Generalization of the primary tone phase variation method: An exclusive way of isolating the frequency-following response components

Federico Lucchetti, Paul Deltenre, Paul Avan, Fabrice Giraudet, Xiaoya Fan, and Antoine Nonclercq

Citation: The Journal of the Acoustical Society of America **144**, 2400 (2018); doi: 10.1121/1.5063821
View online: https://doi.org/10.1121/1.5063821
View Table of Contents: http://asa.scitation.org/toc/jas/144/4
Published by the Acoustical Society of America

Generalization of the primary tone phase variation method: An exclusive way of isolating the frequency-following response components

Federico Lucchetti,[1,a)] Paul Deltenre,[1] Paul Avan,[3] Fabrice Giraudet,[3] Xiaoya Fan,[2] and Antoine Nonclercq[2]

[1]*Laboratoire de Neurophysiologie Sensorielle et Cognitive CP403/22, Brugmann Hospital, Place Van Gehuchten 4, Brussels, B1060, Belgium*

[2]*Bio-, Electro- and Mechanical Systems CP165/56, Université Libre de Bruxelles, Avenue F. D. Roosevelt, 50 Brussels, B1050, Belgium*

[3]*Laboratory of Neurosensory Biophysics Unité mixte de recherche, Institut national de la santé et de la recherche médicale 1107, University Clermont Auvergne, 28 Place Henri Dunant, BP38 Clermont-Ferrand, Cedex 1, F63001, France*

(Received 19 April 2018; revised 17 August 2018; accepted 28 August 2018; published online 26 October 2018)

The primary tone phase variation (PTPV) technique combines selective sub-averaging with systematic variation of the phases of multitone stimuli. Each response component having a known phase relationship with the stimulus components phases can be isolated in the time domain. The method was generalized to the frequency-following response (FFR) evoked by a two-tone (f_1 and f_2) stimulus comprising both linear and non-linear, as well as transient components. The generalized PTPV technique isolated each spectral component present in the FFR, including those sharing the same frequency, allowing comparison of their latencies. After isolation of the envelope component $f_2 - f_1$ from its harmonic distortion $2f_2 - 2f_1$ and from the transient auditory brainstem response, a computerized analysis of instantaneous amplitudes and phases was applied in order to objectively determine the onset and offset latencies of the response components. The successive activation of two generators separated by 3.7 ms could be detected in all ($N = 12$) awake adult normal subjects, but in none ($N = 10$) of the sleeping/sedated children with normal hearing thresholds. The method offers an unprecedented way of disentangling the various FFR subcomponents. These results open the way for renewed investigations of the FFR components in both human and animal research as well as for clinical applications. © *2018 Acoustical Society of America*. https://doi.org/10.1121/1.5063821

[PXJ] Pages: 2400–2412

I. INTRODUCTION

A. The primary tone phase variation technique and the frequency-following response

In the field of auditory research, the primary tone phase variation (PTPV) technique has, up to now, been applied to distortion product otoacoustic emissions (DPOAEs) in order to selectively extract the distortion products evoked by pairs of pure tones, the so-called primaries (Whitehead *et al.*, 1996). This method greatly facilitates analysis of the properties of the targeted spectral component (SC) isolated in the temporal domain. The PTPV technique requires *a priori* knowledge of the phase response behaviour of the SCs expected to be evoked by a multitone stimulus. Each SC having a known phase relationship with the stimulus components phases can be selectively isolated from all the other ones using time-domain ensemble sub-averaging. By systematically varying the phases of the primaries between stimulus presentations and computing one subaverage in which the phase of the target SC is maintained constant, the latter will average normally, whereas the other SCs will cancel out. By applying a version of this technique targeting the acoustic cubic distortion tone (CDT) recorded in the ear canal, Whitehead *et al.* (1996) showed that the CDT onset latency could directly be measured from the averaged temporal waveform. They were also able to compute the CDT rise time and detect phase and amplitude changes occurring several milliseconds after onset, suggesting interaction of multiple generators with different onset latencies. The frequency-following response (FFR) is another type of physiological response that could benefit from the PTPV technique. This short-latency electrical evoked potential reproduces, as its name implies, the frequencies contained in sustained stimuli as well as, when multi-frequency stimuli are used with appropriate frequency ratios, distortion products of cochlear or even central origin (Pandya and Krishnan, 2004). The FFR is clearly more complex than DPOAEs, since according to the details of the recording technique, it may contain cochlear pre-neural activity, i.e., the pre-synaptic cochlear microphonic (CM; Shaheen *et al.*, 2015), in addition to neural linear and non-linear components (Elsisy and Krishnan, 2008; Smith *et al.*, 2017) generated in different levels of the auditory pathways, from the cochlear nerve to the cortex. Component latencies are

[a)]Also at: Bio-, Electro- and Mechanical Systems CP165/56, Université Libre de Bruxelles, Avenue F. D. Roosevelt, 50 Brussels, B1050, Belgium. Electronic mail: federico.lucchetti@ulb.ac.be

logically expected to progressively increase from peripheral to central generators. The FFR also contains the transient auditory brainstem response (ABR) evoked by stimulus onset and offset (Kraus and Nicol, 2005). The sustained neural components of the FFR result from the neural phase-locking process and thus offer a means to probe the quality of the temporal code capturing the spectro-temporal properties of auditory stimuli at various levels of the auditory pathways. Complex stimuli as simple as a pair of tones evoke FFRs exhibiting phase-locking to both the stimulus envelope (ENV) and its temporal fine structure (TFS), i.e., to the instantaneous acoustic pressure variations. For stimuli without energy at the ENV frequency (e.g., pairs of closely spaced primaries, harmonic series with missing fundamental, amplitude modulated pure tones), most of the neural ENV spectral energy is introduced within the auditory system by the half-wave rectification process at the inner hair cell synapse. Therefore, the FFR-ENV, also called the envelope following response (EFR), is usually considered as being of neural origin, but weaker pre-neural components cannot be excluded from near-field recordings close to the cochlea (Shaheen *et al.*, 2015). The surface recorded FFR-TFS reflects both pre-neural (CM) and neural activity. A burgeoning literature reflects the growing interest currently devoted by researchers and clinicians to the FFR as a tool to objectively investigate the quality of neural phase-locking (Aiken and Picton, 2008; Ananthakrishnan *et al.*, 2016), which is thought to play a critical role in many aspects of normal and impaired hearing (Henry *et al.*, 2014; Kale and Heinz, 2010; Lorenzi *et al.*, 2006; Moore, 2008; Zhong *et al.*, 2014). The FFR is currently considered as revealing the integrity of sound processing in the brain. Ongoing FFR studies apply to the interconnected fields of learning and ecological oral communication, the latter encompassing the topic of understanding under adverse listening conditions. Music perception and the effects of musical training and experience are also subjects of FFR studies. A recent review detailing the scope of FFR research and applications can be found in Kraus *et al.* (2017). The currently recommended method to disentangle the various surface-recorded FFR components has been described by Aiken and Picton (2008). In this method, two response subsets are evoked by spectrally identical stimuli of opposite acoustic polarity and identical level, referred to as rarefaction (R) and condensation (C) according to the direction of their initial pressure change. By, respectively, adding and subtracting the two subsets, the $R+C$ response highlights the FFR-ENV, whereas the $R-C$ highlights the FFR-TFS. The former, being invariant to stimulus polarity, is eliminated by the subtraction process. The latter, being driven by the acoustic pressure fluctuations, is severely attenuated by the addition process. However, the R and C addition/subtraction method is not completely satisfactory when confronted with signals like multitone evoked FFRs that contain more than one component, each of which can be either sensitive or insensitive to stimulus polarity. The PTPV technique resolves this problem, but requires as many sets of phase-rotated stimuli as the number of targeted SCs. Here, we generalized the PTPV basic principle in order to extract several SCs with a single set of stimuli. The generalized primary tone phase variation (gPTPV) technique relies on the following mathematical rationale. A multicomponent response signal $R(t)$ such as the multitone evoked FFR is a linear superposition of N narrowband SCs $\{Y_1,\ldots,Y_N\}$ such as

$$R(t)=\sum_{l=1}^{N}Y_l(t)=\sum_{l=1}^{N}A_l(t)\exp(i\Phi_l(t)). \quad (1)$$

Each of these SCs is characterized by its instantaneous amplitude $A_l(t)$ (IA) and instantaneous phase $\Phi_l(t)$ (IP). The latter is the sum of the integrated instantaneous frequency (IF) and an arbitrary phase offset. Knowing that FFR signals emerge from the cochlear mechanoelectrical transduction process and the phase-locking of auditory neurons onto the stimulus periodicities, time-varying properties of the response are dependent on the stimulus spectro-temporal properties. Hence, the FFR evoked by a two-tone stimulus of frequencies f_1 and f_2 will be composed of SCs of mean frequency $f_l=\alpha_1 f_1+\alpha_2 f_2$ $(\alpha_1,\alpha_2\in\mathbb{Z})$, where the additional SCs not present in the stimulus arise from the non-linear properties of the auditory system. Similarly, if stimulus components are subjected to phase shifts ϕ_1 and ϕ_2, SC waveforms will be subjected to an equivalent phase shift $\Phi_l(t)\mapsto\Phi_l(t)+\phi_l$ so that $\phi_l=\alpha_1\phi_1+\alpha_2\phi_2$. The latter equations relating the stimulus to the evoked response can straightforwardly be generalized for any stimulus composed of an arbitrary number of tones. The gPTPV method as a generalization of the opposite polarity stimulation becomes self-evident. The order of a SC is defined as $|\alpha_1|+|\alpha_2|$. If $\phi^R=\alpha_1\phi_1+\alpha_2\phi_2$ is the phase of an $|\alpha_1|+|\alpha_2|$ order SC evoked by a negative polarity stimulus R, then the phase of the same SC evoked by a pair of primaries of opposite polarity C is rotated by $\phi^C=\phi^R+\pi(\alpha_1+\alpha_2)$. Due to this unique phase relationship to the primary phases, odd order SCs $(\alpha_1+\alpha_2=2n+1,n\in\mathbb{N})$ belong to the FFR-TFS, even order SCs constitute the FFR-ENV $(\alpha_1+\alpha_2=2n,n\in\mathbb{N})$. When stimulating with a set of M successive two-tone stimuli (s_k) indexed by k with the same intensities A_1 and A_2 and spectral contents f_1 and f_2 but with different phases ϕ_1 and ϕ_2, $s_k(t,\phi_1,\phi_2)=A_1\sin(2\pi f_1+\phi_1(k))+A_2\sin(2\pi f_2+\phi_2(k))$, Eq. (1) becomes

$$R_k(t)=\sum_{l=1}^{N}\underbrace{A_l(t)\exp(i\Phi_l(t))}_{=Y_l}\underbrace{\exp(i\phi_l(k))}_{=\Omega_l(k)}=\mathbf{Y}\cdot\mathbf{\Omega}(k), \quad (2)$$

so that the phase of each SC is shifted by a constant factor $\phi_l(k)$. The phase response relating input and output responses of the non-linear transfer function has been well characterized by Billings and Zhang (1994). The latter reference demonstrates the assertion that the phase of a SC (output), whose frequency is a linear combination of primary frequencies (inputs), will be the same linear combination of the primary phases. Based on this *a priori* knowledge relating stimuli to responses, a set of M rotated stimuli can be designed, evoking M responses with rotated phases contained in the phase rotation matrix $\mathbf{\Omega}(k)$. If M SCs are to be expected in the FFR

response, at least M stimuli with different primary phases need to be created and Eq. (2) can be inverted. Henceforth, every SC can be obtained as a linear combination of the evoked responses weighted by coefficients of the inverse phase rotation matrix, $\mathbf{Y} = \mathbf{\Omega}^{-1}(k)R_k(t)$, thus, generalizing the basic PTPV method. The primary aim of the present proof-of-concept study was to demonstrate the efficacy of the gPTPV method to isolate each of the expected FFR SCs. The secondary aim was to explore methods for the objective assessment of the FFR components once they have been isolated in the time domain. This latter development was, as a first step, devoted to the EFR evoked by a two-tone stimulus, for which we sought to objectively detect the activation of serial generators and the timing of their activation relative to the envelope onset and the transient ABR wave V.

II. METHOD

A. Stimulus structure

The two-tone stimulus had an ENV frequency at $f_0 = 217$ Hz, which is within the range of human glottal fundamentals and appears to be among the ones yielding the highest EFR amplitudes and phase-locking values (Shinn-Cunningham *et al.*, 2017). EFRs at such frequencies also have the advantage of being dominated by activity of subcortical generators (Joris *et al.*, 2016), known to better resist the deleterious effects of sleep or sedation (Aoyagi *et al.*, 1993). The latter conditions are mandatory to obtain the FFR sub-μV signals with appropriate signal-to-noise (S/N) ratios in many clinical populations.

1. Frequency content

The stimuli contained the third and fourth harmonics of a missing fundamental at f_0. The temporal representation of the stimulus can be seen in Fig. 1, where the upper part shows stimulus waveforms for different phase relations indexed by k. Stimuli of opposite polarity are grouped together (red and blue).

2. Targeted SCs

The SCs expected to be evoked (see Table I) were those at the primary frequencies $f_1 = 651$ Hz and $f_2 = 868$ Hz to which the even ($f_2 - f_1 = 217$ Hz) and odd ($2f_1 - f_2 = 434$ Hz) orders intermodulation products would be added by auditory non-linearities. Pilot recordings obtained in normal subjects using the standard R and C addition/subtraction method having shown occasional spectral energy at 434 and 651 Hz in the FFR-ENV response (even order SCs, respectively, $2f_2 - 2f_1$ and $3f_2 - 3f_1$), these were also targeted for isolation. The inconsistent presence of a component at 1085 Hz (odd order CDT_{21} at $2f_2 - f_1$) having also been observed in the FFR-TFS waveform, this SC was also targeted for isolation as well as the transient ABR onset response (see Table I).

3. Starting phases

The starting phases defining the set of stimuli were determined on the basis of the SCs that were to be isolated.

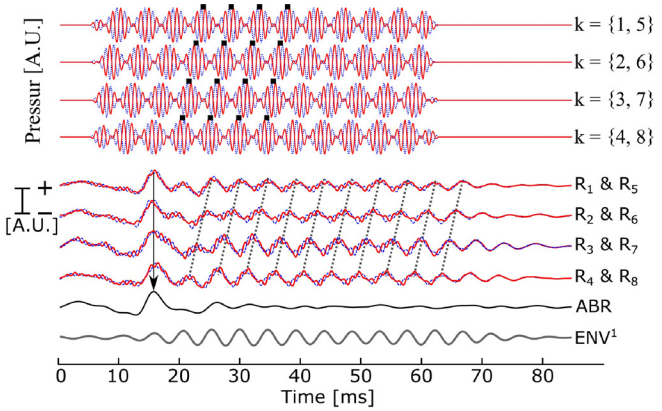

FIG. 1. (Color online) Schematic representation of the PTPV method for the isolation of the ABR and ENV^1 components (ordinates in arbitrary units). The upper part shows stimulus waveforms for different phase relations indexed by k. Stimuli of opposite polarity are superimposed: continuous traces correspond to phase conditions $k = \{1,2,3,4\}$, dotted traces correspond to phase conditions $k = \{5,6,7,8\}$. Black squares mark the temporal positions of the envelope maxima, highlighting the constant phase variation ($\Delta\phi = -\pi/2$) of the missing fundamental as phases of the primaries are varied. The lower part shows their corresponding schematic evoked responses: $\{R_1,R_2,R_3,R_4\}$ (continuous) and $\{R_5,R_6,R_7,R_8\}$ (dotted). The phase of the envelope of each evoked response reproduces the missing fundamental phase shift as illustrated by the obliquity of the dotted lines passing through the envelope peaks. By summing all eight responses after rotating the phases of each pair of evoked responses according to the gPTPV method, the ENV^1 component of all eight responses are brought back in phase so that they can be constructively averaged together (ENV^1 waveform). By contrast the ABR phase is insensitive to stimuli phase variation (down pointing arrow) and a straightforward averaging of the eight evoked responses cancels both the FFR-TFS and FFR-ENV components and isolates the ABR (ABR waveform).

Eight stimuli, individually indexed by k, were generated with the appropriate different starting phases to isolate the eight targeted components. Table I indicated, for each value of k, the phase of each primary and the expected phase of the response components. The last four stimuli $k = \{5,\ldots,8\}$ phases were shifted by a constant factor π with respect to the first four. This created a set of stimuli where the first four waveforms were in opposite polarity to their last four counterparts. This switched the polarity of the FFR-TFS SCs in order to disentangle it from the FFR-ENV components. By computing the phase rotation matrix $\mathbf{\Omega}$, applying its inverse [MATLAB *inv*(), the MathWorks, Natick, MA] to the eight evoked responses $\mathbf{R} = \{R_1,\ldots,R_k\ldots,R_8\}$ we obtain the following set of equations to isolate each SC mentioned above:

$$\begin{pmatrix} \text{ABR} \\ \text{ENV}^1 \\ \text{ENV}^2 \\ \text{ENV}^3 \\ F_1 \\ F_2 \\ \text{CDT}_{12} \\ \text{CDT}_{21} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i & 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i & 1 & i & -1 & -i \\ 1 & -i & -1 & i & -1 & i & 1 & -i \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & i & -1 & -i & -1 & -i & 1 & i \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \end{pmatrix}. \quad (3)$$

ABR corresponds to the transitory response, ENV^1 to the missing fundamental or stimulus envelope, ENV^2 and ENV^3, respectively, refer to the first and second harmonic

TABLE I. Targeted SCs belonging to the FFR-ENV, FFR-TFS components along with the transient ABR. SCs are classified with respect to their (α_1,α_2) values and frequency relation $f_{\alpha 1,\alpha 2}$. Phase values $\phi_{\alpha_1\alpha_2}$ of the primaries and the evoked response are shown as a function of the stimulus index k and represented by an arrow where $\uparrow = 0$, $\leftarrow = \pi/2$, $\downarrow = \pi$, and $\rightarrow = 3\pi/2$.

	Stimulus		ABR	FFR-ENV			FFR-TFS			
(α_1,α_2)			(0, 0)	(−1,1)	(−2,2)	(−3,3)	(1,0)	(0,1)	(2,−1)	(−1,2)
SC			ABR	ENV^1	ENV^2	ENV^3	F_1	F_2	CDT_{12}	CDT_{21}
$f_{\alpha_1\alpha_2}$	f_1	f_2	/	$f_2 - f_1$	$2(f_2 - f_1)$	$3(f_2 - f_1)$	f_1	f_2	$2f_1 - f_2$	$2f_2 - f_1$
f (Hz)	651	868	/	217	434	651	651	868	434	1085
$\phi_{\alpha_1\alpha_2}$	ϕ_1	ϕ_2	0	$\phi_2 - f_1$	$2(\phi_2 - \phi_1)$	$3(\phi_2 - \phi_1)$	ϕ_1	ϕ_2	$2\phi_1 - \phi_2$	$2\phi_2 - \phi_1$
$k = 1$	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
$k = 2$	←	↓	↑	←	↓	→	←	↓	↑	→
$k = 3$	↓	↑	↑	↓	↑	↓	↓	↑	↑	↓
$k = 4$	→	↓	↑	→	↓	←	→	↓	↑	←
$k = 5$	↓	↓	↑	↑	↑	↑	↓	↓	↓	↓
$k = 6$	→	↑	↑	←	↓	→	→	↑	↓	←
$k = 7$	↑	↓	↑	↓	↑	↓	↑	↓	↓	↑
$k = 8$	←	↑	↑	→	↓	←	←	↑	↓	→

distortions of the envelope, CDT_{12} and CDT_{21} refer to the CDTs, and F_1 and F_2 refer to the linear frequency response. The output of the phase rotation matrix had to be scaled down by a factor of 8 to ensure correct subaverages amplitudes. The multiplication of real valued signals $R_k(t)$ by the imaginary unit i was performed by applying the Hilbert transform (HT) $\mathcal{H}$ on the former multiplied by (–1), such as $iR = e^{i(\pi/2)}R \mapsto -\mathcal{H}(R)$, which geometrically translates to a $-\pi/2$ phase rotation of the waveform for positive frequency values (Schreier and Scharf, 2010). The process of SC isolation can be graphically understood from Fig. 1, illustrating the phase behaviour of FFR responses with respect to a change in primary phases. For each phase condition k the envelope maxima of the two-tone stimulus shift backward by a constant factor of $\pi/2$, inducing the same phase lag in the ENV^1 component of the evoked responses. The effect of applying the inverse phase rotation matrix, which is at the heart of the generalization process, is to compensate these phase lags prior to averaging the ENV^1 component. By contrast the ABR peak that remains stable across all phase conditions can be directly isolated without correction (all the multiplicands in its associated row of the inverse matrix are equal to one), whereas all the other components are canceled due to their constant phase variations (see Table I).

B. Instrumentation

The computing workload required for stimuli synthesis and presentation, data acquisition and averaging, as well as real-time dynamic fast Fourier transform (FFT) analysis was distributed between the host computer running LABVIEW 2014 software and the Digital Signal Processing processors of a Tucker-Davis Technologies (TDT, Alachua) RZ6 platform. The two systems were interfaced through *Active X* technology. Stimuli were delivered monaurally at an overall intensity of 85 dB sound pressure level (SPL) through electromagnetically shielded insert earphones (Etymotic ER-300). Stimulus Onset Asynchrony was 105.7 ms. Stimulus waveforms were generated with a sampling frequency of 24 414 Hz and presented as tone bursts with a duration of 57 ms, including a $\cos^2$ rise and fall time of 1 ms. Electrical responses were recorded differentially through a pair of Ag/AgCl cupula surface electrodes positioned at the vertex and on the spinous process of the seventh cervical vertebra (C_Z - Cv_7), with the ground electrode placed on the forehead. Signals were amplified and digitized using a differential preamplifier (R14PA 4-channel Medusa Preamp, TDT) with a sampling frequency of 24 414 Hz and a recording time of 84 ms. The delay of the electrical stimulus was adjusted so that the onset of the acoustical stimulus in the ear canal occurred 5 ms after acquisition triggering. The digitized EEG signal was transferred via optic fibre to the TDT RZ6 interface where the averaging was controlled by LABVIEW 2014. Signals were digitally bandpass filtered between 100 and 3000 Hz with a 12 dB/octave attenuation. Signal acquisition was performed in an acoustically and electromagnetically shielded room.

C. Participants

The FFR data reported in the present work have been collected in two groups of subjects. Group I comprised 12 (8 females) young normal unpaid adults having volunteered to participate. Their ages ranged from 23 to 29 years. Group II comprised ten children (two girls) who were referred for objective physiological evaluation after age-adapted psycho-acoustical methods had proved impossible or unreliable but whom exhibited normal electrophysiological thresholds and otoacoustic emissions (OAEs), thus, excluding an audibility defect. Group II subjects had ages ranging from 6 weeks to 5 years. Most of them were referred for speech retardation, some with a suspicion or a diagnosis of autistic spectrum disorder. The younger ones were referred after repeated failures of neonatal hearing screening tests. Inclusion criteria for group I required bilateral normal hearing as defined by pure-tone hearing thresholds below 20 dB hearing level (HL) at octave frequencies from 0.25 to 8 kHz, normal 226 Hz tympanograms (Interacoustics Titan, software ver.3.00.49, Middelfart) and normal DPOAEs (GSI Audera, software ver.2.1.0.8, Eden Prairie) for f_2 frequencies from 598 to 8004 Hz, with five f_2 values/

octave. The f_2/f_1 ratio was 1.2 and their respective levels were 55 and 65 dB SPL. The minimal S/N ratio required to recognize the presence of a distortion product at any f_2 frequency was 6 dB combined with an absolute level reaching at least 0 dB SPL. Exclusion criteria were: presence of tinnitus, history of ototoxic exposure, and family history of hearing disorders. Stimulation side was randomly assigned. Inclusion criteria for group II relied entirely on the results of physiological testing comprising normal tympanograms with age-adjusted probe tone (Interacoustics Titan, software ver.3.00.49), normal click-evoked ABRs, normal audiograms estimated from auditory steady-state evoked potentials (ASSEP) thresholds and normal DPOAEs. DPOAEs and evoked potentials were obtained on a GSI Audera system (software ver. 2.1.0.8). Click (100 μs) thresholds had to be equal to or better than 20 dB nHL and estimated audiometric thresholds had to be equal to or better than 20 dB HL at 0.5, 1, 2 and 4 kHz. DPOAEs were obtained at five f_2 frequencies (996, 1418, 2004, 2824, and 3996 Hz) using the same stimulation parameters and recognition criteria as for group I. Exclusion criteria were abnormal ABR interpeak intervals (I-V, I-III, and III-V), history of ototoxic exposure, and family history of hearing disorders. Stimulation side was defined on the basis of the lowest average click and ASSEP-estimated thresholds, a few children having demonstrated unilateral hearing loss at the issue of their physiological evaluations. Group II recordings were obtained during post-feeding natural sleep for subjects below six months of age and under sedation with intravenous propofol (10 μg kg^{-1} hour^{-1}), under spontaneous respiration according to the institutional protocol for comprehensive objective evaluation of auditory function in older difficult-to-test children. The study protocol had been approved by the local ethical committee (Ref. No. CE 2017/140).

D. Signal averaging

For group I, individual trials evoked by a given phase condition were rejected from the final average if their peak amplitude exceeded $\pm 14\,\mu$V. Rejected trials were repeated so that a total of $N_{\text{Avg}} = 10\,000$ stimuli contributed to the final responses. For group II, responses were averaged $N_{\text{Avg}} = 6000$ times, individual trials with peak amplitude exceeding $\pm 9\,\mu$V were rejected and replaced. For both groups, two FFR computations were performed per subject: the first resorted to the classical alternate polarity stimulation where $R = R_1$ and $C = R_5$ were averaged each N_{Avg} times depending on subjects group. The second was accomplished by applying the gPTPV method described above, each subaverage $\{R_1,\ldots,R_8\}$ results from $N_s = N_{\text{Avg}}/8$ accepted stimuli. This allowed the comparison of results obtained by both methods. Each subaverage computation was shadowed by a plus-minus averaging procedure (Schimmel, 1967), where all subaverages $\sum_{k=1}^{8} R_k$ were "plus-minus" averaged together so that every SC was canceled out in order to estimate the post-averaging residual noise $s_{\text{noise}}(t) = (1/N_s)\sum_{s=1}^{N_s}(-1)^s \sum_{k=1}^{8} R_k$. This allowed computation of individual components signal-to-noise (S/N) ratio levels based on root mean square ratios.

E. Signal analysis

1. Spectral analysis

Signals were Hamming-windowed and subjected to a FFT. Mean spectral amplitudes per frequency bin were computed by averaging three neighbouring bins. The detection threshold of a SC was set to 2.5 standard deviations of the mean spectral amplitude above the noise floor.

2. IA and phase

The IA and IP values of the signal were estimated using two different transforms: the HT and the complex wavelet transform (CWT; Le Van Quyen *et al.*, 2001). The first consists of computing the convolution of the gPTPV waveform $s(t)$ with a $1/t$ function, enabling to express the original signal into an analytical representation $\xi(t) = s(t) + i\tilde{s}(t) = A(t)\exp[i\phi(t)]$, where $\tilde{s}(\tau)$ is the Hilbert transform of $s(t)$. The CWT transform of the signal was obtained by convolving the gPTPV waveform with a complex Morlet waveform defined at frequency f and time t by

$$\begin{cases} W_s(\tau,f) = \int_{-\infty}^{+\infty} dt\, s(t)\Psi^*_{\tau f}(t) \\ \Psi_{\tau f}(t) = \sqrt{f}\exp(i2\pi f(t-\tau))\exp\left(-\dfrac{(t-\tau)^2}{2\sigma^2}\right), \end{cases} \quad (4)$$

where $\Psi^*_{\tau f}(t)$ is the complex conjugate of the wavelet function and $\sigma = N_{\text{cycles}}/6f$ with N_{cycles}, the number of cycles of the wavelet, set to 1. For a gPTPV waveform of mean frequency $f = 220$ Hz, this limited the CWT analysis frequency range to $[f - 4f/N_{\text{cycles}}, f + 4f/N_{\text{cycles}}] = [0, 880]$ Hz (Le Van Quyen *et al.*, 2001). IA and IP values were extracted by, respectively, computing the modulus and phase angle of the complex valued functions (4). The phase difference $\Delta\phi_f(t)$ was defined as $|\phi_f(t) - 2\pi tf|$ where $\phi_f(t)$ is the IP of the gPTPV waveform and f is its mean frequency.

3. ENV[1] onset and offset latencies

Onset (t_{ON}) and offset (t_{OFF}) latencies were determined by imposing two conditions on the IA and phase difference values. The first condition stemmed from the following rationale: t_{ON} and t_{OFF} determine the period during which the response maximally "phase-locks" to the stimulus. In this temporal window, the phase difference remains constant so that the windowed phase-locking value $PLV_f(t_1,t_2) = (1/t_2 - t_1)\sum_{t=t_1}^{t_2} |\exp(i\Delta\phi_f(t))|$ reaches a maximum for $t_1 = t_{\text{ON}}$ and $t_2 = t_{\text{OFF}}$. The second condition refined the t_1 and t_2 values by imposing that at t_{ON} the IA had to rise above 10% of its nominal value and fall below 10% at t_{OFF} (Whitehead *et al.*, 1996).

4. ABR wave V peak latency

The peak latency of ABR wave V was visually determined by one of the authors accustomed to ABR tracings scoring. The rule used was to select the latest vertex positive peak before the major negative through crossing the baseline.

5. Phase analysis and SC generators

If several sequential generators with different activation latencies contributed to a given FFR SC, we expected to observe a discontinuity of the phase difference (phase shift) at time t_ϕ at which a new generator is activated. We should then also observe a prolongation of the duration of the response $L_{RESP} = t_{ON} - t_{OFF}$ with respect to the duration of the stimulus $L_{STIM} = 57$ ms. The overall duration L_{RESP} was expected to be equal to L_{STIM} plus the activation latency of the latest generator.

F. Dependence on noise level

Phase estimation by means of the HT and CWT methods was expected to be heavily affected by the presence of noise. To estimate the minimal signal-to-noise ratio level S / N needed to warrant reliable results from both methods by testing for false and true detection rates, we generated ENV^1 waveforms embedded in various levels of noise for 2 groups of 12 *in silico* subjects. The noise structure was obtained from the 12 group I subjects by the plus-minus averaging procedure. Two scenarios, illustrated in Fig. 2, have been set up. First, false detection rates were evaluated from a set of 12 single fixed amplitude sinusoids at frequency $f_{ENV^1} = 217$ Hz with onset latencies t_{ON} picked randomly from a uniform distribution between 9 and 14 ms. Second, true detection rates were evaluated from a set of 12 composite signals resulting from the sum of 2 similar fixed amplitude sinusoids shifted in time one relative to the other. The earliest ones had onset latencies randomly generated between 9 and 14 ms. The second ones had onset latencies $t_\phi = t_{ON} + \delta t$, where δt was randomly picked from a uniform distribution between 1 and 6 ms to simulate the onset of a second generator. Both sinusoids had a duration of 57 ms. Onset latency and phase shift detections were performed on both data sets across increasing levels of noise so as to achieve various S / N ratios covering a 1–6.5 interval by 0.1 steps. In the first simulation, the probability of falsely detecting one or more phase shifts (false positive) for the first data set was computed $P[N_\phi > 0] = (1/12)\sum_j^{12} \delta_j$, where $\delta_j = 1$ if the number of detected phase shifts $N_\phi > 0$ else $\delta_j = 0$ if no phase shifts have been detected. For the second data set, the probability of correctly detecting t_ϕ within a temporal range of +/–0.5 ms (true positive) was estimated by $P[t_M \in [t_\phi - 0.5\, t_\phi + 0.5]] = (1/12)\sum_j^{12} \delta_j$, where t_M is the measured phase shift time. If the phase shift is correctly detected within the temporal window then $\delta_j = 1$ else $\delta_j = 0$.

III. RESULTS

A. Minimal S/N ratio level

Figure 2 illustrates the performance of both methods of phase shift detection as a function of S/N ratio level. Above 5.5 dB, both methods reached 100% of true detections and no false one. This result helped us to verify that individual averaged waveforms were sufficiently clean to be submitted to the phase analysis.

B. SC isolation

Figures 3 and 4 illustrate temporal and spectral domains representations of the FFR obtained by the standard (Fig. 3) vs the gPTPV (Fig. 4) methods in the same subject whose data are representative of group I. The comparison of Figs. 3 and 4 clearly demonstrates the superiority of the gPTPV method. The standard method has sorted the various SC between the two resulting $R + C$ and $R - C$ waveforms according to their relative sensitivity to stimulus acoustic polarity so that the component at the envelope frequency was segregated in the $R + C$ waveform, whereas components at the primaries and CDT_{12} frequencies were more prominent in the $R - C$ waveforms. However, each resulting tracing contained three SCs, two of them (at the CDT_{12} and F_1 frequencies) being found in both spectra. On the contrary, the gPTPV method completely isolated each component on the basis of their specific phase-frequency relationship with the stimulus so that components sharing the same frequency (CDT_{12} and ENV^2) were separated. The transitory onset response (note also the presence of a much weaker offset one) was also isolated from all the other components. All SC components, as well as the ABR, were present in every subject of both groups except the ENV^3 (seen in three group II subjects and in none from group I), and the CDT_{21} (seen in two group II subjects and in none from group I).

C. Fixed phases vs gPTPV

Although the gPTPV method is intrinsically linear, we experimentally verified that the decomposition it accomplished was effectively performed as a linear process devoid of distortion in the frequency domain. We compared the

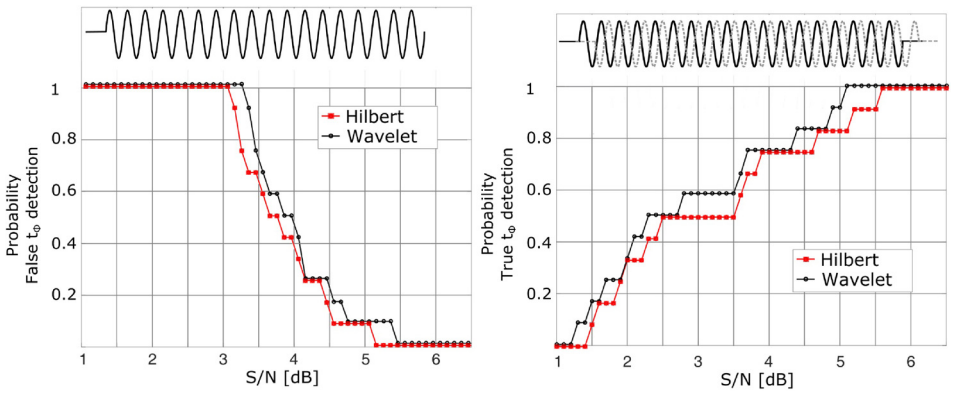

FIG. 2. (Color online) (Left) Scenario 1: Probability of detecting a false phase shift within a single sinusoid (continuous) at onset latency t_{ON} as a function of the S/N ratio. (Right) Scenario 2: Probability of detecting a true phase shift t_ϕ within a signal comprised of a sum two sinusoids of equal frequency with different onset latencies t_{ON} (continuous) and t_ϕ (dotted). Results are shown for the HT (circles) and the CWT (squares).

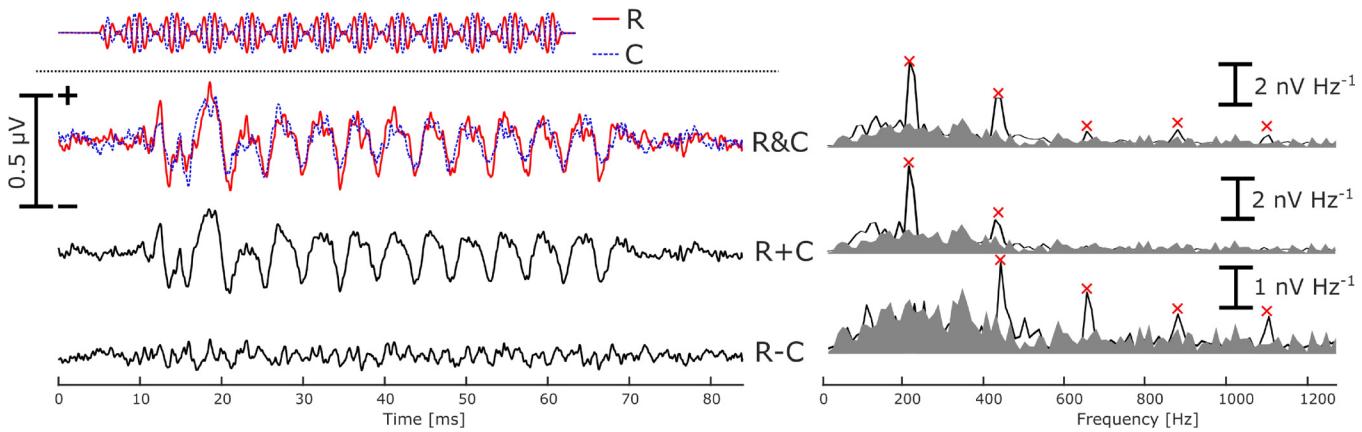

FIG. 3. (Color online) Results of the standard *R* and *C* addition/subtraction method. (Top left) Temporal waveforms of the stimulus composed of the two primaries $f_1 = 651$ Hz and $f_2 = 868$ Hz. Rarefaction (continuous trace) and condensation (dotted trace) polarities are superimposed. (Bottom left) Superimposed FFRs evoked by the *R* and *C* stimuli and standard addition and subtraction of opposite polarities responses, respectively, enhancing the FFR-ENV and FFR-TFS components. (Bottom right) FFTs of the above temporal waveforms. Filled area represents the noise floor. Spectral peaks significantly rising above noise are marked with an "X."

gPTPV waveforms $Y_l(t)$) with the response evoked by a fixed phase stimulus $R(t)$, by computing their difference $\Delta(t) = R(t) - \sum_l^N Y_l(t)$ for each subject from groups I and II. A FFT was applied on each $\Delta(t)$ waveform and then averaged across subjects (Fig. 5). No spurious SC were detected in both groups.

D. Phase shifts and response durations

The temporal waveforms of the ENV[1] component isolated by the gPTPV technique were submitted to our dual criterion method of detecting multiple sequential generators for every subject. The overall (i.e., computed across the entire recording epoch) S/N ratio of the isolated ENV[1] SC of all subjects was above 5.5 dB, the minimum level ensuring optimal detection according to simulation results. The latencies of phase shift occurrences relative to the onset latencies $t_\phi - t_{ON}$ were measured, and response durations were compared to the stimulus one. Both the HT and CWT methods were independently applied to the data and their results compared. Figure 6 illustrates the analysis based on the HT method in one representative group I subject. A first phase shift occurring at $t_{\phi_1} = 4.4$ ms after component onset, signals the activation of a second generator phase-locking onto

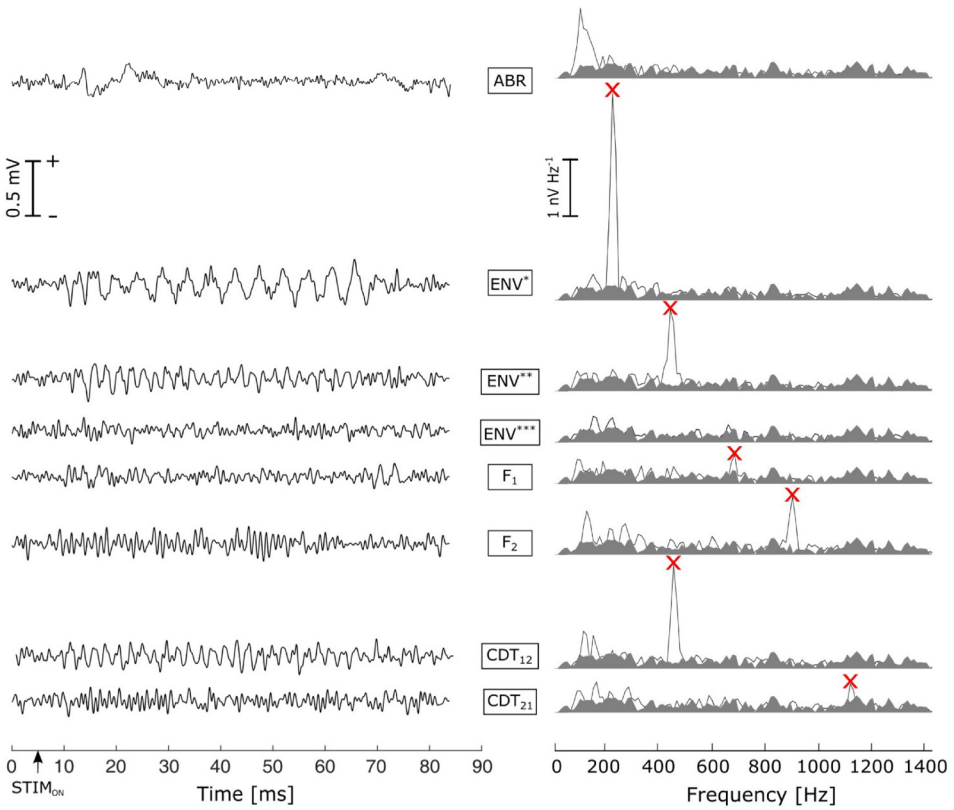

FIG. 4. (Color online) Results of the gPTPV method. (Left) gPTPV temporal waveforms, computed using Eq. (3) with their associated spectral representation (right). Filled area in the spectra represents the noise floor. Spectral peak detection is marked with an "X."

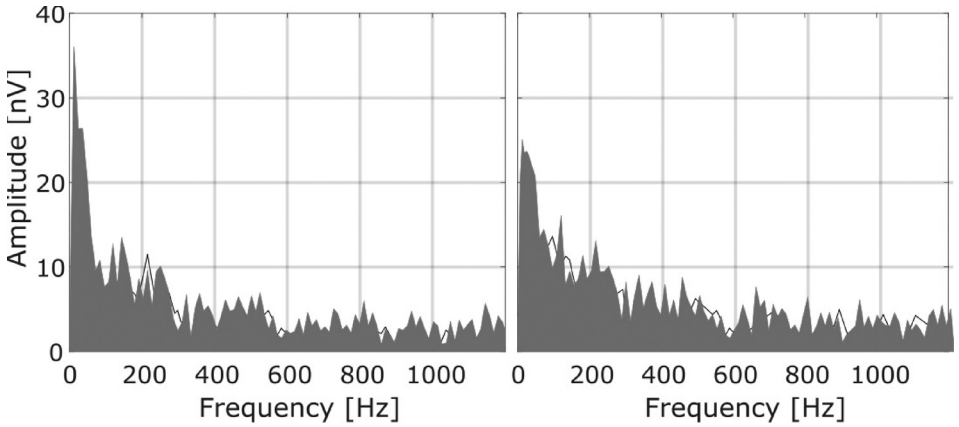

FIG. 5. Averaged spectra of the temporal difference $R(t) - \sum_l Y_l(t)$ for the adult group (left) and children group (right). Averaged noise spectra is depicted by the filled area.

the stimulus envelope with a longer latency as confirmed by the prolonged overall duration of the response $t_{OFF} - t_{ON}$ exceeding that of the stimulus by 2.8 ms. A second phase shift was detected at $t_{\phi_2} \approxeq 55.1$ ms after onset and 4.7 ms before offset. This second phase shift thus marked the offset of the first generator. The CWT-based method yielded exactly the same results. Figure 7 summarizes the results for the two groups by providing the probability density functions (bin width = 0.5 ms) of the relative phase shift latencies and response durations differences. All group I subjects (Fig. 7, top) showed a short latency phase shift t_{ϕ_1} associated with a corresponding response prolongation. Only 5 of the 12 subjects showed a late phase shift t_{ϕ_2} occurring at the expected latency of response termination of the first component. The presence of two clusters of phase shift latencies was obvious and confirmed by the MATLAB *kmeans*() function. Within each cluster, the values were normally distributed ($p < 0.005$ after the Kolmogorov-Smirnov test). Their means and standard deviations were, respectively, 3.7 ± 0.7 ms (HT), 3.8 ms $\pm$ 0.9 ms (CWT) for the first phase shift and 56.1 ± 0.9 ms (HT) and 56.9 ± 0.8 ms (CWT) for the second one. Outside these two clusters, other phase shifts were randomly distributed in time across subjects between 10 and 54 ms after response onset. The response durations relative to stimulus length also distributed normally ($p < 0.005$) with respective means of 3.2 ± 0.7 ms and 3.5 ± 0.7 ms for the HT and the CWT methods, respectively. Phase shift latencies within the first cluster $t_{\phi 1} - t_{ON}$ were highly correlated (Pearson correlation coefficient = 0.95) with the prolongation of the response values $L_{RESP} - L_{STIM}$, for both the HT and CWT methods. A linear regression analysis (least square approach) of the scattering of the response prolongation as a function of the relative phase shift times gave a slope of 0.83 ± 0.21 ($p < 0.05$) with a y-intercept of –0.82 ms for the HT. For the CWT we derived a slope of 1.03 ± 0.49 ($p < 0.05$) with a y-intercept of –0.84 ms. Examination of the IA and PLV graphs gave a hint about the probable cause for the inconstant presence of the late phase shift heralding the first component offset: as can be seen from Fig. 6 (bottom), both IA and PLV values decline during the second half of the response. The S/N ratio level was therefore expected to be lower at the time of the first component offset than at its onset. This was verified by computing a local S/N ratio level within a temporal window of 5 ms centered on the mode of the latency distributions of the early and late phase shifts t_{ϕ_1} and t_{ϕ_2}. The ratio remained above 4.5 dB for all short latency phase shifts and above 4.1 dB for the five late detected late ones but fell below 2.3 dB for the seven subjects where t_{ϕ_2}

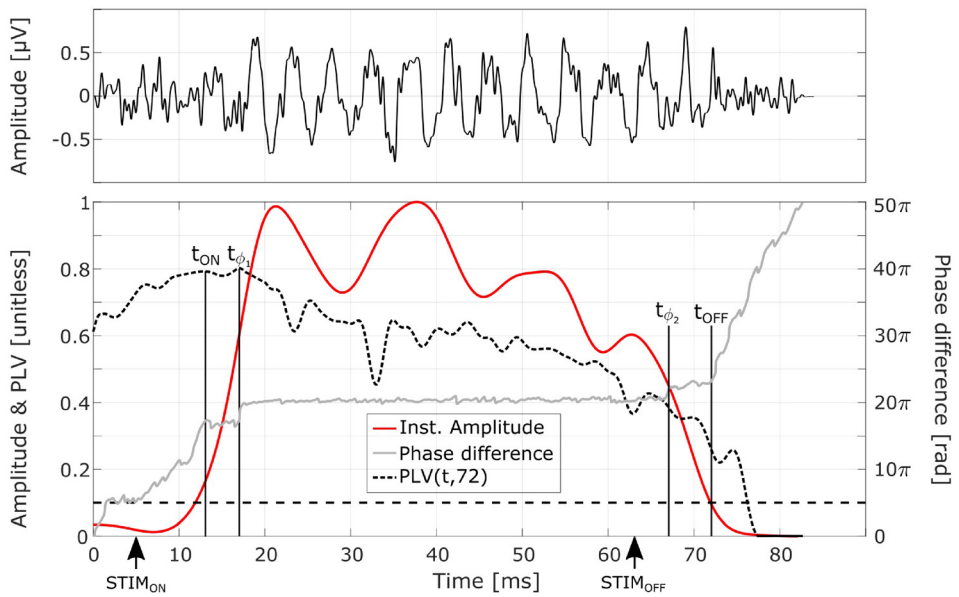

FIG. 6. (Color online) (Top) ENV[1] waveform, has been isolated using the gPTPV method. (Bottom) IA and IP analyses using the HT of the ENV[1] component shown. Within the time span during which the IA (continuous curve) stays above 10% (horizontal dashed line) of its nominal value, the phase difference values graph (light grey) shows two distinct stationary regions where phase-locking between the response and the stimulus occurs. The first plateau is localized between the onset latency ($t_{ON} = 13.1$ ms) and the first phase shift ($t_{\phi_1} = 17.2$ ms), the second between the phase shift and the offset of the response ($t_{OFF} = 72$ ms). A second phase shift is visible at time t_{ϕ_2}. The windowed PLV (t_1,t_2) curve (dashed black curve) highlights two local maxima appearing at times $(t_1,t_2) = (t_{ON},t_{OFF})$ and (t_{ϕ_2},t_{OFF}).

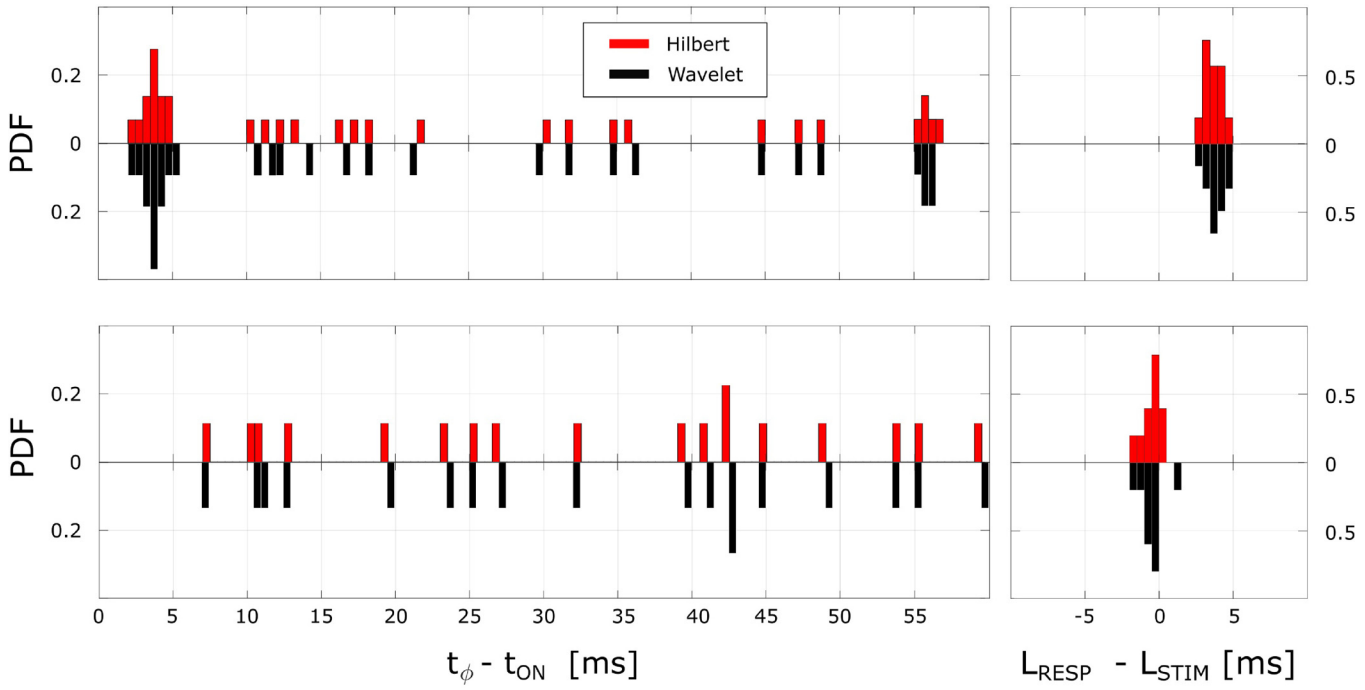

FIG. 7. (Color online) Probability density functions depicting the distributions of the derived parameters. (Leftboxes) Observed phase shift latencies relative to measured component onset ($t_\phi - t_{ON}$). (Right boxes) Observed response durations minus stimulus duration ($L_{RESP} - L_{STIM}$). (Top) Adult group (group I). (Top) Children group (group II). Results are shown for HT and CWT (see the legend).

remained undetected. None of the children (Fig. 7, bottom) showed a short latency phase shift within the distribution observed in group I and no value clustering was found. Moreover, response durations were not prolonged, the means of the differences between response and stimulus durations being, for both methods, very close to 0 with small standard deviation (respectively, 0.5 ± 0.2 and -0.4 ± 0.5 for the HT and CWT methods; see Fig. 7).

E. ENV^1 vs ABR latencies

The onset latency of the ENV^1 component and the peak latency of the transient ABR wave *V* were measured relative to the acoustical stimulus onset and one relative to the other in both groups. The means ±standard deviation (SD) values of the distributions of ENV^1 latencies were, respectively, of 5.2 ± 0.8 ms and 7.9 ± 0.7 ms in groups I and II. For the wave *V* peak, the values were 7.5 ± 1 ms and 8.2 ± 0.4 ms. In group I, the ENV^1 onset preceded the ABR peak by 2.1 ± 0.7 ms, whereas with values of 0.32 ± 0.64 ms, this delay was much shorter in the children group. This difference between the two groups was highly significant ($p < 0.0001$; two-tailed unpaired *t*-test). With respective values of 0.38 and 0.43, the correlation coefficient (Pearson) between the ENV^1 onset and ABR wave *V* latencies did not reach significance in either group ($p = 0.17$ and 0.21).

IV. DISCUSSION

A. Proof of concept

The gPTPV method successfully segregated every expected FFR SC as well as the transient ABR. Furthermore, it enabled differentiation of SCs sharing the same frequency, an accomplishment that would have been impossible to achieve by resorting to classical filtering methods. The concept that the gPTPV method should be useful for FFR analysis was therefore validated by the experimental data. The gPTPV works on the basis that the FFR contains oscillatory components (both sustained: SCs and transient: ABR) characterized by their time-varying amplitudes and phases, and can hence be expressed as a sum of components [see Eq. (1)], forming a set of an orthogonal basis. The basic transform principle is similar to the Fourier transform but by manipulating the stimulus phase and resorting to subaveraging, the gPTPV method yields superior results neither confined to the spectral domain nor limited to narrow-band SCs: it provides distinct temporal domain representations of both transient and sustained components. Whether or not the evoked sustained components share the same frequency, they will be isolated anyway, provided that their phase responses are different. Moreover, as experimentally confirmed in Sec. III C, the gPTPV method recombines responses by merely summing/subtracting or phase-rotating them in the temporal domain without amplitude or phase distortion. The linearity of the method was substantiated by the demonstration that no spurious SCs have been introduced. Henceforth, FFR amplitudes are invariant to stimulus phase and generalize results from previous studies showing that the standard *R* and *C* stimulus polarities evoke identical FFR amplitudes (Aiken and Purcell, 2013; Xu and Ye, 2014). In other words, the time-varying properties, such as the IA and IP, are conserved.

B. EFR temporal structure

After the gPTPV technique had disentangled the EFR main oscillatory component ENV^1 from its higher order harmonics and the transitory ABR response that appeared

concomitantly to EFR onset, we were able to perform accurate phase analyses revealing the component temporal structure. It showed that in normal hearing adults, the ENV[1] component of the EFR contains two successive different stationary portions, the later one starting with a mean latency of 3.7 ms relative to the onset of the first one. The total duration of the response was prolonged with respect to the stimulus duration by 3.1–3.4 ms. With correlation coefficient and slope values close to one, the correlation and linear regression analyses performed on response prolongation ($L_{RESP} - L_{STIM}$) vs relative first phase shift latencies ($t_{\phi 1} - t_{ON}$) led to the conclusion that two successive generators, each producing a response having the stimulus duration, are activated 3.7 ms one after the other. The negative intercept value (≈–0.8 ms) of the regression line is likely due to measurements errors of the phase estimation methods combined with imprecision in the value of the stimulus effective duration since it had $\cos^2$ rise and fall times of 1 ms. Whereas a phase shift heralding a second component onset was present in all adult subjects, only five of them exhibited a phase shift signaling the end of the first component. This was clearly due to the progressive decline of IA and PLV values and the subsequent reduction of the effectiveness of the phase analysis method toward the end of the response. The progressive decline of its amplitude is a well-known FFR feature (Worden and Marsh, 1968) attributed to neural adaptation. Remarkably, none of the children showed evidence of a second, later ENV[1] component. Three hypotheses can be proposed to explain this group difference. The first one is that the reason for which patients were referred to objective audiological evaluation was associated with an FFR anomaly. Most of them suffered from language delays, some of them with autistic traits, but three of them who had been referred for repeated failure of newborn hearing screening, eventually proved to have normal hearing and gave no further reason for concern. Therefore, these three children had no identified explanation for a missing late component, except that they were the youngest among their group (1.3, 1.7, and 13 months of age) so that a maturational effect may also have contributed to the results. The second hypothesis indeed invokes a maturational effect: the FFR is known to be influenced by auditory experience (Krishnan *et al.*, 2005; Wong *et al.*, 2007) with indications that the auditory brainstem still matures through adolescence (Krizman *et al.*, 2015). The third hypothesis is that sleep, either natural or induced, suppressed the second later component. None of our adult subjects went to sleep during the recordings, whereas all children were asleep, most under sedation. Since the FFR components susceptibility to the wake/sleep state is known to increase with their latency (Aoyagi *et al.*, 1993), this is a plausible explanation for the lack of the second component in group II. Whichever proper explanation(s) will eventually emerge, the systematic difference between the two subject groups who differ by several features offering plausible albeit yet unproven causes for the difference, underscores the usefulness of the temporal structure analysis method. We also showed that whatever the subjects group, the results describing the component temporal structure were independent on the phase estimation method, which is consistent with the results obtained by Le Van Quyen *et al.* (2001). Although the *in silico* study (see Sec. III A) yielded useful results to understand the experimental data, it failed to capture all the features of the real recordings. Whereas the surrogate signal was perfectly stationary, the neural response entrained by a stationary stimulus is expected to adapt and show mild fluctuations of its IF around the SC mean frequency value. With the method used here, if the IF leaves the frequency analysis window, a phase shift will be detected. (Hurtado *et al.*, 2004). This, combined with a progressive reduction of S/N ratio levels is the likely explanation for the spurious false detections randomly scattered over time. In the same vein, the phase shift associated with the first component offset went undetected in subjects achieving too low a S/N ratio level within the time window during which the computation was performed.

C. EFR generators

Several studies attempted to infer the generation site of the EFR by measuring the group delay of the response (King *et al.*, 2016; Shaheen *et al.*, 2015; Shinn-Cunningham *et al.*, 2017). This requires the use of signal filtering and phase smoothing, both techniques being known to bias the computed phase-gradient delay (Shera and Bergevin, 2012). Moreover, when the signal is generated by multiple sources with different latencies, group delay measurements yield a weighted average between both latencies, their weights being determined by the relative amplitudes of each source emission. The gPTPV technique resolves these issues by transforming the total FFR response into separate narrow band SCs, consequently enabling, provided that S/N ratio levels are under control, an accurate phase shift and signal duration analysis to effectively decompose EFR generators in the time domain. In the present study, the latter decomposition resorted to the HT and CWT methods. In the present study, the complete separation of the ENV[1] component, brought about by the gPTPV method, allowed uncontaminated measurement of its onset latency. The ENV[1] onset occurred prior to the ABR wave *V*. The absence of correlation between the onset latency of the ENV[1] component and the peak latency of the transient ABR wave *V*, as well as their much greater difference observed in the adult group, is in line with previous argumentation stating that these two responses are "functionally distinct" (Bidelman, 2015). This expression was used to mean that the FFR does not represent a convolution of the ABR wavelet with the stimulus' periodicities (Bidelman, 2015). The observed latency discrepancy could also signify that that the two responses come from functionally independent parallel auditory projections ascending the brainstem. Our results showing that the ENV[1] onset occurred prior to the ABR wave *V*, might be taken as suggesting a more caudal origin for the former. One has, however, to be cautious when comparing latencies of transient and sustained components: the *ABR* V_{th} peak is nothing more than an easily identified morphological landmark within a component with an overall duration of more than 1 ms. As discussed above, the gPTPV method isolates SCs on the basis of their phase response. This is clearly not

equivalent to isolating the response of a given physiological mechanism. No definite clue as to the precise physiological origin of the first and second ENV[1] components can be derived from the present data. Given the widely accepted notion that the FFR is generated in several parts of the ascending auditory pathways, the proposed combination (Sec. IV B) of relatively peripheral with a more central generator is the first to come to mind. However, it may also be hypothesized that the $f_2 - f_1$ SC can result from neural phase-locking to both the envelope of the stimulus and to pre-neural cochlear quadratic distortion. Given the primary frequencies used in the present study, such a dual contribution could lead to a first component followed several ms later by neural phase-locking to the cochlear distortion product that has to travel to its apical tonotopic location. We believe that such a mechanism is unlikely to be the cause for the two EFR components observed in our study because the pre-neural quadratic distortion tone is weak in the human species, being rarely observed in DPOAEs. Moreover, evidence for two EFR components was restricted to the adult group. Since the children group had normal cochlear function and since cochlear size and tonotopy are already adult-like after birth, it is unlikely that this hypothesis explains the difference between the two groups. This difference is better explained by the known greater sensitivity of the more central generator to sleep and sedation. However, the situation could be quite different in some mammal species like the rabbit (Whitehead *et al.*, 1996) and mouse (Verpy *et al.*, 2008), which exhibit a greater variety of pre-neural high amplitude distortion tones.

D. gPTPV vs band-pass filtering

Separating the transient ABR waveform from the ENV[1] component was a necessary condition for onset latency and early phase shift measurements. As can be seen from Fig. 4, the entire ABR overlaps partially the ENV[1] component in the temporal and spectral domain. Since IP estimation methods can only give meaningful results when operating on narrow band signals, prior removal of the ABR as well as higher order harmonics from the ENV[1] was mandatory. It is worth considering whether more easily available classical bandpass filtering methods could not achieve the same results. We did not pursue this approach beyond very limited pilot trials that yielded deterring results. After zero-phase bandpass filtering (forward second-order followed by a backward second-order Butterworth filter with cut-off frequency $f_{ENV} \pm 100$ Hz), the HT failed to show the early phase shift that was conspicuous after the gPTPV conditioning. Additionally, despite the minimally invasive nature of this kind of filter (low order and zero-phase), the estimated IP, locally showed several instances of negative slope values implying negative frequencies, a manifest physical impossibility. Moreover identification of an appropriate filtering method can be very laborious since a considerable amount of parameters (order, low and high cut-off frequency, type,…) have to be chosen and fine-tuned in order to preserve one among several neighbouring frequencies without altering its onset behaviour. This is avoided by the gPTPV method, which does not require any parameter adjustment.

E. Subsequent phase analysis

The HT and CWT methods used in the present study for IA and IP estimation may bring filtering related issues. Fundamentally, no spectro-temporal windowing is required to get the analytical representation of a signal using the HT. However, for the HT to operate properly in terms of signal amplitude and phase extraction, the IA should not oscillate faster than the frequency of its main frequency component. Therefore, according to their spectral content, some input signals can require prior filtering. As the gPTPV waveforms are purely narrowband signals, the HT can be applied without the use of filtering techniques. As to the CWT method, it might imply phase smoothing depending on the spectro-temporal resolution of the transform. In the present study, the temporal width of the wavelet window has been set so as to limit its frequency analysis to the [0,880] Hz region where the temporal resolution is optimal. The EFR response at center frequency 217 Hz should therefore be relatively unaffected by smoothing artefacts inherent to the filtering properties of the CWT method. This assumption is substantiated by the fact that the onset latency response duration and phase analysis results did not significantly differ between the HT and the CWT methods.

F. Limitations of the gPTPV method

1. Intrinsic limitations

Two limitations of the gPTPV method can be encountered when several non-linearities contribute to the same SC. On the one hand, if the contributing non-linearities have different phase responses, which is generally the case if their respective orders differ (Billings and Zhang, 1994), the phase of the output response cannot be predicted from the phases of the primaries. On the other hand, if they share the same phase response the gPTPV method will yield a single output waveform, irrespective of the number of contributing non-linearities. The multiplicity of the causative non-linearities will be revealed only if they have distinguishable different latencies, according to the temporal resolution of the analysis technique.

2. Stimuli related limitations

Although the present study was limited to the use of two-tone stimuli, there is no reason for which the gPTPV technique could not be applied to sinusoidally amplitude modulated pure-tone carriers. This type of stimuli has been extensively used to characterize the behaviour of specific peripheral and central auditory neuron types (Joris *et al.*, 2016). Being composed of a sum of three pure tones, it is straightforward to control the starting phases of their constituents and construct an appropriate rotation matrix. By contrast, it is difficult to imagine how the gPTPV method could be applied to natural or even synthesized speech stimuli that are frequently used in FFR studies. Naturalistic speech stimuli contain too many frequencies and would hence require

the computation of a considerable amount of starting phases for the gPTPV to be applicable. However, a peculiar form of simplified speech stimulus, the so-called sine-wave speech analog that reproduces the temporal dynamic of the spectral peaks of the speech formant frequencies (Oxenham *et al.*, 2004), are worth being explored in combination with the gPTPV method, provided that the phase of each stimulus component is constantly under control. Transposed stimuli are also increasingly used in EFR recordings in order to probe the phase-locking response of high characteristic frequency neurons to low-frequency envelopes. Such stimuli are traditionally synthesized in the time domain by multiplying a high-frequency carrier with a half-wave rectified and low-pass filtered low-frequency sinusoid. This allows comparisons of temporal and tonotopic information by resorting to stimuli with identical peripheral auditory representation but tonotopically different neurons (Oxenham *et al.*, 2004). Synthesis of such transposed stimuli in the frequency domain requires too many primary frequencies (at least seven) to achieve a flat envelope between rectified half-waves. Thereafter, a considerable amount of phases and frequencies of the output components have to be predicted to apply the gPTPV. For FFR studies in which the goal is merely to probe the phase-locking behaviour of high center frequency neurons, without the need for a comparison between apical and basal responses, a high-frequency carrier modulated by a low-frequency sinusoid should do the job as demonstrated by the routine clinical recordings of the frequency specific auditory steady-state evoked responses for audiometric applications. Note that for stimuli with more than two primaries there is no penalty linked to the number of stimulus components in terms of averaging time. As illustrated by Fig. 1 and Eq. (3), all evoked responses R_k contribute to each isolated component by virtue of applying the inverse phase rotation matrix. This is the equivalent of obtaining the FFR-ENV by adding the R and C subaverages in the standard method.

G. Prospects

The present proof of concept study opens the way for renewed investigations of the FFR components in both human and animal work. The issue of FFR generators is not without unresolved controversies, the combination of the gPTPV technique with all other methods (recording derivations, maturational studies, natural or experimental lesion studies, specific pathological mechanisms with known molecular defects, training,…) is expected to significantly improve our understanding of this complex response.

ACKNOWLEDGMENTS

This research was financially supported by the Belgian Kids Fund for Pediatric Research (F.L.), the Belgian Fonds National de la Recherche Scientifique (Grant No. J.0092.13; P.D.), and the Brugmann Foundation (P.D.).

Aiken, S. J., and Picton, T. W. (**2008**). "Envelope and spectral frequency-following responses to vowel sounds," Hear. Res. **245**(1-2), 35–47.

Aiken, S. J., and Purcell, D. (**2013**). "Sensitivity to stimulus polarity in speech-evoked frequency-following responses," in *Proceedings of Meetings on Acoustics ICA2013* (Acoustical Society of America, Melville, NY), Vol. 19, p. 050121.

Ananthakrishnan, S., Krishnan, A., and Bartlett, E. (**2016**). "Human frequency following response: Neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss," Ear Hear. **37**(2), e91.

Aoyagi, M., Kiren, T., Kim, Y., Suzuki, Y., Fuse,T., and Koike,Y. (**1993**). "Optimal modulation frequency for amplitude-modulation following response in young children during sleep," Hear. Res. **65**(1-2), 253–261.

Bidelman, G. M. (**2015**). "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR," Hear. Res. **323**, 68–80.

Billings, S. A., and Zhang, S. A. (**1994**). "Analysing non-linear systems in the frequency domain–II. The phase response," Mech. Syst, Signal. Process. **8**, 45–62.

Elsisy, H., and Krishnan, A. (**2008**). "Comparison of the acoustic and neural distortion product at 2f1-f2 in normal-hearing adults," Int. J. Audiol. **47**, 431–438.

Henry, K. S., Kale, K. S., and Heinz, M. G. (**2014**). "Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers," Front. Syst. Neurosci. **8**, 20.

Hurtado, J. M., Rubchinsky, L. L., and Sigvardt, K. A. (**2004**). "Statistical method for detection of phase-locking episodes in neural oscillations," J. Neurophysiol. **91**(4), 1883–1898.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiolog. Rev. **84**(2), 541–577.

Kale, S., and Heinz, K. A. (**2010**). "Envelope coding in auditory nerve fibers following noise-induced hearing loss," J. Assoc. Res. Otolaryngol. **11**(4), 657–673.

King, A., Hopkins, K., and Plack, K. (**2016**). "Differential group delay of the frequency following response measured vertically and horizontally," J. Assoc. Res. Otolaryngol. **17**(2), 133–143.

Kraus, N., Anderson, S., and White-Schwoch, T. (**2017**). "The frequency-following response: A window into human communication," in *Frequency-Following Response* (Springer, Cham).

Kraus, N., and Nicol, T. (**2005**). "Brainstem origins for cortical 'what' and 'where' pathways in the auditory system," Trends Neurosci. **28**(4), 176–181.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (**2005**). "Encoding of pitch in the human brainstem is sensitive to language experience," Cogn. Brain Res. **25**(1), 161–168.

Krizman, J., Tierney, A., Fitzroy, A. B., Skoe, E., Amar, J., and Kraus, N. (**2015**). "Continued maturation of auditory brainstem function during adolescence: A longitudinal approach," Clin. Neurophysiol. **126**(12), 2348–2355.

Le Van Quyen, M., Foucher, J., Lachaux, J.-P., Rodriguez, E., Lutz, A., Martinerie, J., and Varela, F. J. (**2001**). "Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony," J. Neurosci. Methods **111**(2), 83–98.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**(49), 18866–18869.

Moore, B. C. (**2008**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**(4), 399–406.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**(5), 1421–1425.

Pandya, P. K., and Krishnan, A. (**2004**). "Human frequency-following response correlates of the distortion product at 2f1-f2," J. Am. Acad. Audiol. **15**(3), 184–197.

Schimmel, H. (**1967**). "The (±) reference: Accuracy of estimated mean components in average response studies," Science **157**(3784), 92–94.

Schreier, P. J., and Scharf, L. L. (**2010**). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals* (Cambridge University Press, Cambridge, UK).

Shaheen, L. A., Valero, M. D., and Liberman, M. C. (**2015**). "Towards a diagnosis of cochlear neuropathy with envelope following responses," J. Assoc. Res. Otolaryngol. **16**(6), 727–745.

Shera, C. A., and Bergevin, C. (**2012**). "Obtaining reliable phase-gradient delays from otoacoustic emission data," J. Acoust. Soc. Am. **132**(2), 927–943.

Shinn-Cunningham, B., Varghese, L., Wang, L., and Bharadwaj, H. (**2017**). "Individual differences in temporal perception and their implications for everyday listening," in *The Frequency-Following Response* (Springer, Cham), pp. 159–192.
Smith, S. B., Ichiba, K., Velenovsky, D. S., and Cone, B. (**2017**). "Efferent modulation of pre-neural and neural distortion products," Hear. Res. **356**, 25–34.
Verpy, E., Weil, D., Leibovici, M., Goodyear, R. J., Hamard, G., Houdon, C., Lefèvre, G. M., Hardelin, J. P., Richardson, G. P., Avan, P., and Petit, C. (**2008**). "Stereocilin-deficient mice reveal the origin of cochlear waveform distortions," Nature **456**(7219), 255–258.
Whitehead, M., Stagner, B., Martin, G., and Lonsbury-Martin, B. (**1996**). "Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency," J. Acoust. Soc. Am. **100**(3), 1663–1679.
Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci. **10**(4), 420–422.
Worden, F., and Marsh, J. (**1968**). "Frequency-following (microphonic-like) neural responses evoked by sound," Electroencephalogr. Clin. Neurophysiol. **25**(1), 42–52.
Xu, Q., and Ye, D. (**2014**). "Stimulus polarity effects on individual components of speech-evoked frequency following response," in *2014 7th International Conference on Biomedical Engineering and Informatics (BMEI)* (IEEE, Piscataway, NJ), pp. 383–387.
Zhong, Z., Henry, K. S., and Heinz, K. S. (**2014**). "Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas," Hear. Res. **309**, 55–62.