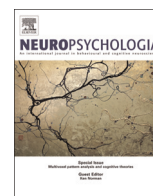




ELSEVIER

Contents lists available at ScienceDirect

## Neuropsychologia

journal homepage: [www.elsevier.com/locate/neuropsychologia](http://www.elsevier.com/locate/neuropsychologia)

# Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept

Gavin M. Bidelman<sup>a,b,\*</sup>, Claude Alain<sup>c,d,e</sup>

<sup>a</sup> Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

<sup>b</sup> School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Avenue, Memphis, TN 38105, USA

<sup>c</sup> Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada

<sup>d</sup> Department of Psychology, University of Toronto, Toronto, ON, Canada

<sup>e</sup> Institute of Medical Sciences, University of Toronto, ON, Canada

## ARTICLE INFO

## Article history:

Received 19 September 2014

Received in revised form

18 December 2014

Accepted 22 December 2014

Available online 24 December 2014

## Keywords:

Auditory scene analysis

Streaming

Sensory systems

Event-related potentials (ERPs)

Frequency-following response (FFR)

Object-related negativity (ORN)

Sound classification

Reaction time

## ABSTRACT

Natural soundscapes often contain multiple sound sources at any given time. Numerous studies have reported that in human observers, the perception and identification of concurrent sounds is paralleled by specific changes in cortical event-related potentials (ERPs). Although these studies provide a window into the cerebral mechanisms governing sound segregation, little is known about the subcortical neural architecture and hierarchy of neurocomputations that lead to this robust perceptual process. Using computational modeling, scalp-recorded brainstem/cortical ERPs, and human psychophysics, we demonstrate that a primary cue for sound segregation, i.e., harmonicity, is encoded at the auditory nerve level within tens of milliseconds after the onset of sound and is maintained, largely untransformed, in phase-locked activity of the rostral brainstem. As then indexed by auditory cortical responses, (in)harmonicity is coded in the signature and magnitude of the cortical object-related negativity (ORN) response (150–200 ms). The salience of the resulting percept is then captured in a discrete, categorical-like coding scheme by a late negativity response (N5; ~500 ms latency), just prior to the elicitation of a behavioral judgment. Subcortical activity correlated with cortical evoked responses such that weaker phase-locked brainstem responses (lower neural harmonicity) generated larger ORN amplitude, reflecting the cortical registration of multiple sound objects. Studying multiple brain indices simultaneously helps illuminate the mechanisms and time-course of neural processing underlying concurrent sound segregation and may lead to further development and refinement of physiologically driven models of auditory scene analysis.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

In social gatherings, human listeners must perceptually integrate the sounds originating from one person's voice, e.g., fundamental frequency ( $f_0$ ) and harmonics, and segregate those of other talkers. The segregation of concurrent speech sounds is promoted by differences in a number of acoustic cues including harmonicity, spatial location, and onset asynchrony (Alain, 2007; Carlyon, 2004; Ciocca, 2008). In the laboratory, the perception of concurrent sound objects can be parametrically induced by mistuning one spectral component (i.e., harmonic) from an otherwise periodic harmonic complex tone. Sounds that are harmonically related (i.e., integer multiple of a fundamental) evoke unitary pitch

\* Correspondence to: School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Avenue, Memphis, TN 38105, USA.  
Fax: +1 901 525 1282.

E-mail address: [g.bidelman@memphis.edu](mailto:g.bidelman@memphis.edu) (G.M. Bidelman).

percepts and are thus heard as belonging to a common source, whereas inharmonicity tends to cue the presence of multiple auditory objects. Low order harmonics that are mistuned by about 4–6% of their original value are “heard out” from the harmonic complex so that listeners perceive two sounds: a complex tone and another sound with a pure-tone quality (e.g., Alain et al., 2001; Darwin et al., 1994; Moore et al., 1986).

Evidence from animal neurophysiology and human scalp-recordings of event-related brain potentials (ERPs) suggest that concurrent sound segregation involves low-level processes that take place along the ascending auditory pathway including the auditory nerve (Sinex et al., 2003), cochlear nucleus (Sinex, 2008), the inferior colliculus (Sinex et al., 2005, 2002), and the primary and associative auditory cortices (Dyson and Alain, 2004; Fishman et al., 2014; Hautus and Johnson, 2005). The detection of a mistuned harmonic in an otherwise periodic signal may involve neurons sensitive to equal spacing between tonal elements (Roberts, 1998; Roberts and Brunstrom, 1998) which, in turn, could be

used to build a harmonic sieve or template (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009; Goldstein, 1973; Hartmann et al., 1990; Lin and Hartmann, 1998). That is, neurons sensitive to frequency periodicity could act as a series of filters that allow harmonically related partials (i.e., frequencies sharing a common  $f_0$ ) to group together with the  $f_0$  and mistuned partials to form separate neural representations and thus, the perception of multiple auditory objects (Darwin et al., 1994, 1995).

Furthermore, evidence from ERP studies suggests that the perception of concurrent sound objects may occur independently of attention (Alain and Izenberg, 2003; Alain et al., 2002; Dyson et al., 2005) and likely involves primary and associative auditory cortices (Arnott et al., 2011; Dyson and Alain, 2004). While the amplitude of the cortical ERPs can differ between passive and active listening conditions (Alain et al., 2001), studies that have directly examined the effects of selective attention have not revealed an effect of attention load on the object-related negativity (ORN) amplitude or latency, a component of the ERPs reflecting the early cortical processing of harmonic mistuning (Alain and Izenberg, 2003). Thus, while arousal may change the ORN, the response largely reflects pre-attentive processing of concurrent sound objects in the early auditory cortices.

In addition to the ORN, a second modulation is often observed (referred to as P400) and emerges when a listener is asked to indicate behaviorally whether they hear one or two concurrent sounds. Accordingly, the later P400 is likely related to attention-dependent cognitive processing of a second auditory object, while the ORN would likely reflect automatic, attention-independent processing of the mistuned harmonic (Alain, 2007; Alain et al., 2001; Hautus and Johnson, 2005). Together, the ORN and P400 provide stimulus- and perception/response-related cortical indices of listeners' ability to segregate concurrent sounds based on periodicity cues, which among other auditory functions, is important for understanding speech in noisy situations (e.g., Binns and Culling, 2007; Swaminathan and Heinz, 2012).

While the relationship between neural responses from the human auditory cortex and concurrent sound perception has been examined, the possible contribution of subcortical brainstem structures has so far remained unexplored in humans. Furthermore, the relationship between brainstem, cortical evoked responses, and perception remains to be determined. Here, we measured brainstem and cortical ERPs within the same participants elicited by harmonic stimuli that do and do not promote concurrent sound segregation. This approach has been helpful to tease apart neural events across the auditory system that index the encoding of physical stimulus characteristics from those related to perception (Bidelman et al., 2013). This approach also allows us to evaluate both primitive (i.e., pre-attentive or automatic) and more complex (i.e., abstract, endogenous) neural codes underlying concurrent sound segregation. Participants were presented with harmonic complex tones where all tonal elements were either in tune (i.e., integer multiple of the fundamental) or included a mistuned component. Human ERPs were also complemented by computational modeling, which allowed us to investigate whether or not neural correlates of concurrent sound perception are present as early as the auditory nerve (AN). Comparisons between AN, brainstem, and cortical responses allowed us to trace the brain correlates and hierarchical signal transformations underlying auditory scene analysis from the periphery through perception and in turn, assess whether these various neural responses correlate with listeners' concurrent sound perception. We hypothesized that inharmonicity would be well-represented in AN and brainstem activity, and that cortical evoked responses would better correlate with listeners' perception. These findings would be consistent with our previous reports, demonstrating that cortical activity reflects correlates of (speech) perception whereas brainstem

responses largely code stimulus-related information (e.g., Bidelman et al., 2013).

## 2. Methods and materials

### 2.1. Participants

Ten English-speaking young adults (6 females; age:  $25.8 \pm 3.0$  years) participated in the experiment. All were right-handed (Oldfield, 1971) and reported a collegiate level of education ( $18.7 \pm 2.7$  years). Audiometric screening confirmed normal hearing sensitivity (i.e.,  $\leq 25$  dB HL) at octave frequencies between 250 and 8000 Hz. Participants reported no previous history of neurological or psychiatric illnesses. None had more than three years of formal musical training [for effects of musical training on concurrent sound segregation, see Zendel and Alain (2009, 2013)]. All participants gave written, informed consent in compliance with an experimental protocol approved by the Baycrest Centre Research Ethics Committee and were paid for their time.

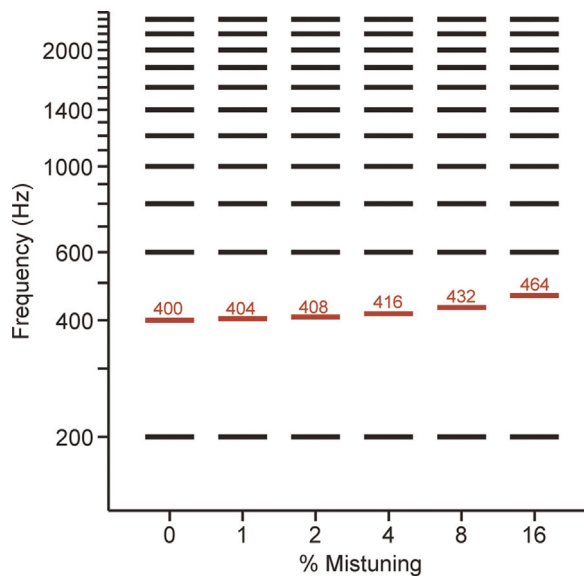
### 2.2. Stimuli

Harmonic complex stimuli were similar to those used in our previous reports (e.g., Alain et al., 2001, 2012; Zendel and Alain, 2009). They were generated by combining 12 pure-tones at integer multiples of a  $f_0=200$  Hz fundamental frequency added in sine phase. Each tone was 100 ms in duration including 10 ms of rise-fall ramping to reduce acoustic spectral splatter and minimize onset responses. Mistuning was achieved by selectively shifting the second harmonic of each complex upward by 1%, 2%, 4%, 8%, or 16% of its original value of 400 Hz (Fig. 1). An additional perfectly harmonic complex (0% mistuning) was included as a control stimulus. Stimulus presentation was controlled by MATLAB<sup>®</sup> 2011 (The MathWorks) routed to a GSI audiometer via a TDT RP2 interface (Tucker-Davis Technologies) and delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-3 A, Etymotic Research). Stimulus intensity was calibrated using a Larson-Davis sound pressure level (SPL) meter (Model 824) measured with a 2 cc, artificial ear coupler (AEC1001). Left and right ear channels were calibrated separately.

### 2.3. Electrophysiological recording and data preprocessing

Data acquisition and response evaluation were similar to previous reports from our laboratory (e.g., Bidelman et al., 2013, 2014a, 2014b). During ERP recording, listeners heard 200 randomly ordered exemplars of each token and were required on each trial, to indicate whether they heard "1-" or "2-sounds" (binary response) as quickly as possible. Participants were told that the stimuli would either sound like a single pitch ("1-sound") or a pitch with an additional faint sounding pure-tone ("2-sounds") (Alain et al., 2001). The next stimulus was then initiated between 400 and 600 ms after the participant's behavioral response (jittered in 20-ms steps, rectangular distribution). Thus, the interstimulus interval (ISI) for cortical ERP presentations was, on average, between  $\sim 900$  and 1100 ms. An additional 2000 trials (ISI=150 ms) were then collected in order to detect sub-microvolt brainstem ERPs (Bidelman et al., 2013). Given that early brainstem ERPs are largely unaffected by attention (Hillyard and Picton, 1979; Picton and Hillyard, 1974; Picton et al., 1971; Woods and Hillyard, 1978), participants watched a self-selected movie with subtitles during blocks of brainstem recording to facilitate a calm yet wakeful state. Electrophysiological recording lasted  $\sim 2$  hrs.

The continuous electroencephalogram (EEG) was recorded between electrodes placed on the high forehead at the hairline



**Fig. 1.** Stimuli used to probe concurrent sound perception. Harmonic complex tones ( $f_0=200$  Hz) were parametrically mistuned by increasing the frequency of the second harmonic from 0% to 16% of its original value. Across this continuum, listeners' percept changes from hearing only one (0%, 1% condition) to hearing two auditory objects (8%, 16%).

referenced to linked mastoids. A third electrode on the mid-forehead served as common ground. This montage is optimal for recording evoked responses of both subcortical and cortical origin (Bidelman et al., 2013, 2014a; Krishnan et al., 2012; Musacchia et al., 2008). Contact impedances were maintained below 1 k $\Omega$  throughout the duration of the experiment. The EEG was digitized at 20 kHz and bandpass filtered online between 0.05 and 3500 Hz (SynAmps2, NeuroScan). Traces were then segmented (cortical ERP:  $-100$  to 600 ms; brainstem ERP:  $-40$  to 210 ms), baselined to the pre-stimulus interval, and subsequently averaged in the time domain to obtain ERPs for each condition (Delorme and Makeig, 2004). Trials exceeding a  $\pm 50$   $\mu$ V threshold were rejected as artifacts prior to averaging. Grand average evoked responses were then bandpass filtered in different frequency bands to isolate subcortical (175–2500 Hz) and cortical (1–30 Hz) evoked potentials, respectively (Bidelman et al., 2013; Krishnan et al., 2012; Musacchia et al., 2008).

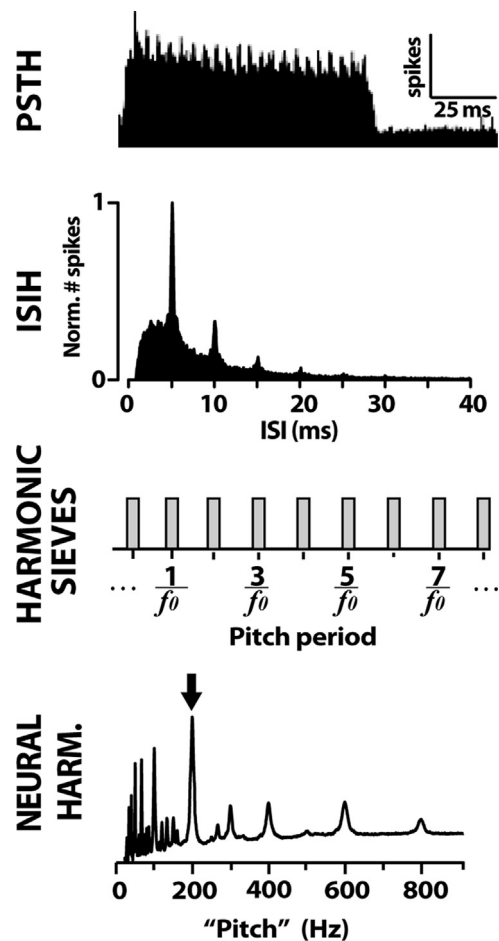
## 2.4. Data analysis

### 2.4.1. Behavioral data

Participants' responses were quantified by calculating the number of stimuli listeners reported hearing two auditory objects from among the total presentations for each stimulus condition. Reaction times (i.e., classification speeds) were computed separately for each participant as the mean response latency across trials within a given condition. Implausible reaction times (RTs) shorter than 200 ms or exceeding 1500 ms were discarded as outliers and excluded from further analysis (Bidelman et al., 2013).

### 2.4.2. Brainstem ERPs

For each participant and condition, Fast Fourier Transforms (FFTs; length=20,000 points) were first computed from each brainstem time waveform to index the magnitude of spectral information contained in subcortical activity. "Neural harmony" (i.e., neural pitch salience) was then estimated from each spectrum using a harmonic template analysis whereby a series of sieves selected spectral activity at  $f_0$  and its integer multiples [for details, see Bidelman and Heinz (2011)]. For quantifying relevant activity,



**Fig. 2.** Procedure for computing "neural harmony" from auditory nerve (AN) responses. Single-unit responses were generated by presenting stimuli (100 stimulus repetitions per stimulus type) to a computational model of the AN (Ibrahim and Bruce, 2010; Zilany et al., 2009) using 70 model fibers (CFs: 125–16,000 Hz). (First and second rows) From individual fiber spike trains, interspike interval histograms (ISIHs) were first computed to index pitch periodicities contained in individual fibers. Single-fiber ISIHs were then summed to create a pooled, population-level ISIH indexing the various periodicities coded across the AN ensemble. (Third row) Each pooled ISIH was then passed through a series of periodic sieves each reflecting a single pitch template (i.e.,  $f_0$ ). The magnitude at the output of a single sieve reflects the salience of pitch-relevant information for the corresponding  $f_0$  pitch. (Bottom row) Analyzing the output across all possible sieve templates ( $f_0=25$ –1000 Hz) results in a running salience profile for a particular stimulus. The maximal salience of the pitch salience curve (arrow,  $f_0=200$  Hz) was taken as an estimate of neural harmony, i.e., the degree of harmony for each mistuned stimulus input to the model. See text for details.

this approach considers all relevant harmonics and thus mimics the fact that perceptually, listeners combine information across harmonics to construct a unitary pitch percept (Gockel et al., 2007; Goldstein, 1973; Terhardt et al., 1982). The sieve was composed of 5 Hz wide bins situated at the  $f_0$  of the stimuli (200 Hz) and its integer multiples (i.e.,  $2f_0$ ,  $3f_0$ ,  $4f_0$ ,  $nf_0$ ; for  $nf_0 < 1000$  Hz). For each condition, the degree of  $f_0$  salience was estimated by dividing the mean density of activity falling within the sieve's bins by the mean density of activity in the whole FFT, providing a contrast between activity related to the pitch ( $f_0$ ) and background energy unrelated to pitch (Cedolin and Delgutte, 2005). This measure provides an index of overall harmonicity contained in neural phase-locked responses to complex stimuli (Bidelman and Krishnan, 2009; see also Fig. 2 bottom panel and AN response analyses), i.e., how well brainstem response spectra match a single harmonic series. It should be noted that the sieve analysis assumes that pitch-bearing information in the brainstem frequency following



response (FFR) is largely contained in a temporal code reflecting neural phase-locking to stimulus harmonics. That a temporal code dominates the neural encoding of pitch is consistent with a large body of neurophysiological data, which demonstrate that time-domain neural representations account for a wide variety of perceptual pitch attributes (Bidelman et al., 2011; Bidelman and Heinz, 2011; Cariani and Delgutte, 1996a, 1996b; Cedolin and Delgutte, 2005). However, the brainstem FFR and analysis used here does not address the importance of place cues (i.e., tonotopy), which have been shown to be important in behavioral studies of pitch perception (e.g., Oxenham et al., 2004).

#### 2.4.3. Cortical ERPs

For each participant and condition, the ORN was isolated by subtracting ERP responses to the tuned stimuli (0%) from those elicited by the various mistuned conditions (Alain et al., 2001). From these difference waves, ORN amplitude and latency were measured from ERP traces as the maximum negativity between 125 and 175 ms (Alain et al., 2001, 2012). The N5 wave was recorded as the most prominent negativity between 450 and 550 ms post-stimulus onset from the original ERP waveforms.

### 2.5. Modeling neural correlates of concurrent sound segregation in auditory nerve (AN)

#### 2.5.1. AN model architecture

We used a biologically plausible, computational model of the AN (Ibrahim and Bruce, 2010; Zilany et al., 2009) to investigate the encoding of harmonic stimuli at the earliest stage of neural processing along the auditory pathway. This model represents the latest extension of a well-established model rigorously tested against actual physiological AN responses to tones, broadband noise, speech-like sounds, and musical stimuli (Bidelman and Heinz, 2011; Trainor et al., 2014; Zilany and Bruce, 2006, 2007). The model incorporates several important nonlinearities observed in the auditory periphery, including cochlear filtering, level-dependent gain (i.e., compression) and bandwidth control, as well as two-tone suppression. Recent improvements to the model introduced power-law dynamics and long-term adaptation into the synapse between the inner hair cell and AN fibers, yielding more accurate responses to temporal features of complex sound (e.g., amplitude modulation, forward masking) (Zilany et al., 2009). Model threshold tuning curves have been well fit to the characteristic frequency (CF)-dependent variation in threshold and bandwidth for high-spontaneous rate (SR) fibers in normal-hearing mammals (Miller et al., 1997). Original model parameters were fit to single-unit data recorded in cat (Zilany and Bruce, 2006, 2007). However, more recent modifications (Ibrahim and Bruce, 2010; Zilany et al., 2014)—adopted presently—incorporate human middle-ear filtering (Pascal et al., 1998) and increased basilar membrane frequency selectivity to reflect newer (i.e., sharper) estimates of human cochlear tuning (Joris et al., 2011; Shera et al., 2002). The stochastic nature of AN responses is accounted for by a modified non-homogenous Poisson process, which includes effects of both absolute and relative refractory periods and captures the major stochastic properties of AN responses (e.g., Young and Barta, 1986). A detailed treatment of this model is beyond the scope of this present implementation. Essentially, for a given model neuron and acoustic input (sound waveform measured in Pascals), the model produces a realistic set of spike-trains (i.e., neural action potentials) that closely mimic those obtained from actual animal preparations. A thorough discussion of this computational model and its parameters can be found in the original reports (Zilany and Bruce, 2007; Zilany et al., 2009, 2014).

#### 2.5.2. AN neural harmonicity quantification

A schematic of the various steps in generating and analyzing AN spike data is illustrated in Fig. 2 and mimic our previous modeling studies (Bidelman and Heinz, 2011; Trainor et al., 2014). Identical stimuli to those presented in the behavioral and ERP experiments were used for AN simulations. Single-unit responses were generated by presenting harmonic stimuli to the AN model. An ensemble of 70 high-SR (> 50 spikes/s) AN fibers was simulated with characteristic frequencies (CFs) spaced logarithmically between 125 and 16,000 Hz, covering a broad extent of the cochlear partition. A representative single-fiber post-stimulus time histogram (PSTH) (constructed using 100 repetitions of each stimulus type; 0.1 ms bins; 150 ms epoch window) is shown in Fig. 2 'PSTH.' The PSTH quantifies the time-varying neural discharge pattern of each fiber response. From individual fiber spike trains, first-order interspike interval histograms (ISIHS) were then computed to index pitch periodicities contained in individual fiber responses. Fiber ISIHS were then pooled across all CFs to estimate a population-level ISIH indexing the various periodicities ("all-order" spike intervals) coded across the entire AN ensemble (Cariani and Delgutte, 1996a). Each pooled ISIH was then passed through a harmonic sieve analyzer.

This analysis was similar in nearly all respects to that used for the brainstem ERP data with the exception that the sieve was constructed of harmonically-related pitch periods (harmonics and sub-harmonics) rather than frequency. In other words, it was the time-domain analog of the spectral analysis performed on the brainstem data. The change in analysis domain from time (AN responses) to frequency (brainstem response) was motivated by the fact that the auditory system must convert temporal to rate-place information to counteract the reduction in neural phase-locking along the ascending pathway (Cariani and Delgutte, 1996a). For a given pitch template with fundamental period  $T$  ( $T=1/f_0$ ), sieve windows (100  $\mu$ s wide bins) were situated at  $T$  and its integer multiples (i.e.,  $\dots T/2, T, 2T, \dots, nT$ ), for all  $nT < 50$  ms. The magnitude at the output of a single sieve reflects the salience of pitch-relevant information for the corresponding  $f_0$  pitch contained in the AN response. Analyzing the output across all possible sieve templates ( $f_0=25\text{--}1000$  Hz) results in a running pitch salience curve for a particular stimulus (see also Fig. 9A). The maximal salience of the pitch salience profile was taken as an estimate of neural harmonicity for a given stimulus, a metric closely associated with the degree of neural pitch salience (Bidelman and Heinz, 2011; Cedolin and Delgutte, 2005). Our previous work has shown that this measure of AN harmonicity is strongly associated with complex pitch percepts (Bidelman and Heinz, 2011). In the present study, it is used to assess how neural harmonicity changes in the AN as a function of mistuning. While the AN model does account for the stochastic nature of AN fibers' spontaneous discharge rate, its output is largely deterministic in response to a repeated stimulus. Hence, statistical analysis (e.g., ANOVA) are not appropriate for the AN model data as responses contain no variance in their measurement.

### 2.6. Statistical analysis

One-way, mixed model ANOVAs (SAS<sup>®</sup> 9.4) were conducted on all dependent variables. Stimulus mistuning (6 levels: 0%, 1%, 2%, 4%, 8%, 16%) functioned as the within-subjects factor and participants served as a random factor. Tukey–Kramer multiple comparisons controlled Type I error inflation. Correlational analyses (Spearman's- $r$ ,  $df=58$  and two-tailed tests unless otherwise noted) were used to explore the relationship between brain (AN, brainstem/cortical ERPs) and behavioral (perceptual judgments, RT) measures. An a priori significance level was set at  $\alpha=0.05$ .

### 3. Results

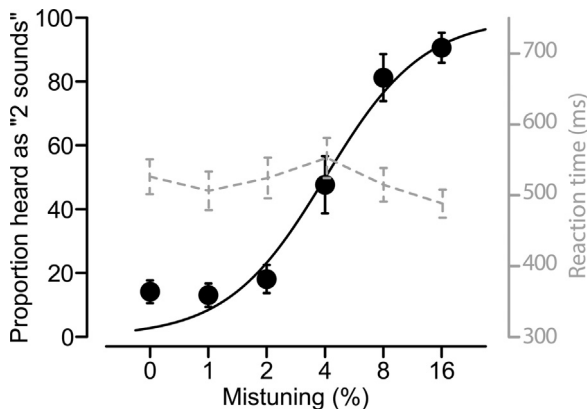
#### 3.1. Behavioral data

Behavioral judgments and corresponding RTs as a function of harmonicity are shown in Fig. 3. An ANOVA revealed a main effect of mistuning on behavioral judgments [ $F_{5, 45}=47.54, p < 0.001$ ]. Participants tended to hear a single auditory object with lower degrees of mistuning (0%, 1%) and multiple sound objects with larger mistuning (8%, 16%). Additionally, we found a significant modulation in the RT of perceptual classification dependent on stimulus mistuning [ $F_{5, 45}=4.24, p=0.003$ ]. This effect was attributable to participants being slower at labeling ambiguous stimuli with moderate levels of mistuning (2%, 4%) relative to those with a clear single (0%) or dual (16%) percept [contrast: (2%, 4%) vs. (0%, 16%);  $t_{45}=2.98, p=0.005$ ].

#### 3.2. Cortical correlates of concurrent sound segregation

Cortical ERPs elicited by the continuum of mistuned stimuli are shown in Fig. 4. Neuroelectric responses revealed typical morphology of the cortical auditory ERPs, featuring prominent waves in the first ~150 ms after sound onset (i.e., P1–N1–P2 complex). Compared to tuned stimuli (0% condition), mistuned stimuli yielded an increased negativity between 100 and 200 ms after sound onset. This component, the object-related negativity (ORN) (Alain et al., 2001), is revealed in the difference wave between ERPs elicited by tuned and mistuned stimuli (dash-dotted trace, Fig. 4B). An ANOVA showed a main effect of harmonic mistuning on ORN magnitudes [ $F_{4, 36}=6.82, p < 0.001$ ], which grew monotonically with increasing inharmonicity of the stimulus [linear contrast:  $t_{36}=4.95, p < 0.001$ ] (Fig. 4C).

In addition to the ORN, a prominent negative deflection appearing at a latency of ~500 ms (N5), followed the salience (i.e., ambiguity) of listeners' perception (Fig. 4D). An ANOVA revealed a main effect of mistuning on N5 magnitude [ $F_{5, 45}=7.75, p < 0.001$ ]. Follow-up contrasts revealed that the N5 wave was most robust for stimuli heard as salient single- or double-object percepts (e.g., 0% and 16%) compared to weak [1 and 8% conditions:  $t_{45}=4.49, p < 0.001$ ] or ambiguous [2 and 4%:  $t_{45}=5.87, p < 0.001$ ] stimulus conditions where listeners are less confident in the number of auditory objects present. This observation was confirmed by a quadratic contrast, which showed that the N5 produced a "U-shaped" pattern of response amplitude [ $t_{45}=6.04, p < 0.001$ ], that



**Fig. 3.** The number of perceived auditory objects depends on stimulus mistuning. Plotted here are the proportion of trials reported as containing "2 sound objects" (black) and corresponding reaction times (gray). Listeners hear only one auditory object for perfectly harmonic tones and those with low degrees of mistuning (0–2%). Perception flips at ~4% mistuning and listeners begin to hear two auditory objects. A strong dual-object percept is obtained at high degrees of mistuning (8–16%). Error bars =  $\pm 1$  s.e.m.

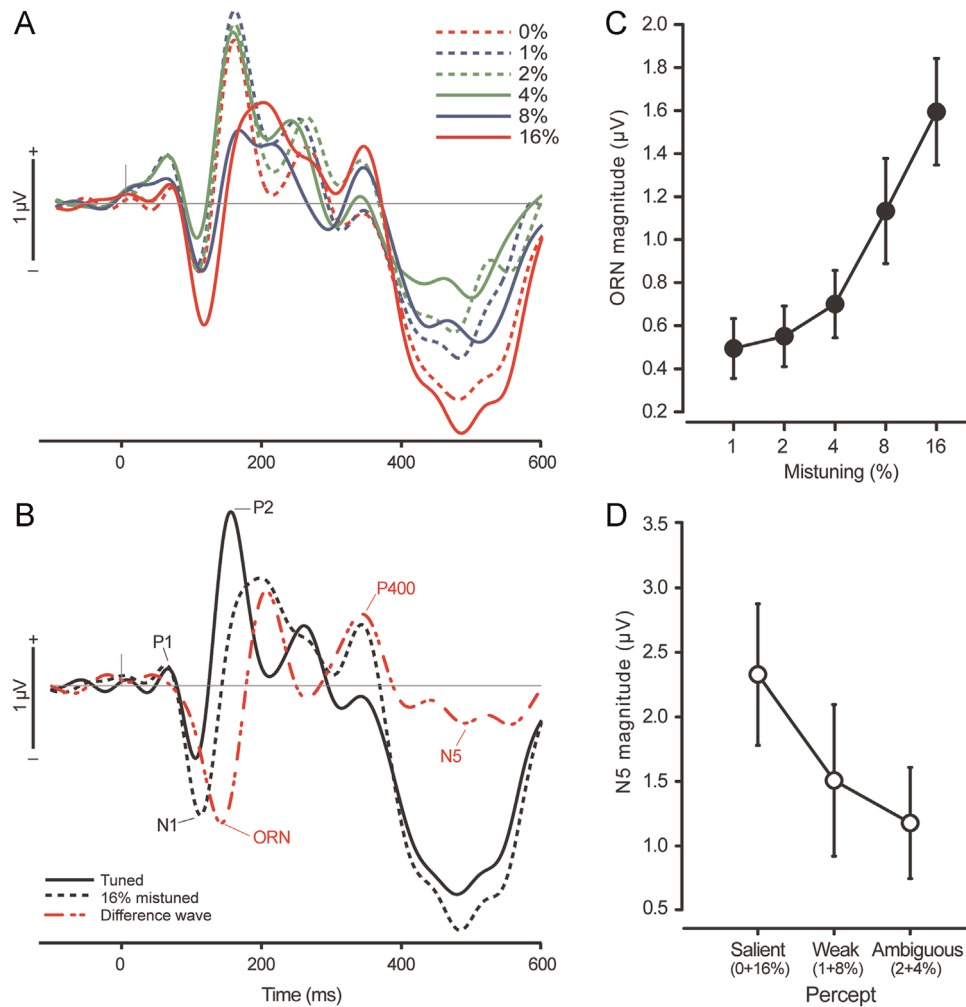
is, the response was strongest for salient stimuli and weaker for more ambiguous, mid-point stimuli along the mistuning continuum. Together, these findings suggest that during concurrent sound segregation, the cortex performs an early registration of inharmonicity as indexed by the continuous coding in the ORN but later signals a more categorical, decision-based response that reflects the salience of the auditory percept (N5 wave).

Scatterplots of listeners' ORN magnitudes plotted against their behavioral responses (across all mistuned harmonic stimuli) are shown in Fig. 5. Listeners' subjective judgment for hearing one or two sound objects was predicted based on their ORN amplitude; pooled across the entire sample, larger cortical responses corresponded with a higher tendency to hear two sound objects [ $r=0.53, p < 0.0001, df=48$ ] (Fig. 5A). Similarly, ORN amplitudes predicted behavioral RTs for determining the number of auditory objects present in the stimulus [ $r=-0.53, p=0.0018, df=48$ ] (Fig. 5B). That is, larger ORN responses were associated with faster behavior classification. To verify that these group-level correspondences also held at the single-subject level, we recomputed ORN-behavior correlations individually for each participant (i.e., within-subjects correlations) and tested the corresponding correlation coefficients against a null hypothesis of zero mean group correlation (one sample *t*-test). Results confirmed a strong correspondence between listeners' individual ORN amplitudes and their subjective behavioral judgments [ $t_9=5.19, p=0.0006$ ] and RTs [ $t_9=-3.29, p=0.0094$ ]. No linear correspondence was observed between N5 amplitude and listeners' psychometric judgments [ $r=0.09, p=0.48$ ] nor their RTs [ $r=-0.07, p=0.57$ ]. Similarly, there were no correlations between the ORN latency and behavioral measures ( $ps > 0.15$ ). However, N5 latency was correlated with behavioral RTs across the aggregate sample [ $r=-0.27, p=0.03$ ] but this correlation was not consistent when assessed via within-subjects analysis [ $t_9=-1.39, p=0.19$ ].

#### 3.3. Brainstem correlates of concurrent sound segregation

Brainstem potentials and their corresponding spectral characteristics are shown in Fig. 6. Time waveforms reveal periodic, phase-locked neural activity that occurs over the duration of the stimulus presentation. Response spectra reveal robust response energy at the  $f_0$  (200 Hz) and its integer multiples. As expected,  $f_0$  amplitude of brainstem responses was invariant to changes in harmonicity [data not shown;  $F_{5, 45}=1.62, p=0.17$ ]. This is consistent with the notion that brainstem  $f_0$  amplitudes reflect the neural encoding for stimulus envelope repetition rate (Gockel et al., 2011; Hall, 1979; Wile and Balaban, 2007) which, despite harmonic mistuning, remains common across our stimulus continuum. In contrast, mistuning the second harmonic (400 Hz) weakened brainstem phase-locking to the upper harmonics of the complex tone at and above the mistuned component (> 400 Hz). The weaker encoding for this pitch-bearing information with increasing mistuning is consistent with the larger pitch ambiguity that arises from mistuned stimuli (Darwin et al., 1994, 1995; Moore et al., 1985). That mistuning-related changes are more prominent in the upper harmonics (but not  $f_0$ ) of brainstem responses is consistent with single-unit recordings in chinchilla AN fibers, which demonstrate a higher sensitivity in the neural encoding of fine structure (cf. upper harmonics) compared to stimulus envelope cues (cf.  $f_0$ ) (Kale et al., 2014). This observation was confirmed by an ANOVA, which revealed a significant effect of mistuning on brainstem neural harmonicity [ $F_{5, 45}=3.03, p=0.0194$ ] (Fig. 6C). Brainstem ERPs contained more robust harmonicity for stimuli containing lower (0–2%) relative to higher (4–16%) degrees of mistuning [contrast (1%, 2%, 4%) vs. (4%, 8%, 16%):  $t_{45}=3.29, p=0.0058$ ].

Lastly, inter- and intra-subject pairwise correlations between



**Fig. 4.** Cortical correlates of concurrent sound identification. (A) Grand averaged ERP responses recorded at a frontal electrode site ( $\sim$ Fpz) for each mistuning condition. (B) Derivation of the object-related negativity (ORN) response. Responses to tuned and 16% mistuned stimuli are shown in solid and dotted traces, respectively. The dash-dotted line shows the ORN, derived as the difference wave between ERPs elicited by the tuned and mistuned stimuli. (C) ORN magnitude grows monotonically with increasing mistuning suggesting this wave indexes the automatic registration of stimulus inharmonicity. (D) The N5 response, a negativity emerging  $\sim$ 500 ms post-stimulus (see B), indexes the salience of the resulting percept. Stimuli yielding strong mono-(0%) or dual-object (16%) percepts elicit larger N5 responses followed by stimuli yielding weaker (1%, 8%) percepts. Ambiguous sounds sometimes heard as containing either one or two auditory objects elicit even smaller N5 responses. Error bars =  $\pm$  1 s.e.m.

brainstem harmonicity and behavioral measures indicated that subcortical activity could predict behavioral judgments of the number of sound objects [ $r = -0.31$ ,  $p = 0.017$ ; intra-subject correlations:  $t_9 = -3.33$ ,  $p = 0.01$ ] but not RTs [ $r = 0.15$ ,  $p = 0.23$ ; intra-subject correlations:  $t_9 = 1.06$ ,  $p = 0.32$ ]. The negative correlation implies that weaker brainstem harmonicity tends to promote the hearing of a multiple auditory objects (i.e., segregated case). The brainstem-behavior link observed here is consistent with previous studies, which demonstrate that consonance judgments of complex musical pitches is predicted based on the neural encoding of harmonicity at the level of the brainstem (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009).

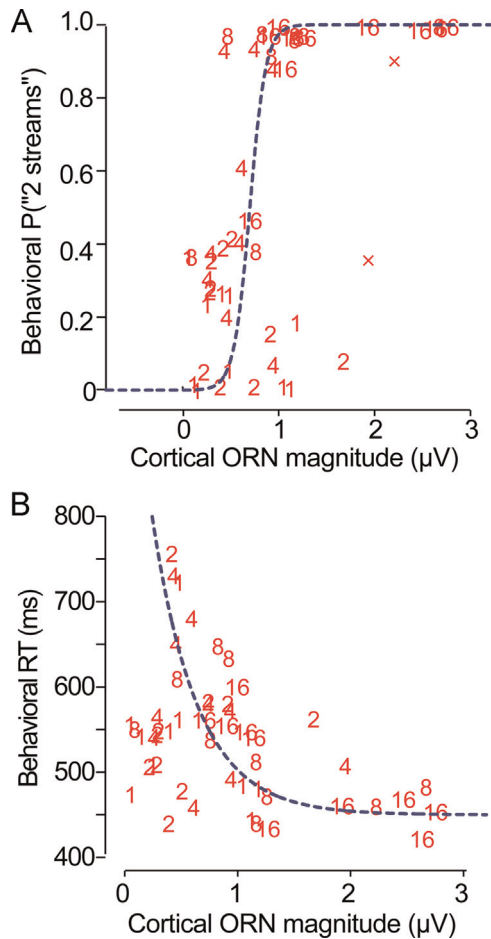
#### 3.4. Correspondence between brainstem/cortical responses and behavior

To assess relations between subcortical and cortical sound encoding, we examined the correlations between brainstem and cortical ERP measures (Fig. 7). Across all mistuning conditions, brainstem harmonicity was negatively correlated with cortical ORN magnitudes [ $r = -0.33$ ,  $p = 0.009$ ]. This group-level (i.e., between subjects) correlation was confirmed with intra-individual correlational analysis, which showed a strong negative correspondence

between listeners' individual brainstem harmonicity and their ORN amplitude [ $t$ -test against zero mean group correlation:  $t_9 = -2.72$ ,  $p = 0.023$ ]. Thus, both inter- and intra-subject analyses revealed a link whereby mistuned harmonic sounds that elicited weaker neural phase-locked responses in the brainstem (i.e., lower neural harmonicity) generated larger ORN amplitudes at a cortical level of processing—reflecting an increased responsiveness to stimulus mistuning.

We used multiple least-squares regression to determine the extent to which brainstem and cortical brain responses could predict behavioral judgments (i.e., reporting “1-” or “2-sounds”) and RT in the concurrent sound classification task (Bidelman et al., 2014b). For each of the two behavioral measures, we constructed a regression model consisting of the main neural effects of interest:  $\Psi = \beta_1 BS + \beta_2 ORN_{amp} + \beta_3 ORN_{lat} + \beta_4 N5_{amp} + \beta_5 N5_{lat}$ , where  $\Psi$  represents a listener's behavioral classification judgment (or RT),  $BS$  is the brainstem neural harmonicity (i.e., Fig. 6C), and  $ORN_{amp/lat}$  and  $N5_{amp/lat}$  are the amplitude and latency of the ORN and N5 waves of the cortical evoked response, respectively (i.e., Fig. 4).  $\beta_1 - \beta_5$  represent to-be-estimated scalar coefficients, computed via least-squares analysis, for the weighting of each of these neural factors in the regression model. Regression coefficients were standardized (total variance = 1) to equate the scales





**Fig. 5.** Brain-behavioral correspondences in concurrent sound identification. (A) Cortical ERP ORN magnitudes predict listeners' subjective behavioral judgments for hearing two auditory objects. Each participant's brain response (across all mistuning conditions) is plotted against their corresponding behavioral response. Individual data points are represented by numbers, corresponding to the stimulus mistuning condition. A positive association indicates that larger ORN responses relate to a higher preponderance for hearing two auditory objects. Dotted line shows a sigmoidal fit to the aggregate data:  $P = 1 / (1 + e^{12(x - 0.7)})$ . Data points denoted as 'x' were excluded prior to curve fitting. (B) ORN magnitudes predict behavioral reaction times during auditory object decisions. On average, larger ORN responses (8%, 16% conditions) are associated with faster decisions in behavior classification. Dotted line shows an exponential decay fit to the aggregate data:  $RT = 500e^{-2.5(x - 0.1)} + 450$ .

between variables and allow us to estimate their individual predictive power on human behavior (Bidelman et al., 2014b). Adjusted  $R^2$  accounts for the number of regressors by penalizing the fit as additional explanatory variables are added to a model and can be useful in feature selection. Thus, unlike standard  $R^2$ , adj- $R^2$  was used to assess model fits as it increases only if additional terms improve a model more than expected by chance.

The resultant regression function for RTs was:  $\psi_{RT} = 0.09BS - 0.40ORN_{amp} + 0.26ORN_{lat} + 0.05N5_{amp} - 0.27N5_{lat}$  (adj- $R^2 = 0.21$ ); the function for the subjective classification judgment was:  $\psi_{Behav} = -0.17BS + 0.59ORN_{amp} - 0.01ORN_{lat} + 0.17N5_{amp} - 0.23N5_{lat}$  (adj- $R^2 = 0.45$ ) (bold coefficients denote significant predictor variables,  $p < 0.05$ ). Together, these models suggest that both the speed of listeners' response and their subjective judgment of concurrent sound objects is at least partially predicted based on the amplitude and latency of listeners' neurophysiological responses. Furthermore, the significant predictive power of the cortical response metrics (ORN and N5 amplitude/latency) suggests that both early (exogenous) and late (endogenous) cortical brain activity can account for listeners'

perception of concurrent sound objects. While brainstem activity provides a robust correlate to index changes in harmonicity (Fig. 6C), it is not as reliable of a predictor of behavior as cortical brain activity.

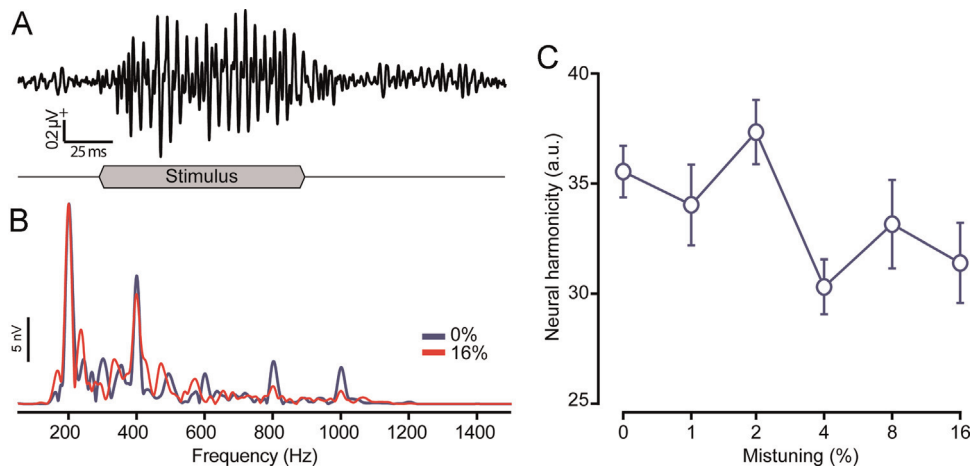
### 3.5. AN correlates of concurrent sound segregation

AN neural harmonicity is shown as a function of mistuning in Fig. 8. Pitch salience profiles index the degree of harmonicity evoked by each stimulus at the level of the auditory nerve. Maximal neural harmonicity was achieved for sieve templates proximal to 200 Hz, the  $f_0$  of our stimulus set. Across the stimulus continuum, neural harmonicity decreases with increasing mistuning paralleling the response pattern observed in brainstem potentials (cf. Fig. 6C). As with brainstem ERPs, this measure of AN harmonicity was inversely related to listeners' behavioral judgments. That is, more harmonic stimuli evoked larger AN responses and were also heard as only a single auditory object; highly mistuned stimuli generated lower AN harmonicity, which coincided with the perception of two auditory objects.

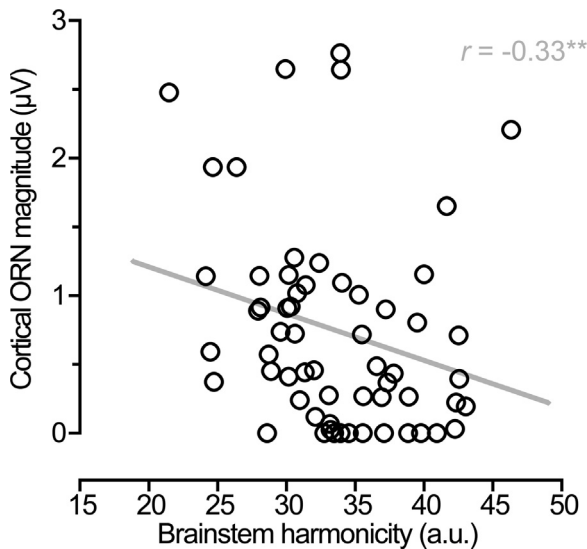
### 3.6. Correspondence between AN, ERP, and behavioral responses

Having established a relationship between brainstem and cortical correlates of harmonic mistuning, we further investigated whether even earlier brain activity (AN) could predict later cortical ORN and subsequent behavioral responses related to concurrent sound perception. Because AN responses could not be measured in our participants, we used the same model simulations for all participants as an approximation of their likely AN activity. AN data was also chosen for the subcortical regressor given that brainstem ERPs (group mean) and AN responses showed a near parallel pattern of change with increasing mistuning (cf. Figs. 6C and 8B) and were correlated with one another [ $r_{AN, BS} = 0.82$ ,  $p = 0.02$ ,  $df = 4$ , one-tailed test]. Correlations between model AN pitch harmonicity, ORN amplitude (group average), and behavioral responses (group average) are shown in Fig. 9. We found significant correspondence between AN and cortical responses (Fig. 9A:  $r_{AN, ORN} = -0.94$ ,  $p = 0.006$ ,  $df = 4$ ), AN and behavioral responses (Fig. 9B:  $r_{AN, Behavior} = -0.93$ ,  $p = 0.009$ ,  $df = 4$ ), and ORN amplitude and behavior (Fig. 9C:  $r_{ORN, Behavior} = 0.91$ ,  $p = 0.01$ ,  $df = 4$ ).

In addition to pairwise bivariate correlations, we examined the correspondence between both neural responses and behavioral judgments when considering all three variables simultaneously. Fig. 9D shows a 3-dimensional scatterplot where each point in space represents the grand average AN, cortical, and behavioral responses to mistuned harmonic stimuli in the  $x$ -,  $y$ -, and  $z$ -dimension, respectively. We fit a plane to the data using multivariate regression with multiple predictor variables as implemented via the "regress" function in MATLAB. This analysis revealed a significant link between all three measures ( $r_{all} = 0.93$ ,  $p = 0.04$ ,  $df = 4$ ). The improvement in predicting listeners' behavior using both neural measures (instead of each singly) was confirmed through the comparison of multiple regression analyses (AN alone: adj- $R^2 = 0.85$ ; ORN alone: adj- $R^2 = 0.84$ ; AN+ORN: adj- $R^2 = 0.88$ ). The significant fit of the plane in multidimensional space implies that a listener's perception of mistuned harmonic stimuli is predicted, at least in part, by how well stimulus (in) harmonicity is encoded at very early subcortical and cortical stages of the auditory pathway. That is, weaker representation of harmonicity in the auditory nerve (and brainstem) and thus stronger ORN responses in cortex are associated with a tendency to hear concurrent sound objects (upper portion of the fitted plane), whereas stronger harmonicity coded in AN yields smaller ORN responses and the perception of only a single sound event (lower portion of the plane).



**Fig. 6.** Brainstem correlates of concurrent sound identification. (A) Prototypical brainstem response time waveform (0% tuned condition). Note the phase-locked (chopping) response, which occurs over the duration of the stimulus presentation. (B) Brainstem response spectra for the tuned (0%) and mistuned (16%) stimuli. Note the response energy at the fundamental frequency ( $f_0$ : 200 Hz) and its integer multiples.  $f_0$  Amplitude—reflecting encoding for the stimulus envelope repetition rate—remains unchanged across mistuning. In contrast, mistuning weakens brainstem phase-locking to the upper harmonics of the complex tone above the mistuned component ( $> 400$  Hz). (C) Neural harmonicity extracted from brainstem response spectra via harmonic sieve analysis (Bidelman and Krishnan, 2009; Cedolin and Delgutte, 2005). Neural harmonicity decreases with increasing mistuning, indexing the larger inharmonicity of these stimuli. Error bars =  $\pm 1$  s.e.m.



**Fig. 7.** Correspondence between subcortical and cortical measures of concurrent sound perception. Brainstem neural harmonicity is negatively associated with cortical ORN amplitude. Each participant's brain response (across all mistuning conditions) is plotted against their corresponding behavior response. Less harmonic, mistuned stimuli are heard as multiple auditory objects. These sounds elicit weaker neural phase-locked responses in the brainstem (lower neural harmonicity), which corresponds with a larger ORN, reflecting the cortical registration of multiple sound objects.  $*p < 0.05$ .

#### 4. Discussion

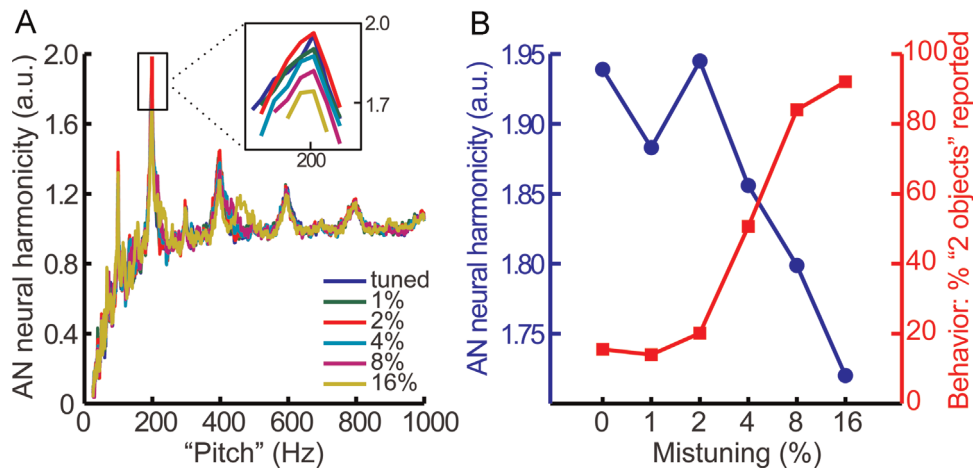
The multifaceted approach used in the present study allows us to bridge findings from animal (e.g., Sinex, 2008; Sinex et al., 2003) and human ERP research (e.g., Alain, 2007; Alain et al., 2001; Dyson and Alain, 2004) by revealing the interplay between subcortical and cortical brain activity associated with concurrent sound perception. Here, we show that concurrent sound perception involves a hierarchy of neurocomputations, recruiting both primitive (e.g., automatic encoding of inharmonicity in the brainstem) and abstract processes (e.g., cortical registration of multi-object salience) operating at multiple levels within the auditory system.

#### 4.1. Subcortical correlates of concurrent sound segregation

Our data show that neural activity in the auditory nerve and brainstem provided an accurate (faithful) neural transcription of stimulus harmonicity, which inversely correlated with human behavior. In both cases, neural harmonicity systematically followed changes in acoustic harmonicity, suggesting that subcortical structures carry adequate information for one of the dominant cues for concurrent sound perception. These results are in line with previous studies examining the neural encoding of vowel phonemes, which demonstrate that pre-attentive brainstem processing provides a high-fidelity representation of the acoustic signal (Bidelman et al., 2013). Moreover, the close parallel between model AN and brainstem response properties further suggests that human brainstem activity, presumably generated in the midbrain inferior colliculi (IC) (Bidelman, under review; Sohmer et al., 1977), maintains stimulus representation already established in earlier peripheral auditory structures. These findings are consistent with single-unit recording in animal models (Pressnitzer et al., 2008), which suggest that the auditory periphery plays a more important role in aspects of complex auditory scene analysis than conventionally thought (cf. Nelken, 2004). Indeed, previous work has revealed that response properties of the cochlear nucleus of the guinea pig contribute to the strength and temporal formation of auditory streaming (i.e., sequential sound segregation) (Pressnitzer et al., 2008). Our findings extend these results by demonstrating similar neural correlates in young adult listeners for concurrent sound segregation beginning at the output of the cochlea. Moreover, these data further argue against the notion that the neural organization of the auditory scene starts at a cortical level (cf. Nelken, 2004).

Harmonic pitch sieve analyses demonstrated that in both AN and brainstem representations, the degree of neural harmonicity was related to stimulus mistuning; high neural harmonicity was related to tuned stimuli that promote single auditory object perception while low harmonicity was related to mistuned stimuli, which listeners hear as multiple objects. These findings indicate that the perception of multiple sound objects could be formed using brainstem information and computing a fit with, or violation from, harmonic templates. Under this premise, the auditory system might decipher the number of auditory objects via a spectral comparator mechanism, which evaluates the incoming acoustic



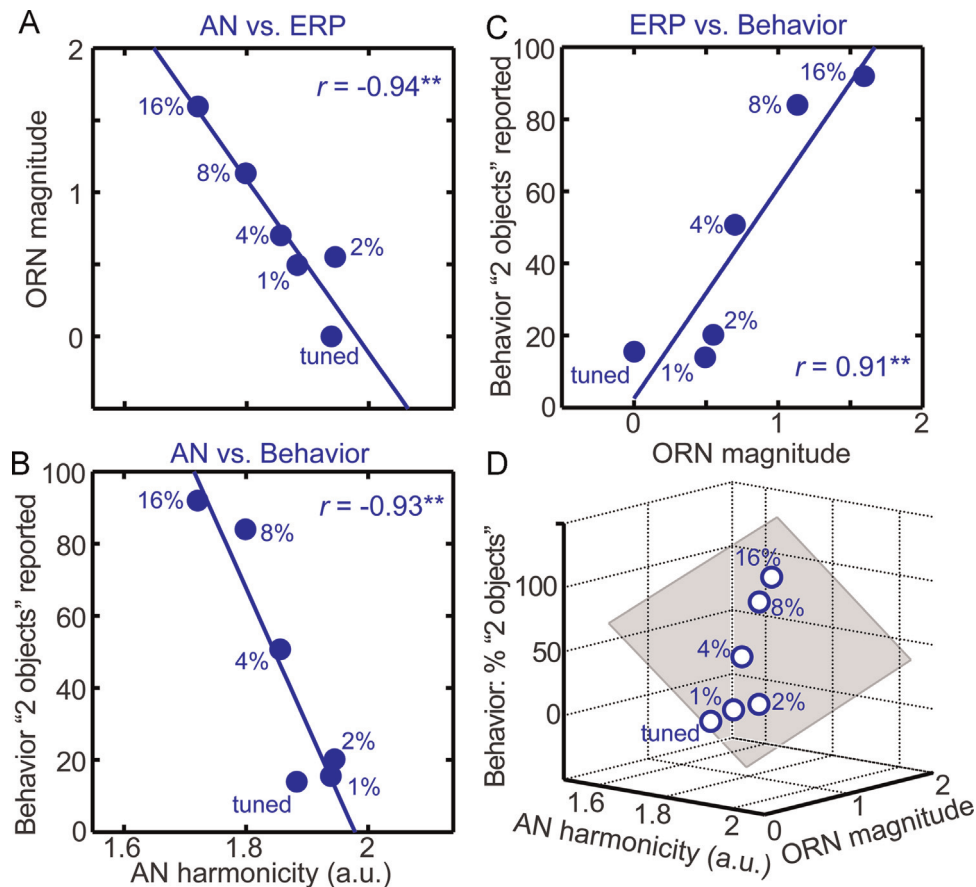


**Fig. 8.** Neural harmonicity computed from AN responses. (A) Pitch saliency profiles quantify the degree of neural harmonicity for all possible pitch-relevant information contained in AN. Maximal saliency occurs near 200 Hz, the  $f_0$  of the acoustic stimuli (see Fig. 1), and is modulated based on the degree of acoustic mistuning (inset). (B) AN response harmonicity shows an inverse relation with listeners' perception. Stimuli with higher mistuning are less proximal to a perfect harmonic series, elicit weaker AN harmonicity, and tend to be heard as multiple auditory objects.

information against internalized pitch templates. Evidence from both animal (Baumann et al., 2011; Langner, 1997) and human electrophysiological data (Barton et al., 2012; Bidelman and Krishnan, 2009; Krishnan et al., 2012) reveal that periodicity cues are maintained (relatively intact but with some attenuation) in

neural phase-locked activity up through the upper brainstem, providing ample information for pitch-related computations.

Previous modeling work has shown that the pitch of a mistuned harmonic complex can be accounted for by modal shifts in the distribution of interspike intervals at the level of the AN; small



**Fig. 9.** Correspondence between auditory nerve, cortical, and behavioral correlates of concurrent sound perception. (A) AN encoding of harmonicity is negatively correlated with ORN magnitudes (group mean, Fig. 4). Stimuli with higher mistuning (8%, 16%) elicit weaker AN harmonicity and larger ORN responses from auditory cortex. (B) Similarly, AN harmonicity is correlated with human behavioral judgments for hearing concurrent sound objects. More acoustically harmonic stimuli (0%, 1%) elicit larger AN responses and are heard as only a single auditory object. (C) Cortical ORN correlates positively with behavior. Larger ORN corresponds with a higher likelihood of hearing two sounds. (D) Multidimensional correlations between AN, cortical, and behavioral processing for concurrent sound perception. Weaker representation of harmonicity in the auditory nerve and thus stronger ORN responses in cortex are associated with a tendency to hear multiple auditory objects (upper portion of the plane), whereas stronger harmonicity coded in AN yields smaller ORNs and the perception of only a single sound event (lower portion of the plane).  $^{**}p < 0.01$ .

mistunings elicit periodic spike times at  $1/f_0$  (i.e., fundamental pitch period) and the sensation of a single pitch whereas large mistuning elicits complex periodicities (jittered around  $1/f_0$ ) that modulate the perceived pitch (Meddis and O'Mard, 1997). Our harmonic pitch sieve approach builds on these studies by incorporating harmonically related phase-locked activity at  $f_0$  as well as its harmonics and thus, is likely more sensitive than approaches based solely on  $f_0$ . The sieve approach was motivated by the fact that listeners combine harmonic information across spectral channels to form a unitary pitch percept (Gockel et al., 2007). Harmonic sieve templates akin to those used in the current study have provided a theoretical model that successfully describe a wide variety of pitch (Cedolin and Delgutte, 2005; Goldstein, 1973) and other auditory percepts, including the perceptual attributes of musical consonance and dissonance (Bidelman and Heinz, 2011; Bidelman and Krishnan, 2009). Perceptual studies estimate that the tolerance of such templates follow a Gaussian function with standard deviation of  $\sim 4\%$  (Darwin et al., 1994; Moore et al., 1985). Interestingly, this falls in line with both our AN and brainstem harmonicity data, which decline at similar degrees of mistuning. That is, mistuning outside  $\sim 3\text{--}4\%$  fails to activate the sieve as they “miss” a harmonically-related bin and thus weaken the salience of the pitch representation.

While the biological basis of such “harmonic-” or “pitch-templates” is somewhat speculative, pitch templates akin to the periodic sieve technique used here have been recently observed in animal studies. Multiunit recordings in marmoset reveal the existence of “harmonic templates” in lateral primary auditory cortex (A1) and surrounding belt areas in the form of pitch sensitive neurons (Bendor and Wang, 2005). These units respond to sounds that elicit similar pitch percepts despite spectral dissimilarity in their physical acoustics (Bendor and Wang, 2005, 2010). Interestingly, these neurons also display multi-peaked tuning with peaks occurring at harmonically-related frequencies (Bendor et al., 2012; see also, Kikuchi et al., 2014). It is conceivable that the multi-peaked nature of these units' response properties could form the physiological basis for a putative harmonic template. While current evidence for harmonic templates in subcortical structures are weak, midbrain neurons do show tuning to best modulation frequencies (i.e., “periodotopy”) (Langner and Schreiner, 1988), mapped orthogonal to the tonotopic frequency axis (Baumann et al., 2011). These organizational characteristics could lay the foundation for harmonic templates that are established at lower, subcortical levels of the auditory system. Indeed, it is postulated that subcortical harmonic templates may arise due to the combined effects of intrinsic cochlear filtering and general neuronal periodicity detection mechanisms (cf. Ebeling, 2008; Shamma and Klein, 2000). In some sense then, concurrent sound segregation (at least with respect to harmonic mistuning) may be somewhat epiphenomenal, a byproduct of the auditory systems' natural tendency to compute and assess violations in harmonic structure or pitch.

#### 4.2. Cortical correlates of concurrent sound segregation

Previous work using scalp-recorded ERPs suggests that the auditory cortex plays a key role in concurrent sound segregation (e.g., Alain, 2007; Alain et al., 2001). Our results replicate and extend these earlier studies by showing a clear relationship between early encoding of (in)harmonicity and ORN amplitude. Multivariate regression analysis revealed that while subcortical responses encoded exogenous stimulus properties and did correlate with perceptual judgments, they were not as good of a predictor of listeners' actual perception as cortical activity (see also, Bidelman et al., 2013; Gockel et al., 2011). In agreement with previous research (Alain et al., 2001; Dyson and Alain, 2004), the

likelihood of hearing two concurrent sound objects increased with increasing inharmonicity and correlated with ORN amplitude. That the ORN indexes concurrent sound percepts is further supported by the correspondence between its amplitude and RTs, which revealed that larger ORN coincided with faster behavioral decisions regarding the number of auditory objects. Moreover, in a prior study, single-trial analysis of ORN responses showed that for an identical yet ambiguous stimulus (4% mistuning), cortical ERP amplitudes were modulated dependent on listeners' perceptual judgment. That is, larger ORN was observed when listeners perceived two sound objects and was smaller when they heard only one (Alain et al., 2013). Collectively, these findings support the notion that even early cortical processing is more than an echo of the stimulus; it reflects listener's subjective auditory experience and perception of the auditory scene (Alain et al., 2013; Bidelman et al., 2013).

Unfortunately, our data cannot fully speak to the underlying mechanisms responsible for generating the changes observed in the scalp-recorded ORN. Physiologically, the ORN could reflect activity of auditory cortical neurons sensitive to inharmonicity (or lack of harmonicity). Noninvasive dipole source modeling of EEG data suggests ORN neural generators near Heschl's gyrus (Arnott et al., 2011). Importantly, the sources of the ORN are different than those found for the obligatory N1 wave (Arnott et al., 2011), which is elicited by transient stimulus onset. Furthermore, a study using functional magnetic resonance imaging while listeners segregated and identified two concurrent vowels revealed recruitment of a thalamo-cortical network comprising the primary auditory cortex when listeners accurately identified both vowels (Alain et al., 2005). Converging evidence from animal research also supports the role of primary auditory cortices in concurrent sound segregation. Indeed, recent work by Fishman and colleagues have explored the neural correlates of concurrent sound segregation based on inharmonicity in A1 of macaque monkeys (Fishman and Steinschneider, 2010; Fishman et al., 2014). Their studies reveal not only neural correlates of “pop out” of mistuned harmonics in A1 (in the form of both increased firing rate and temporal response pattern changes) but also potential intracortical homologs of the ORN localized to A1. Changes in both the local firing rate and temporal discharge pattern with increasing mistuning may underlie the modulations we observe in the far-field, scalp-recorded ORN.

In this regard, it is useful to compare our ORN data with the near-field recordings of Fishman and Steinschneider (2010) who demonstrated enhanced responses (i.e., “pop-out”) to the mistuned harmonic in A1 of non-human primates. Accordingly, when a component is mistuned, one might expect there to be greater power in the spectra of brainstem responses at the frequency of the mistuned component. Yet, we did not observe any indication of “pop-out” in either model AN or human brainstem responses. Consequently, our data lead us to infer that while subcortical representations register stimulus (in-)harmonicity, the *perceptual salience* of segregating sound objects is better captured by later cortical activity (ORN and N5) (see also, Alain et al., 2013). The notion that cortical but not subcortical activity reflects perceptual events is also evident in our regression analyses, which demonstrated that behavioral judgments were predicted by neural activity at a cortical (but not subcortical) level of processing.

In the present study, the ORN was also accompanied by a later negative wave that peaked around 500 ms after sound onset (N5). In contrast to the ORN, this deflection was modulated not by stimulus properties per se, but rather, the perceptual decision. Stimuli offering strong single (0%) or dual object percepts (16%) elicited larger N5 amplitude than those producing weak (1%, 8%) or ambiguous (2%, 4%) percepts. To our knowledge, this is the first study to observe this type of late response in a mistuned harmonic

paradigm. It is interesting to note that this wave follows an attention-dependent component (P400) previously observed in mistuned harmonic paradigms that is associated with perceptual judgment (Alain et al., 2001). The P400 has been linked to controlled processes covering a widely distributed neural network including medial temporal lobe and posterior association cortices (Alain et al., 2001). It is possible that the N5 is somewhat influenced by the presence of the P400, a P3-like component that appears only when listeners actively attend and perform a concurrent sound segregation task (Alain et al., 2001). In the present study we did not analyze the P400 as it was not as apparent in all participants and across conditions as the N5. However, the influence of this component on the N5 is probably minimal as the P400 showed maximal differences at the extreme ends of the continuum (0% and 16% mistuning), yet, these two conditions produced similar N5 responses (Fig. 4A). This indicates at least a partial function independence of these two components. In so much as the N5 wave reflects post-perceptual processing, we speculate that this deflection may index the salience and/or confidence of participants' responses as the amplitude of this wave was greater for stimuli that could easily be classified and systematically diminished with increasing stimulus uncertainty.

#### 4.3. Limitations and future directions

Limitations of the current study are worth noting. Most notably, in the current study, brainstem ERPs were recorded during passive listening whereas cortical ERPs were obtained during an active task. Thus, it is possible that the stronger correspondence between cortical responses and behavior was due to active attention and task engagement in those blocks. An active task would not have been possible during collection of brainstem responses; the number of trials needed to detect the brainstem ERP (many thousands) would have made recording prohibitively slow. While results are equivocal, the majority of prior studies have demonstrated that the human brainstem response is largely unaffected by attention (Galbraith and Kane, 1993; Hillyard and Picton, 1979; Okamoto et al., 2011; Picton and Hillyard, 1974; Picton et al., 1971; Woods and Hillyard, 1978; but see Galbraith et al., 1998; Lehmann and Schonwiesner, 2014), so attentional effects are unlikely to alter our brainstem findings. Similarly, while the cortical auditory ERPs are modulated by attention (e.g., Alain, 2007; Hillyard and Picton, 1979), the ORN response is elicited under both passive and active listening with similar amplitude and latency characteristics (Alain, 2007; Alain et al., 2001). Attention alone is unlikely then to account for the differential effects between our brainstem and ORN data.

Exposure duration also differed between brainstem and cortical ERP blocks, which was longer in the former condition. However, this is unlikely to explain differences between ERP classes given that brainstem responses show much less habituation than cortical potentials and are stable over long periods of time and subject states (Salamy and McKean, 1977). In the current study, the patterns of changes in brainstem and cortical ORN responses did show some similarity, only a difference in sign. This could imply that brainstem codes harmonicity while ORN activity reflects inharmonicity (rather than the absence of a process reflecting harmonicity). While this is open for interpretation, regression analyses revealed that cortical responses were the primary predictor of listeners' perception. Therefore, our data are broadly consistent with the notion that subcortical and cortical auditory processing—while both relevant to building aspects of perception—offer different predictive power and correlations with behavior. A dissociation between cortex and brainstem is further supported by our previous study which demonstrated that listeners' perception of speech material was accounted for by cortical but not

subcortical auditory neural activity (Bidelman et al., 2013). Nevertheless, future studies are needed to more precisely delineate the role of brainstem neural representations in concurrent sound segregation and how attention might modulate various responses across the auditory pathway.

#### 4.4. Tracing the neural chronometry of concurrent sound segregation

Collectively, our findings suggest a series of hierarchical neural computations that are involved in concurrent sound perception. The neural signature of scene analysis seems to begin early in the auditory periphery and is carried forward along the ascending auditory pathway. Continuous coding of harmonicity maintained from the AN and brainstem are then output to early auditory cortical structures which perform an early, pre-attentive registration of acoustic inharmonicity, as indexed by the continuous coding of the ORN. The inverse relation between subcortical and cortical responses suggests a signal transformation between brainstem and the auditory cortex whereby the former codes acoustic periodicity/harmonicity and the later, its complement (i.e., inharmonicity). Later stages of auditory-cognitive processing may then utilize ORN-coded inharmonicity to signal a more categorical, decision-based response built on the salience of the inharmonicity cue (N5 wave).

Together, our findings establish that the earliest reliable changes in neural activity associated with distinguishing simultaneous auditory events occurs milliseconds after sound enters the ear (auditory-nerve) and that these neural correlates are maintained in successively more abstract forms hundreds of milliseconds later. This premise converges with the notion that continuous, yet redundant information relayed from lower levels of the auditory system is successively pruned so as to allow easier readout of signal identity in higher brain areas responsible for generating perception (Chechik et al., 2006). Indeed, multiunit (Perez et al., 2013) and ERP studies (Bidelman et al., 2013, 2014a, 2014b) that have directly compared responses in brainstem and early sensory cortex reveal that the neural code for speech becomes more distinct/abstract as the information ascends the central auditory pathway. It is possible that the current data reflect a similar pruning operation: continuous harmonicity coded in lower structures (AN, brainstem, ORN) is cast into a representation that reflects the salience of eventual percept (e.g., N5 wave).

The notion of a progressive change in neural representations with the increasing level of processing agrees well with single-unit data of Sinex and colleagues, who reported distinct changes in the neural correlates of harmonic mistuning from peripheral to progressively more central auditory nuclei (Sinex, 2008; Sinex et al., 2003, 2002). In their animal studies, auditory nerve coded primarily the individual stimulus components of mistuned complex tones, while more stereotyped responses were observed in mid-brain neurons, where responses diverged from stimulus acoustics (contrasting AN) and largely coded the mistuned feature. The transformation between AN and rostral brainstem representations of harmonic mistuning suggests a fundamental change in neuro-computation between the periphery (AN) and brainstem (IC). In this regard, our ERP data are consistent with the notion of a transformation in the neural correlates of concurrent sound segregation along the ascending auditory pathway. However, in our data, the locus of this transformation appears to reside between the brainstem and cerebral cortex—rather than AN and brainstem as suggested by Sinex et al. (2003). These differences might be attributable to our examination of population level activity—in both AN modeling and scalp-recorded brainstem responses—compared to single-unit data (Sinex et al., 2003, 2002) or the granularity of neural representation that is reflected in the different measurement techniques (their single-unit receptive fields



vs. our far-field potentials). Nevertheless, these studies collectively agree that (i) neural correlates of concurrent sound segregation are present among even the most peripheral auditory structures, brainstem, and cortex and (ii) there is a fundamental transformation in signal processing when traversing progressively higher stages of the auditory pathway.

In summary, our comprehensive approach examining multi-level auditory responses helps reaffirm the notion that the detection of co-occurring auditory objects (i) rests critically on processing of frequency periodicity in both subcortical and cortical stages of the auditory system and (ii) subcortical processing provides an accurate and detailed acoustic analysis while cortical processing plays the dominant role in perceptual analysis of the auditory scene.

## Acknowledgments

Portions of this work were supported by grants from the Canadian Institutes of Health Research (MOP 106619) and the Natural Sciences and Engineering Research Council of Canada (194536) awarded to C.A. and the American Hearing Research Foundation (AHRF) and American Academy of Audiology Foundation (AAF) awarded to G.M.B.

## References

- Alain, C., 2007. Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear. Res.* 229, 225–236.
- Alain, C., Arnott, S.R., Picton, T.W., 2001. Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *J. Exp. Psychol.: Hum. Percept. Perform.* 27, 1072–1089.
- Alain, C., Izenberg, A., 2003. Effects of attentional load on auditory scene analysis. *J. Cogn. Neurosci.* 15, 1063–1073.
- Alain, C., McDonald, K., Van Roon, P., 2012. Effects of age and background noise on processing a mistuned harmonic in an otherwise periodic complex sound. *Hear. Res.* 283, 126–135.
- Alain, C., Reinke, K., McDonald, K.L., Chau, W., Tam, F., Pacurar, A., Graham, S., 2005. Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage* 26, 592–599.
- Alain, C., Roye, A., Arnott, S.R., 2013. Middle- and long-latency auditory evoked potentials: What are they telling us on central auditory disorders?. In: Ceesia, G.G. (Ed.), *Handbook of Clinical Neurophysiology: Disorders of Peripheral and Central Auditory Processing*, vol. 10. Elsevier, London, United Kingdom, pp. 177–199.
- Alain, C., Schuler, B.M., McDonald, K.L., 2002. Neural activity associated with distinguishing concurrent auditory objects. *J. Acoust. Soc. Am.* 111, 990–995.
- Arnott, S.R., Bardouille, T., Ross, B., Alain, C., 2011. Neural generators underlying concurrent sound segregation. *Brain Res.* 1387, 116–124.
- Barton, B., Venezia, J.H., Saberi, K., Hickok, G., Brewer, A.A., 2012. Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc. Natl. Acad. Sci. USA* 109, 20738–20743.
- Baumann, S., Griffiths, T.D., Sun, L., Petkov, C.I., Thiele, A., Rees, A., 2011. Orthogonal representation of sound dimensions in the primate midbrain. *Nat. Neurosci.* 14, 423–425.
- Bendor, D., Osmani, M.S., Wang, X., 2012. Dual-pitch processing mechanisms in primate auditory cortex. *J. Neurosci.* 32, 16149–16161.
- Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Bendor, D., Wang, X., 2010. Neural coding of periodicity in marmoset auditory cortex. *J. Neurophysiol.* 103, 1809–1822.
- Bidelman, G.M. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hear. Res.* (under review).
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434.
- Bidelman, G.M., Heinz, M.G., 2011. Auditory-nerve responses predict pitch attributes related to musical consonance–dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.* 130, 1488–1502.
- Bidelman, G.M., Krishnan, A., 2009. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171.
- Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212.
- Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014a. Age-related changes in the subcortical–cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35, 2526–2540.
- Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014b. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673.
- Binns, C., Culling, J.F., 2007. The role of fundamental frequency contours in the perception of speech against interfering speech. *J. Acoust. Soc. Am.* 128, 1765–1776.
- Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J. Neurophysiol.* 76, 1717–1734.
- Carlyon, R.P., 2004. How the brain separates sounds. *Trends Cogn. Sci.* 8, 465–471.
- Cedolin, L., Delgutte, B., 2005. Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J. Neurophysiol.* 94, 347–362.
- Chechik, G., Anderson, M.J., Bar-Yosef, O., Young, E.D., Tishby, N., Nelken, I., 2006. Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51, 359–368.
- Ciocca, V., 2008. The auditory organization of complex sounds. *Front. Biosci.* 13, 148–169.
- Darwin, C.J., Ciocca, V., Sandell, G.R., 1994. Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic. *J. Acoust. Soc. Am.* 95, 2631–2636.
- Darwin, C.J., Hukin, R.W., Al-Khatib, B.Y., 1995. Grouping in pitch perception: evidence for sequential constraints. *J. Acoust. Soc. Am.* 98, 880.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21.
- Dyson, B.J., Alain, C., 2004. Representation of concurrent acoustic objects in primary auditory cortex. *J. Acoust. Soc. Am.* 115, 280–288.
- Dyson, B.J., Alain, C., He, Y., 2005. I've heard it all before: perceptual invariance represented by early cortical auditory-evoked responses. *Brain Res. Cogn. Brain Res.* 23, 457–460.
- Ebeling, M., 2008. Neuronal periodicity detection as a basis for the perception of consonance: a mathematical model of tonal fusion. *J. Acoust. Soc. Am.* 124, 2320–2329.
- Fishman, Y.I., Steinschneider, M., 2010. Neural correlates of auditory scene analysis based on inharmonicity in monkey primary auditory cortex. *J. Neurosci.* 30, 12480–12494.
- Fishman, Y.I., Steinschneider, M., Micheyl, C., 2014. Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: implications for models of auditory scene analysis. *J. Neurosci.* 34, 12425–12443.
- Galbraith, G., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen Jr., T.A., 1998. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893.
- Galbraith, G.C., Kane, J.M., 1993. Brainstem frequency-following responses and cortical event-related potentials during attention. *Percept. Mot. Skills* 76, 1231–1241.
- Gockel, H.E., Carlyon, R.P., Mehta, A., Plack, C.J., 2011. The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *J. Assoc. Res. Otolaryngol.* 12, 767–782.
- Gockel, H.E., Moore, B.C., Carlyon, R.P., Plack, C.J., 2007. Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *J. Acoust. Soc. Am.* 121, 373–382.
- Goldstein, J.L., 1973. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54, 1496–1516.
- Hall 3rd, J.W., 1979. Auditory brainstem frequency following responses to waveform envelope periodicity. *Science* 205, 1297–1299.
- Hartmann, W.M., McAdams, S., Smith, B.K., 1990. Hearing a mistuned harmonic in an otherwise periodic complex tone. *J. Acoust. Soc. Am.* 88, 1712–1724.
- Hautus, M.J., Johnson, B.W., 2005. Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J. Acoust. Soc. Am.* 117, 275–280.
- Hillyard, S.A., Picton, T.W., 1979. Event-related brain potentials and selective information processing in man. In: Desmedt, J.E. (Ed.), *Progress in Clinical Neurophysiology*. Karger, Basel, pp. 1–52.
- Ibrahim, R.A., Bruce, I.C., 2010. Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In: Lopez-Poveda, E.A., Palmer, A.R., Meddis, R. (Eds.), *The Neurophysiological Bases of Auditory Perception*. Springer, New York, pp. 429–438.
- Joris, P.X., Bergevin, C., Kalluri, R., McLaughlin, M., Michelet, P., van der Heijden, M., Spera, C.A., 2011. Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc. Natl. Acad. Sci.* 108, 17516–17520.
- Kale, S., Micheyl, C., Heinz, M.G., 2014. Implications of within-fiber temporal coding for perceptual studies of f0 discrimination and discrimination of harmonic and inharmonic tone complexes. *J. Assoc. Res. Otolaryngol.* 15, 465–482.
- Kikuchi, Y., Horwitz, B., Mishkin, M., Rauschecker, J.P., 2014. Processing of harmonics in the lateral belt of macaque auditory cortex. *Front. Neurosci.* 8, 1–13.
- Krishnan, A., Bidelman, G.M., Smalt, C.J., Ananthakrishnan, S., Gandour, J.T., 2012. Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia* 50, 2849–2859.
- Langner, G., 1997. Neural processing and representation of periodicity pitch. *Acta Oto-laryngol. (Supplement)* 532, 68–76.
- Langner, G., Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60, 1799–1822.
- Lehmann, A., Schonwiesner, M., 2014. Selective attention modulates human



- auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS One* 9, e85442.
- Lin, J.Y., Hartmann, W.M., 1998. The pitch of a mistuned harmonic: evidence for a template model. *J. Acoust. Soc. Am.* 103, 2608–2617.
- Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820.
- Miller, R.L., Schilling, J.R., Franck, K.R., Young, E.D., 1997. Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers. *J. Acoust. Soc. Am.* 101, 3602–3616.
- Moore, B.C., Glasberg, B.R., Peters, R.W., 1986. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* 80, 479–483.
- Moore, B.C.J., Glasberg, B.R., Peters, R.W., 1985. Relative dominance of individual partials in determining the pitch of complex tones. *J. Acoust. Soc. Am.* 77, 1853–1860.
- Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42.
- Nelken, I., 2004. Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480.
- Okamoto, H., Stracke, H., Bermudez, P., Pantev, C., 2011. Sound processing hierarchy within human auditory cortex. *J. Cogn. Neurosci.* 23, 1855–1863.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Oxenham, A.J., Bernstein, J.G.W., Penagos, H., 2004. Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. USA* 101, 1421–1425.
- Pascal, J., Bourgeade, A., Lagier, M., Legros, C., 1998. Linear and nonlinear model of the human middle ear. *J. Acoust. Soc. Am.* 104, 1509–1516.
- Perez, C.A., Engineer, C.T., Jakkamsetti, V., Carraway, R.S., Perry, M.S., Kilgard, M.P., 2013. Different timescales for the neural coding of consonant and vowel sounds. *Cereb. Cortex* 23, 670–683.
- Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr. Clin. Neurophysiol.* 36, 191–199.
- Picton, T.W., Hillyard, S.A., Galambos, R., Schiff, M., 1971. Human auditory attention: a central or peripheral process? *Science* 173, 351–353.
- Pressnitzer, D., Sayles, M., Micheyl, C., Winter, I.M., 2008. Perceptual organization of sound begins in the auditory periphery. *Curr. Biol.* 18, 1124–1128.
- Roberts, B., 1998. Effects of spectral pattern on the perceptual salience of partials in harmonic and frequency-shifted complex tones: a performance measure. *J. Acoust. Soc. Am.* 103, 3588–3596.
- Roberts, B., Brunstrom, J.M., 1998. Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. *J. Acoust. Soc. Am.* 104, 2326–2338.
- Salamy, A., McKean, C.M., 1977. Habituation and dishabituation of cortical and brainstem evoked potentials. *Int. J. Neurosci.* 7, 175–182.
- Shamma, S., Klein, D., 2000. The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.* 107, 2631–2644.
- Shera, C.A., Guinan Jr., J.J., Oxenham, A.J., 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci. USA* 99, 3318–3323.
- Sinex, D.G., 2008. Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. *Hear. Res.* 238, 39–48.
- Sinex, D.G., Guzik, H., Li, H., Henderson, S.J., 2003. Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res.* 182, 130–139.
- Sinex, D.G., Li, H., Velenovsky, D.S., 2005. Prevalence of stereotypical responses to mistuned complex tones in the inferior colliculus. *J. Neurophysiol.* 94, 3523–3537.
- Sinex, D.G., Sabes, J.H., Li, H., 2002. Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hear. Res.* 168, 150–162.
- Sohmer, H., Pratt, H., Kinarti, R., 1977. Sources of frequency-following responses (FFR) in man. *Electroencephalogr. Clin. Neurophysiol.* 42, 656–664.
- Swaminathan, J., Heinz, M.G., 2012. Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *J. Neurosci.* 32, 1747–1756.
- Terhardt, E., Stoll, G., Seewann, M., 1982. Algorithm for the extraction of pitch and pitch salience from complex tonal signals. *J. Acoust. Soc. Am.* 71, 679–687.
- Trainor, L.J., Marie, C., Bruce, I.C., Bidelman, G.M., 2014. Explaining the high voice superiority effect in polyphonic music: evidence from cortical evoked potentials and peripheral auditory models. *Hear. Res.* 308, 60–70.
- Wile, D., Balaban, E., 2007. An auditory neural correlate suggests a mechanism underlying holistic pitch perception. *PLoS One* 2, e369.
- Woods, D.L., Hillyard, S.A., 1978. Attention at the cocktail party: brainstem evoked responses reveal no peripheral gating. In: Otto, D.A. (Ed.), *Multidisciplinary Perspectives in Event-Related Brain Potential Research (EPA 600/9-77-043)*. U.S. Government Printing Office, Washington, DC, pp. 230–233.
- Young, E.D., Barta, P.E., 1986. Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J. Acoust. Soc. Am.* 79, 426–442.
- Zendel, B.R., Alain, C., 2009. Concurrent sound segregation is enhanced in musicians. *J. Cogn. Neurosci.* 21, 1488–1498.
- Zendel, B.R., Alain, C., 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *J. Cogn. Neurosci.* 25, 503–516.
- Zilany, M.S., Bruce, I.C., 2006. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J. Acoust. Soc. Am.* 120, 1446–1466.
- Zilany, M.S., Bruce, I.C., 2007. Representation of the vowel /e/ in normal and impaired auditory nerve fibers: model predictions of responses in cats. *J. Acoust. Soc. Am.* 122, 402–417.
- Zilany, M.S., Bruce, I.C., Nelson, P.C., Carney, L.H., 2009. A phenomenological model of the synapse between the inner hair cell and auditory nerve: long-term adaptation with power-law dynamics. *J. Acoust. Soc. Am.* 126, 2390–2412.
- Zilany, M.S.A., Bruce, I.C., Carney, L.H., 2014. Updated parameters and expanded simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.* 135, 283–286.