig. 6 Occurrence of phantom travelling wave as a function of two-tone amplitude ratios. **(a)** Time-frequency spectrum from EEMD of two-tone complex ($f_1 = 1,363$ Hz, $f_2 = 2,002$ Hz, CDT = 724 Hz) for $a_1/a_2 = 0.1$, 0.3, and 0.9 (top to bottom) respectively. **(b)** Hilbert power spectrum of

the two-tone stimuli at amplitude ratio $a_1/a_2 = 0.1$, 0.3, and 0.9 (top to bottom), respectively. **(c)** The magnitude of the third IMF component (corresponding to phantom) shown independently as a function of two-tone amplitude ratios

differences as a function of different frequency combinations (Fig. 7b). A high positive Pearson correlation coefficient between IMF power and percentage of beats perceived was obtained, suggesting a strong relationship between travelling wave dynamics and psycho-acoustical responses to phantom tones (Fig. 7a,b). Phase coherence of two two-tone

combination complexes was computed as the Pearson correlation coefficient of phase angles for each tone complex. Cross-correlation coefficients between the oscillatory pattern of the two-tone complex and the travelling wave associated with the highest frequency combination tone pair ($f_1/f_2 = 1,636/2,002$ Hz), with equal amplitudes ($a_1 = a_2 = 1$) showed

Fig. 7 Comparison between travelling-wave pattern and listeners' response in identifying the occurrence of phantom tones. **(a)** Psychoacoustical performance in detecting phantom tone occurrence averaged from three normal-hearing listeners plotted as line compared to IMF power spectra (shown in bar; error bars represent ± 1 standard deviation) for $f_1 = 1,363$ Hz (left), 1400 Hz, and 1,636 Hz (right), respectively. **(b)** Comparison graph of human ear perception of phantom occurrence and IMF power as a function of frequency and amplitude ratio of two-tone complex. **(c)** Cross-correlation coefficient between oscillatory mode of two-tone complexes and the travelling

wave incidence associated with $f_1 = 1,636$ Hz, $a_1 = a_2 = 1$. Pearson's (R) correlation are depicted in gray scale for each frequency-amplitude combination, with lighter color (white) indicating stronger phase coherence. Sensitivity index (d') in discriminating phantom tone elicited from corresponding two-tone combinations are shown in colors differing for each frequency combination condition. Each colored circle represents average d' in detecting a pair of amplitude-modulated beat probe. Lighter colored circles represent lower detection sensitivity. Colored lines indicate first degree linear fit to human discrimination sensitivity data

a remarkably similar resemblance with listeners' sensitivity (measured by d') in discriminating the magnitude of phantom tones elicited from corresponding frequency-amplitude combinations (Fig. 7c).

Discussion

We showed that a non-linear time series decomposition revealed the occurrence of a phantom tone travelling wave, traditionally reported as an inner-ear artifact originating in sensory hair cells (Barral & Martin, 2012). Further, we show that a non-linear decomposition approach effectively approximates human auditory perception by respective changes of spectro-temporal dynamics of IMFs that are parallel to the perceptual spectro-temporal properties of phantom tones experienced by listeners. The occurrence of this spatial-temporal travelling wave is interesting for several reasons. First, the original sound source did not contain any indication of this travelling wave component, as the two-tone stimuli were composed of frequencies other than the travelling wave properties. Despite its "absence" within the raw incidence signal, the spectro-temporal properties of the travelling wave can faithfully reflect level and frequency dependent changes to the input sound source. Second, the existence of this spatial-temporal wave has never before been observed on the sound's Fourier spectrum. We believe that the EEMD separation method revealed this travelling-wave component due to its capability to capture instantaneous intra-wave oscillations. Even though the original sound source was linear and stationary, using a decomposition approach, which requires no *a priori* basis, revealed additional components that would be ignored within linear structures. In fact, through utilization of the EEMD approach, we have also identified the existence of travelling waves emanating from similar auditory phenomenon, such as virtual pitch (i.e., phantom fundamental) harmonic complexes, which had no acoustic correlates at the perceived frequency.

The spatial-temporal pattern of the travelling waves revealed by EEMD analysis of two-tone combinations closely paralleled human perceptual experience. Spectro-temporal dynamics of travelling waves corresponded to listeners' perceptual spectro-temporal properties of these tones and reflected the tone's perceptual variations, due to level and frequency dependence of the input sound source. The mean instantaneous speed of the travelling wave was surprising, for it described the perceived frequency of the phantom tone. Further, power spectra of the travelling wave indicated listeners' perceived magnitude of phantom tones. Temporal coherence of the dynamics of oscillatory mode reflected listeners' sensitivity in perceptual discrimination of amplitude-modulated phantom tones. To our knowledge, this is the first application of a time series decomposition approach to describe human

psychoacoustical performance to phantom tones. The fact that IMFs can robustly index human psychoacoustical processing to pitch-related phenomenon means that IMFs can possibly reflect the dynamic processing of human ear mechanics and/or the combination of these mechanics with central cortical processes associated with these types of tones.

The present findings have important implications regarding the true source of the phantom information in complex signals. Current theories on phantom processing share several similarities, including some combination of sensory hair-cell activities and central auditory processing mechanisms to pitch extraction (Ziębakowski, 2012). These theories typically have assumed that phantom percepts exist exclusively intra-aurally, and, thus, significant efforts have been placed on accounting physiologically for the generation of non-linear components (either by resorting to inner ear vibration mechanisms or cortical processing). However, such explanations have not been sufficient in describing phantom and other similar pitch phenomenon (Ashmore, 2008). The origins of these non-linear cochlear mechanics remain debatable, and studies that have shown maxima in places corresponding to spectral components of non-linear distortions have done so only with cooperation from the central auditory system (Robles et al., 1997). Further, some researchers have suggested that outer hair cells, which move (are motile) as they receive efferent information from the brain, can cause some vibration in the basilar membrane as the source of the non-linearity (Ruggero & Rich, 1991; Eguiluz et al., 2000). Thus, it cannot be determined whether the CDT phenomenon is an effect of central mechanism alone or the result of peripheral non-linearities alone. Our findings suggest an alternative view in that perhaps the true source of the phantom sound, in fact, exists extra-aurally as non-linear components within the input signal. This is consistent with Lohri et al. who reported the existence of extra-aural combination tones produced by violins (Lohri, Carral, & Chatziioannou, 2011). If this phantom frequency exists as a non-linear combination, one potential implication could be that the basilar membrane responds directly to the extra-aural component frequency corresponding to phantom percepts, and not generated via non-linear cochlear mechanisms.

The current findings have implications for many areas of cognitive processing, including music and speech, but possibly other areas of perception such as vision where the dynamic stimulus may generate phantom percepts. These distortions products (phantoms) have been used to create controlled auditory illusions of sound sources and synthesize fuzzy effects in technological music, including the creation of illusory localization effects of phantom sources (Kendall et al., 2014). In music applications, many composers and performers are aware of and use CDTs to develop musical layers (Haworth, 2011). These applications manipulate the effects of CDTs either by time-domain processing of amplitude- or frequency-modulated components of the primary tones or by using

certain harmonics for the primary tones. In speech processing, one recent study has shown that the perceptual effects of CDTs influenced vocal speech imitation. Specifically, listeners who were perceptually biased toward hearing the missing fundamental showed a better capacity to imitate the vocal pitch of the model talkers (Postma-Nilsenova & Postma, 2013). This suggests that individual sensitivity to combination tones affects the accuracy of vocal pitch imitation, an important aspect of human interaction.

Conclusions

The empirical mode decomposition method revealed a perceptible auditory component that was not apparent in the original sounds. The spectro-temporal pattern of this extra-aural component predicted listeners' perceptual experience to phantom tones. Our results provide a different view to the true source of auditory phantom perception that extends from the use of linear time series. Findings suggest that phantom tones are not “ghost” percepts, as conventionally believed, but exist extra-aurally as non-linear components in the original signals. The ability of EEMD to characterize human auditory perception to phantom phenomenon, and its potential extension to other speech and music signals, is promising.

Acknowledgements We would like to thank Prof. Wei-Kuang Liang, Prof. Chi-Hung Juan, and Dr. Ling-Yueh Yang for helpful suggestions. We gratefully acknowledge financial support from the Ministry of Science and Technology of Taiwan.

Compliance with ethical standards

Competing interests statement The authors declare that they have no competing financial interests.

References

- Ashmore, J. (2008). Cochlear Outer Hair Cell Motility. *Physiological Reviews*, 88(1), 173.
- Barral, J., & Martin, P. (2012). Phantom tones and suppressive masking by active non-linear oscillation of the hair-cell bundle. *Proceedings of the National Academy of Sciences*, 109(21), E1344–E1351. doi: <https://doi.org/10.1073/pnas.1202426109>
- Bhagat, S. P., & Champlin, C. A. (2004). Evaluation of distortion products produced by the human auditory system. *Hearing Research*, 193(1), 51–67.
- Capranica, R. R., & Moffat, A. J. M. (1980). Nonlinear Properties of the Peripheral Auditory System of Anurans. In A. N. Popper & R. R. Fay (Eds.), *Comparative Studies of Hearing in Vertebrates* (pp. 139–165). New York: Springer New York.
- Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76(3), 1698.
- Chen, J. (2009). Application of empirical mode decomposition in structural health monitoring: Some experience. *Advances in Adaptive Data Analysis*, 1(4), 601–621.
- Chen, S., Su, H., Zhang, R., Tian, J., & Yang, L. (2008). Improving Empirical Mode Decomposition Using Support Vector Machines for Multifocus Image Fusion. *Sensors*, 8(4).
- Cooper, N., & Rhode, W. (1997). Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. *Journal of Neurophysiology*, 78(1), 261–270.
- Cummings, D. A. T., Irizarry, R. A., Huang, N. E., Endy, T. P., Nisalak, A., Ungchusak, K., & Burke, D. S. (2004). Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature*, 427(6972), 344–347.
- Dong, B. (2017). Characterizing resonant component in speech: A different view of tracking fundamental frequency. *Mechanical Systems and Signal Processing*, 88, 318–333.
- Eguiluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A., & Magnasco, M. O. (2000). Essential non-linearities in hearing. *Physical Review Letters*, 84(22), 5232.
- Fastl, H., & Zwicker, E. (2013). *Psychoacoustics: Facts and Models* (Vol. 22): Springer Science & Business Media.
- Goldstein, J. L. (1967). Auditory Nonlinearity. *The Journal of the Acoustical Society of America*, 41(3), 676–699.
- Haworth, C. (2011). Composing with Absent Sound. In *Proceedings of the International Computer Music Conference*, pp. 342–345.
- Huang, N. E. (2001). Computer Implemented Empirical Mode Decomposition Method Apparatus and Article of Manufacture for Two-Dimensional Signals. 6,311,130,B1, 30 Oct 2001.
- Huang, H., & Pan, J. (2006). Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*, 86(4), 792–803.
- Huang, N. E., Chen, X., Lo, M. T., & Wu, Z. (2011). On Hilbert spectral representation: A true time-frequency representation for non-linear and nonstationary data. *Advances in Adaptive Data Analysis*, 3(01n02), 63–93.
- Huang, N. E., Long, S. R., & Shen, Z. (1996). The mechanism for frequency downshift in non-linear wave evolution. *Advances in applied mechanics*, 32, 59–117C.
- Huang, N. E., & Shen, S. S. (2014). *Hilbert–Huang Transform and Its Applications* (Vol. 16): World Scientific.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., ... Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995.
- Huang, N. E., Wu, Z., Pinzón, J. E., Parkinson, C. L., Long, S. R., Blank, K., ... Chen, X. (2009). Reductions of noise and uncertainty in annual global surface temperature anomaly data. *Advances in Adaptive Data Analysis*, 1(3), 447–460.
- Johnsen, N., Bagi, P., & Elberling, C. (1983). Evoked acoustic emissions from the human ear: III. Findings in neonates. *Scandinavian audiology*, 12(1), 17–24.
- Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *The Journal of the Acoustical Society of America*, 64(5), 1386–1391.
- Kendall, G. S., Haworth, C., & Cádiz, R. F. (2014). Sound synthesis with auditory distortion products. *Computer Music Journal*, 38, 5–23.
- Khaldi, K., Boudraa, A. O., & Turki, M. (2016). Voiced/unvoiced speech classification-based adaptive filtering of decomposed empirical modes for speech enhancement. *IET Signal Processing*, 10(1), 69–80.
- Khaldi, K., Turki-Hadj Alouane, M., & Boudraa, A. O. (2010). Voiced speech enhancement based on adaptive filtering of selected intrinsic mode functions. *Advances in Adaptive Data Analysis*, 2(1), 65–80.
- Liang, H., Bressler, S. L., Desimone, R., & Fries, P. (2005). Empirical mode decomposition: a method for analyzing neural data. *Neurocomputing*, 65–66, 801–807.

- Lohri, A., Carral, S., & Chatziioannou, V. (2011). Combination Tones in Violins. *Archives of Acoustics*, 36(4), 727. doi:<https://doi.org/10.2478/v10168-011-0049-1>
- MATLAB and Statistics Toolbox Release (2015). The MathWorks, Inc., Natick, Massachusetts, United States.
- Michalski, W., Bochnia, M., & Dziewiszek, W. (2011). Simultaneous measurement of the DPOAE signal amplitude and phase changes. *Archives of Acoustics*, 36, 3, 499–508.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*: Academic Press.
- Nunes, J. C., Guyot, S., & Deléclle, E. (2005). Texture analysis based on local analysis of the Bidimensional Empirical Mode Decomposition. *Machine Vision and Applications*, 16(3), 177–188.
- Plomp, R. (1965). Detectability Threshold for Combination Tones. *The Journal of the Acoustical Society of America*, 37(6), 1110–1123.
- Postma-Nilsenova, M., & Postma, E. (2013). Auditory perception bias in speech imitation. *Frontiers in Psychology*, 4, 1–8.
- Rilling, G., & Flandrin, P. (2008). One or two frequencies? The empirical mode decomposition answers. *IEEE transactions on signal processing*, 56(1), 85–95.
- Robles, L., Ruggero, M. A., Rich, N. C. (1997). Two-tone distortion on the basilar membrane of the chin-chilla cochlea. *Journal of Neurophysiology*, 77(5), 2385–2399.
- Ruggero, M. A. & Rich, N. C. (1991). Furosemide alters organ of Corti mechanism: Evidence for feedback of outer hair cells upon the basilar membrane. *Journal of Neuroscience*, 11, 1057–1067.
- Ruzmaikin, A., & Feynman, J. (2009). Search for climate trends in satellite data. *Advances in Adaptive Data Analysis*, 1(4), 667–679.
- Sheng, Y. (2010). Wavelet Transform. In *Transforms and Applications Handbook*: CRC Press.
- Smoorenburg, G. F. (1972). Audibility Region of Combination Tones. *The Journal of the Acoustical Society of America*, 52(2B), 603–614.
- Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, 134(1), 628–639.
- van Dijk, P., & Manley, G. A. (2001). Distortion product otoacoustic emissions in the tree frog *Hyla cinerea*. *Hearing Research*, 153(1), 14–22.
- Wu, Z., & Huang, N. E. (2005). Ensemble empirical mode decomposition: A noise-assisted data analysis method, *COLA Technical Reports*, 193, Center for Ocean-Land-Atmosphere Studies, Calverton, MD.
- Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1), 1–41.
- Wu, Z., Huang, N. E., & Chen, X. (2009). The multi-dimensional ensemble empirical mode decomposition method. *Advances in Adaptive Data Analysis*, 1(3), 339–372.
- Wu, Z., Schneider, E. K., Kirtman, B. P., Sarachik, E. S., Huang, N. E., & Tucker, C. J. (2008). The modulated annual cycle: an alternative reference frame for climate anomalies. *Climate Dynamics*, 31(7), 823–841.
- Ziębakowski, T. (2012). Combination Tones in the Model of Central Auditory Processing for Pitch Perception. *Archives of Acoustics*, 37(4), 571.