

## ORIGINAL ARTICLE

# Individual Differences in Human Auditory Processing: Insights From Single-Trial Auditory Midbrain Activity in an Animal Model

Travis White-Schwoch<sup>1</sup>, Trent Nicol<sup>1</sup>, Catherine M. Warrier<sup>1</sup>,  
Daniel A. Abrams<sup>1,4</sup>, and Nina Kraus<sup>1,2,3</sup>

<sup>1</sup>Auditory Neuroscience Laboratory ([www.brainvolts.northwestern.edu](http://www.brainvolts.northwestern.edu)) & Department of Communication Sciences, Northwestern University, Evanston, IL, 60208, USA, <sup>2</sup>Department of Neurobiology & Physiology, Northwestern University, Evanston, IL, 60208, USA, <sup>3</sup>Department of Otolaryngology, Northwestern University, Chicago, IL, 60611, USA, and <sup>4</sup>Present address: Stanford Cognitive & Systems Neuroscience Laboratory, Stanford University, Palo Alto, CA, 94304, USA

Address correspondence to Nina Kraus, 2240 Campus Dr., Evanston, IL 60208, USA. [www.brainvolts.northwestern.edu](http://www.brainvolts.northwestern.edu). Email: [nkraus@northwestern.edu](mailto:nkraus@northwestern.edu)

## Abstract

Auditory-evoked potentials are classically defined as the summations of synchronous firing along the auditory neuraxis. Converging evidence supports a model whereby timing jitter in neural coding compromises listening and causes variable scalp-recorded potentials. Yet the intrinsic noise of human scalp recordings precludes a full understanding of the biological origins of individual differences in listening skills. To delineate the mechanisms contributing to these phenomena, *in vivo* extracellular activity was recorded from inferior colliculus in guinea pigs to speech in quiet and noise. Here we show that trial-by-trial timing jitter is a mechanism contributing to auditory response variability. Identical variability patterns were observed in scalp recordings in human children, implicating jittered timing as a factor underlying reduced coding of dynamic speech features and speech in noise. Moreover, intertrial variability in human listeners is tied to language development. Together, these findings suggest that variable timing in inferior colliculus blurs the neural coding of speech in noise, and propose a consequence of this timing jitter for human behavior. These results hint both at the mechanisms underlying speech processing in general, and at what may go awry in individuals with listening difficulties.

**Key words:** auditory midbrain, auditory processing, development, neural variability, speech in noise

## Introduction

Neurophysiological responses to evoked stimuli have provided a window into individual differences in sensory processing, revealing the imprint of experience, the impact of language and learning problems, and insight into future linguistic abilities and disabilities. These approaches have proven especially fruitful in the auditory system, where speech-evoked neurophysiological responses offer a noninvasive means to evaluate the integrity of neural processing in humans. In particular, the

frequency-following response (FFR) to speech evaluates the precision of time-locked neurophysiological activity elicited by features in the speech signal.

The FFR is generated predominantly by inferior colliculus (IC), the lemniscal midbrain nucleus of the ascending auditory system (for review see [Chandrasekaran and Kraus 2010](#)). An important metric quantifies the intertrial variability of the FFR, which is based on the assumption that variability across responses to the same sound reflects subtle dyssynchronies in

neural firing. This approach effectively characterizes FFRs in individuals with listening difficulties, and suggests that high intertrial variability represents a key neurophysiological mechanism underlying atypical perception in these individuals. For example, variable responses to speech are evident both in children with reading impairment (Hornickel and Kraus 2013; White-Schwoch et al. 2015b) and older adults (Anderson et al. 2012; Ruggles et al. 2012), populations that often experience speech perception difficulties in adverse listening conditions (Dubno et al. 1984; Pichora-Fuller et al. 1995; Bradlow et al. 2003; Ziegler et al. 2005; Gordon-Salant et al. 2010). Extreme variability in subcortical neural firing (auditory neuropathy) causes extreme difficulties understanding speech in noise and rapid temporal cues (Zeng et al. 1999; Kraus et al. 2000).

Animal models support the hypothesis that poor auditory processing can be grounded in imprecise temporal resolution in the central auditory system. Rodent models of auditory aging document extensive declines in inhibitory neurotransmitter receptors in auditory midbrain, thalamus, and cortex, presumably degrading synchronous firing across units (Casparly et al. 1995, 2008, 2013; Milbrandt et al. 1997; Walton et al. 1998; Richardson et al. 2013) and constraining the neural representation of dynamic acoustic cues (Parthasarathy and Bartlett 2011; Cai and Casparly 2015; see also Anderson et al. 2012; Presacco et al. 2015). Similarly, rat models of language impairment, including dyslexia and autism, exhibit less precise spike timing, smaller local field potentials, and poorer physiological categorization of human speech in auditory cortex (Engineer et al. 2014, 2015; Centanni et al. 2014a, 2014b).

Similar phenomena may underlie the difficulty some listeners experience understanding speech in noise, despite normal audiograms. Among these individuals are certain children with language-based learning disabilities (Wright et al. 1997; Bradlow et al. 2003; Ziegler et al. 2005; but see Messaoud-Galusi et al. 2011) and older adults (Pichora-Fuller et al. 1995; Gordon-Salant 2014; Füllgrabe et al. 2015; but see Schoof and Rosen 2014). It has been proposed that variable neural firing underlies these perceptual difficulties, based on the aforementioned neurophysiological approaches (Anderson et al. 2012; Hornickel and Kraus 2013; White-Schwoch et al. 2015a; 2015b) and behavioral evidence that “jittering” periodicity cues in speech tokens impairs young adults’ recognition such that they perform similarly to older adults (Pichora-Fuller et al. 2007). Also noteworthy is that individuals with autism—another population that often experiences listening difficulties (Gervais et al. 2004; Kuhl et al. 2005; Russo et al. 2008; Abrams et al. 2013), including processing speech in noise (Alcántara et al. 2004; Russo et al. 2009)—exhibit variable evoked responses to auditory, visual, and somatosensory stimuli (Dinstein et al. 2012). Thus, variability in sensory coding may be an intrinsic biological constraint on information processing.

Delineating the mechanisms underlying FFR variability has proven challenging given the inherent noise in scalp recordings—these recordings typically require averaging across hundreds, if not thousands, of trials to compute an interpretable evoked response. This limitation precludes a full understanding of the mechanisms underlying FFR variability and its link to auditory perception. Thus, the overarching aim of this report is to investigate local mechanisms contributing to scalp-recorded response properties and, by extension, the biological phenomena that may contribute to individual differences in auditory processing.

A first goal of this report is to directly compare near-field (depth-recorded) evoked responses in an animal model to

far-field (scalp-recorded) responses in humans as a function of stimulus conditions. Animal models offer insight into local neural activity that is unavailable in humans. Thus, investigations in animal models facilitate knowledge of the basic mechanisms underlying response patterns observed in humans. When these response patterns are disrupted in clinical populations, relating them back to “typical” animal models may lead to a better understanding of the specific physiological processes that have gone awry. These comparisons ideally minimize methodological differences between species. To this end, we investigated neurophysiological responses in the guinea pig, an animal that exhibits hearing sensitivity comparable to humans within the frequency spectrum of speech (Fay 1988). This allowed us to deliver identical stimuli to animal and human subjects. We also strived to standardize physiological recording and analysis methodologies across the human and animal study components.

A second goal of this report is to investigate behavioral ramifications of intertrial timing variability in response to speech heard in background noise in children. Our previous work established that preschoolers with poor early language skills exhibit more variable neurophysiological responses to consonant-vowel (CV) transitions in noise than their peers (White-Schwoch et al. 2015b). This aligns with evidence that children with language or literacy impairment exhibit a similar profile (Hornickel et al. 2012; Hornickel and Kraus 2013), and with evidence that a rat model of dyslexia exhibits variable cortical responses to consonant-vowel-consonant speech sounds (Centanni et al. 2014a, 2014b). Here, we aim to extend these findings in humans and tie them to work in the animal model. To our knowledge, this is one of the first efforts to directly compare neurophysiological response variability between humans and an animal model, and thus has potential to conceptually “bridge” local activity, scalp recordings, and emergent language skills.

We employed a comparative neurophysiological approach by eliciting responses to identical speech-like stimuli in quiet and background noise in an animal model and in humans. Although there are striking similarities noted in response patterns to speech sounds between human scalp recordings and near-field recordings in animal models, to our knowledge these have not been directly compared. This knowledge gap limits our understanding of the neural events that may contribute to auditory processing and disorders thereof.

Our central hypothesis is that, in auditory midbrain, timing variability blurs the neural representation of perceptually vulnerable speech features in noise. In turn, we hypothesize that children who exhibit excessive response variability in noise lag behind their peers in language development. Comparisons of patterns of intertrial variability across species may elucidate the mechanisms contributing to aggregate population responses in humans and, by extension, auditory processing abilities and disabilities.

## Materials and Methods

Auditory-neurophysiological responses were elicited to synthesized speech in quiet and noise in human listeners and an animal model. Identical stimuli were used in both study components. In animal experiments, *in vivo* extracellular activity was recorded from several sites in the central nucleus of inferior colliculus (ICc) across laminae, along with simultaneous recordings at the epidural surface. In a parallel human experiment, scalp-recorded responses (FFRs; also known as

auditory brainstem responses to complex sounds, or cABRs) were elicited from a cohort of young children (ages 3–5 years).

Using preschool-aged children as the human cohort presents several advantages. Young children tend to have large FFRs that are likely to be interpretable when subaveraged (see [Data Analysis](#), later). Moreover, given evidence that a life of noise exposure may cause neural degeneration prior to a threshold shift on the audiogram ([Sergeyenko et al. 2013](#)), and that this degeneration may shape FFR response properties ([Ruggles et al. 2012](#); [Plack et al. 2014](#)), using young listeners provides some safeguards against this potential factor. Furthermore, biological mechanisms underlying children's auditory function are of particular theoretical and clinical interest because remediation to improve listening and language skills may be most efficacious during early childhood ([Bishop and Adams 1990](#)), which is an established period of developmental neuroplasticity in human auditory midbrain ([Johnson et al. 2008](#); [Anderson et al. 2015](#); [Skoe et al. 2015](#)).

These children were also tested on a standardized test of phonological processing (knowledge and manipulation of the sound structure of spoken language). Phonological processing is a chief primitive in language development, and phonological deficits are observed in many children with language impairment and/or poor literacy achievement ([Bishop 1997](#)); in many cases, these language impairments overlap with listening difficulties ([Tallal and Piercy 1973](#); [Moore et al. 2010](#); but see [Rosen 2003](#); [Goswami 2014](#)).

Two aspects of neural coding are of interest, both of which quantify variability in neural coding across trials. The first is termed “representational variability” (henceforth, variability) and quantifies the morphological dissimilarity of responses across trials. The second is “timing variability” (henceforth, jitter) and quantifies the timing dissimilarity of responses across trials.

## Ethical Statement

The animal protocol was approved by the Institutional Animal Care and Use Committee of Northwestern University, pursuant to all United States ethical guidelines for laboratory animal welfare (assurance number A3282-01). The human protocol was approved by the Institutional Review Board of Northwestern University, pursuant to the Declaration of Helsinki (assurance number FWA 00001549); parents provided written informed consent and children provided verbal assent.

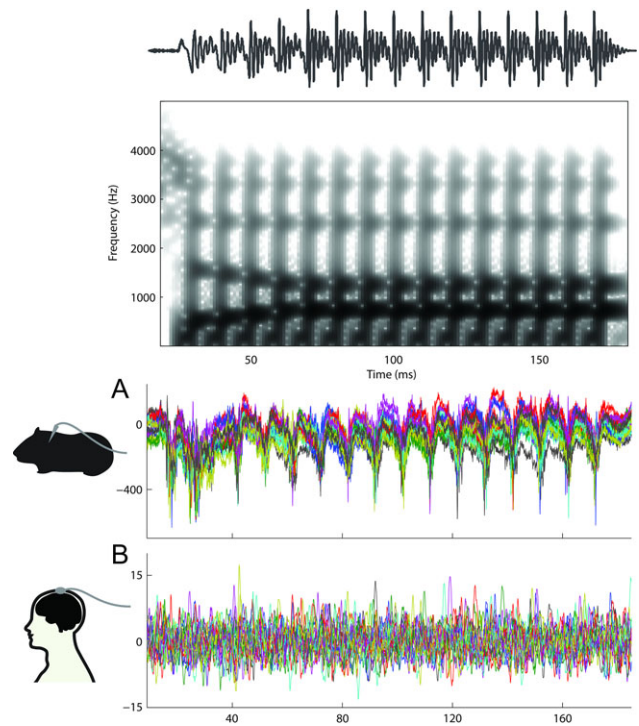
## Stimulus

The stimulus for both the animal and human experiments was a CV syllable [da]. The rationale for using this stimulus is that stop consonants such as /d/ pose special challenges to several groups of listeners, including children with learning problems and older adults. Moreover, it has long been noted that the perceptual challenges posed by stop consonants are exacerbated in noise ([Miller and Nicely 1955](#); see also [Cunningham et al. 2001, 2002](#); [White-Schwoch et al. 2015a](#)). With respect to FFR intertrial variability in humans, previous studies have demonstrated that children with learning problems exhibit excessive variability in response to stop consonants, including in noise ([White-Schwoch et al. 2015b](#); see also [Centanni et al. 2014a, 2014b](#)). Although these sounds are slightly lower in frequency than a typical guinea pig vocalization, they are still well within the range of audibility ([Fay 1988](#)); historically, this approach has proven fruitful in

evaluating how auditory-neurophysiological activity changes along acoustic dimensions that are behaviorally salient to human listeners ([Kraus et al. 1994](#); [McGee et al. 1996](#); [Cunningham et al. 2002](#); [Warrier et al. 2011](#)).

A single [da] token was used, and this stimulus was repeated across trials to determine the extent to which the auditory system represents this stimulus invariably. The [da] was a synthesized 170 ms 6-formant syllable that began with a stop onset burst and had a 5 ms voice onset time (Klatt-based synthesizer, SenSyn, Sensimetrics Corporation, Malden, MA). The CV transition lasted 50 ms and this period was followed by a 120 ms steady-state vowel. The fundamental frequency ( $F_0$ ) was fixed at 100 Hz. During the CV transition, the lower 3 formants shifted ( $F_1$ , 400–720 Hz;  $F_2$ , 1700–1240 Hz;  $F_3$ , 2850–2500 Hz) but were steady for the vowel period. The upper 3 formants were steady throughout the stimulus ( $F_4$ , 3300 Hz;  $F_5$ , 3750 Hz;  $F_6$ , 4900 Hz). A waveform of the [da] is presented in [Figure 1](#) along with a spectrogram (generated in Luscinia, <http://rflachlan.github.io/Luscinia>; last accessed 17 September 2016). On half of the trials the polarity of the [da] was inverted (stimulus waveform multiplied by  $-1$ ) to avoid stimulus artifacts in the neural responses.

The [da] was presented in isolation (the “quiet” condition) and masked by multitalker babble (the “noise” condition). The babble track consisted of 6 talkers and was presented continuously over the speech token. Please refer to [Van Engen and Bradlow \(2007\)](#) for details on its acoustics. Although a single babble track is used and repeated, it is 22 s long (compared with



**Figure 1.** (Top) The waveform of the [da] stimulus is shown in the time domain along with a spectrogram. (A) Single traces of near-field activity recorded from central nucleus of inferior colliculus are illustrated—20 randomly selected trials have been overlaid, each in a different color. Single-trial activity showed phase-locked activity to the stimulus (top) and had a high signal-to-noise ratio. (B) In contrast, recordings made at the human scalp did not have adequate single-trial SNR and had to be averaged across thousands of repetitions. An equal number of randomly selected trials from a human subject are shown here to illustrate that these do not have the SNR appropriate to analyze single trials.



the 170 ms stimulus). The babble track is looped continuously to avoid phase synchrony between the [da] and the babble track. In both the animal and human components of this study, in the noise condition the [da] was presented at a +10 dB signal-to-noise (SNR) ratio.

## Animal Component

### Subjects and Preparation

The subjects in the animal component were 10 pigmented guinea pigs (*Cavia porcellus*; 7 female) weighing 346–803 g (mean, 549 g). Animals were anesthetized before recording with a ketamine hydrochloride (60 mg/kg) and xylazine (8 mg/kg) cocktail. Supplemental doses (15 mg/kg ketamine; 4 mg/kg xylazine) were administered hourly or as needed throughout the recording session. Following anesthetization, subjects were mounted in a stereotaxic device in a soundproofed and electrically shielded booth (IAC Acoustics, Bronx, NY). Body temperature was maintained at 37.5°C with a thermistor-controlled heating pad (Harvard Apparatus, Holliston, MA) on the subject's abdomen.

To confirm normal hearing sensitivity, auditory brainstem responses (ABR) were elicited by a click stimulus at 50 and 80 dB SPL in each subject. To measure ABRs, electromyographic needle electrodes were inserted into the skin midway between the 2 ears (noninverting), on the snout midway between the eyes and nose (inverting), and into loose skin at the neck (ground).

### Surgery

A rostro-caudal incision was made along the scalp surface and tissue was retracted to expose the skull. Following exposure, holes were drilled into the skull under an operating microscope. A section of the dura was removed with a cautery, and mineral oil was used to coat the cortical surface. ICc was accessed using a vertical approach with tungsten microelectrodes (MicroProbes, Gaithersburg, MD). The electrode impedance was approximately 2 M $\Omega$  at 1 kHz, which corresponds to a recording volume conservatively calculated at <0.1 mm<sup>3</sup>, and therefore reflects multiunit activity. An electrode was advanced perpendicular to the cortical surface using a remote-controlled micromanipulator (Märzhäuser Wetzlar GmbH & Co. KG, Wetzlar, Germany), and for all recordings the dorsal/ventral reference of the electrode was determined at a point slightly above cortex surface; this position was maintained for all penetrations within a subject. ICc coordinates were approximately 0.3 mm caudal to the interaural line, 1.5 mm left of the sagittal suture, and 4.0 mm ventral to the surface of the brain. Simultaneous surface recordings were measured with a supradural silver ball electrode placed at the vertex 10 mm caudal to Bregma. The ground electrode (alligator clip) was positioned in loose skin towards the posterior of the scalp and the reference electrode (silver ball) was placed 15 mm rostral of Bregma.

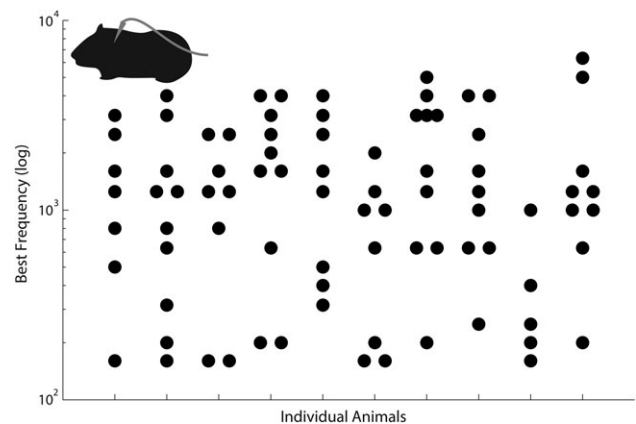
### Neurophysiological Recording

A depth penetration technique was used during electrode advancement. Click stimuli (100  $\mu$ s rectangular pulses) were delivered at 3.5 Hz. Multiple penetrations were made in each subject. A monitoring oscilloscope was inspected and the response size and gross waveform morphology were considered. If the response was small and broad, electrode penetration was continued until the waveform was characteristic of an ICc response, namely, large amplitude with a sharp onset. Location was verified by comparing characteristics of responses

to probe tones and noise to published response properties of ICc neurons (Rees and Palmer 1988; Syka et al. 2000; Liu et al. 2006).

The best frequency (BF) region of each site was determined using a procedure similar to Xie et al. (2007). Specifically, we determined what frequency from 160 to 16 000 Hz (varying in third octaves) elicited the most robust response at each site. Each probe tone was presented at a low intensity (30 dB HL) and was 100 ms in duration with a 10 ms rise-fall time and a 110 ms interstimulus interval (ISI). A total of 30 repetitions of a tone at a given frequency were presented, and the entire group of tones was presented varying in frequency in a pseudorandom order for each recording site. Tuned regions for each subject are summarized in Figure 2. BFs ranged from 160 to 6300 Hz (median: 1250 Hz). It should be noted that this range of BFs only represents a subset of the guinea pig tonotopic axis (Malmierca et al. 1995); this range was selected because it corresponds to the frequencies present in speech. The goal in selecting recording sites was to get a range of BFs across the frequency spectrum of speech; in some cases multiple recordings were made from 2 or more sites with the same BF.

For each recording site in ICc, 300 presentations of the [da] (150 per polarity) were presented using a computer with custom MATLAB programs (Mathworks, Natick, MA). Stimuli were converted to analog signals using a National Instruments D/A converter (National Instruments Corporation, Austin, TX) and delivered via electromagnetically shielded earphones (ER-2; Etymotic Research, Inc., Elk Grove Village, IL) monaurally through hollow ear bars. In quiet the [da] was presented at 75 dBA, and the noise condition the [da] was presented at 75 dBA with the noise at 65 dBA. Speech tokens were presented with a 60 ms ISI. Responses were differentially amplified with a gain of 500 and filtered from 10 to 8000 Hz by 2 Grass P511 amplifiers (Grass Technologies, West Warwick, RI), digitized at 33.333 kHz by an MCC A/D board (Measurement Computing Corporation, Norton, MA), and then saved to second computer running a custom MATLAB acquisition program. The MCC board received triggers from the D/A converter to mark stimulus onsets. Responses were epoched from –40 to 190 ms (re-



**Figure 2.** A schematic illustrating the recordings collected from the animal component of the study. Extracellular EEG was recorded in vivo from several sites in the central nucleus of inferior colliculus. Each animal subject is represented by a column, and each dot represents a single recording from that subject, aligned with the best frequency response of that site. In some cases, multiple penetrations were made in an individual subject that resulted in multiple recordings from sites with the same best frequency.

stimulus onset) and high-pass filtered prior to analysis (100 Hz cut-off, 2nd order Butterworth).

In total there were 84 ICc recordings. Surface recordings were combined from all penetrations within a single subject. Responses to the 2 polarities were added at the final averaging stage.

## Human Component

### Subjects

The subjects in the human component were 50 children, ages 3.4–5.2 years (26 female; mean age: 4.5 years, SD = 0.3). All of the children were native English speakers with no second language experience; no parent reported a history of a neurologic condition or a diagnosis of autism spectrum disorder. Each child passed a screening of peripheral auditory function (normal otoscopy, type A tympanograms, and distortion product otoacoustic emissions > 6 dB above the noise floor from 0.5 to 4 kHz). Additionally, all children had normal click-evoked ABRs (wave V latency <5.84 ms in response to a 100  $\mu$ s square wave click presented at 80.4 dB SPL to the right ear). Children received \$10/h and a cute t-shirt for their participation.

### Neurophysiological Recording

Children were seated comfortably in an electrically shielded and sound-attenuated booth (IAC Acoustics) while watching a film to facilitate a relaxed state. The film was presented at 45 dB SPL so that the children could listen to it through the left ear, a technique often used to measure auditory-neurophysiological responses in children (Cunningham et al. 2000, 2001; White-Schwoch et al. 2015a; 2015b). The stimuli were delivered monaurally to the right ear at via electromagnetically shielded insert earphones (ER-3A; Etymotic Research) with an 81 ms ISI. In quiet the [da] was presented at 80 dB SPL, and the noise condition the [da] was presented at 80 dB SPL with the noise at 70 dB SPL. Responses were recorded with a BioSemi Active2 system (BioSemi, Amsterdam, The Netherlands) with an ActiABR module into LabView 2.0 (National Instruments Corporation, Austin, TX). A vertical recording montage was used with CMS/DRL centered around Fpz, the active electrode at Cz, and the reference on the right earlobe. Responses were digitized at 16.384 kHz with an online high-pass filter at 100 Hz (first order Butterworth) and low-pass filtered at 3200 Hz (5th order sinc filter).

The BioSemi ActiABR records with a highpass filter at 100 Hz. Our laboratory's other data in humans and in the animal model are recorded with open filters. To compare across these recording systems, the BioSemi responses were amplified in the frequency domain for 3 decades below 100 Hz, simulating a recording with open filters. Responses were then bandpass filtered using our standard FFR range, 70–2000 Hz (second order Butterworth, zero phase shift). The responses were segmented into epochs corresponding to the time window over which a stimulus is presented. With stimulus onset as 0 ms, the time range for epochs was –40 to 210 ms. Epochs were then baseline-corrected to the prestimulus period. Epochs with any point amplitude exceeding  $\pm 35 \mu$ V were rejected as artifact (~10–15% of recordings; all final recordings comprised 4000 trials).

### Behavioral Tests

The Children's Evaluation of Language Fundamentals, 2nd Preschool Edition (Pearson, San Antonio, TX) was used to evaluate early language development. In particular, the Phonological Awareness test evaluated children's knowledge of and ability to fluently manipulate the sound structure of spoken language.

The Matrix Reasoning subtest of the Wechsler Preschool and Primary Scale of Intelligence-III (Pearson) was also administered. This provides a measure of nonverbal intelligence that is used as a control behavioral measure (see Rosen 2003 for a discussion of this issue).

## Data Analysis

### Animal Component—ICc recordings

**Variability:** Of interest was the morphological similarity of the neurophysiological responses across stimulus trials; that is, how variably is speech represented by midbrain across trials? Within each ICc recording, the mean of the correlation between all possible pairs of trials was computed. Correlations were computed separately over response regions corresponding to stimulus onset (10 ms window), the CV transition (50 ms window), and the vowel (110 ms window). This is referred to as the intertrial variability measure. Please note that because a correlation value is used for variability, a higher number indicates a less variable (i.e., more consistent) response.

The correlation for responses to the onset, CV transition, and vowel were computed over windows that differed in size. To ensure that the size of the analysis window did not bias results, a complementary sliding window analyses was conducted on the responses. Correlations were run for all possible pairs of trials across the response time region, with 20-ms windows and 1 ms of overlap. Specifically, the first window was centered at 0 ms (i.e., –10 to 10 ms), the second was centered at 1 ms (i.e., –9 to 11 ms) and the last was centered at 190 ms (i.e., 180–200 ms). As above, a higher number indicates a less variable (i.e., more consistent) response.

**Jitter:** In addition, capitalizing on the high SNR offered by ICc recordings, the mean timing difference, or jitter, between all possible pairs of trials was computed. This jitter was calculated using a cross-correlation; that is, 2 trials were shifted in time relative to each other (range of lags  $\pm 7$  ms) to determine the lag at which the correlation between the 2 trials reaches its maximum. The absolute value of the lag was used for statistical purposes. Once again this procedure was applied separately over response regions corresponding to stimulus onset, the CV transition, and the vowel. This is referred to as the intertrial jitter measure. Please note that because a timing lag is used for jitter, a higher value indicates means a more jittered response.

The jitter for responses to the onset, CV transition, and vowel were run over windows that differed in size. To ensure that the size of the analysis window did not bias results, a complementary sliding window analyses was conducted on the responses. Cross-correlations were run for all possible pairs of trials across the response time region, with 20-ms windows and 1 ms of overlap. Specifically, the first window was centered at 0 ms (i.e., –10 to 10 ms), the second was centered at 1 ms (i.e., –9 to 11 ms) and the last was centered at 190 ms (i.e., 180–200 ms). For each window, the absolute value of the lag that achieved the maximum correlation was determined, and the mean of these values across windows and all possible pairs of trials was determined as the intertrial jitter. As above, a higher number indicates a more jittered response.

It should be noted that our method for computing jitter differs somewhat from approaches in single-unit studies, which typically involves taking the standard deviation or coefficient of variation of spike timing (e.g., Mainen and Sejnowski 1995). We chose our analysis approach because we wanted to hew as closely as possible to the techniques used in human neurophysiological studies, which require subaveraging because of

the low SNR of scalp recordings, while still capitalizing on the high SNR of the IC recordings. Moreover, because we are comparing intracollicular responses to broadband population responses in humans, we did not want to bias calculations on our near-field responses to the spiking activity.

**Phase-locking factor:** For the ICc recordings, we supplemented our variability and jitter analyses by calculating the intertrial phase-locking factor (PLF) in response to the [da] in quiet and noise. The advantage of the PLF approach is that it provides time- and frequency-specific information about the variability of extracellular activity. An additional advantage is that this analysis is agnostic to the amplitude of the response. Specifically, the PLF quantifies the consistency of the phase of the response in discrete time–frequency bins. Thus, this analysis allowed us to infer if the variability and jitter effects were driven by responses to specific frequency bins (such as the response to the fundamental frequency or first formant).

PLF was calculated for each ICc response in quiet and noise. A sliding window analysis was used with 40 ms bins run across the response epoch (midpoints were 0–170 ms re stimulus onset). The phase of responses to each polarity was maintained (this is analogous to adding response to alternating polarities, as indicated above). The frequency spectrum of each bin was calculated with a fast Fourier transform (Hanning window), and normalized to unit vectors at each frequency point, thus avoiding any bias of response amplitude. The unit vectors were then averaged, and the length of the resulting vector is the PLF. Mean PLFs were calculated in 20 Hz bins centered on the fundamental and integer harmonics (100–2000 Hz) for responses to the CV transition (20–70 ms) and the vowel (70–170 ms). Because the response to the onset is a transient we do not anticipate phase-locking statistical analyses of PLFs were therefore limited to the CV transition and vowel.

#### Animal Component—Surface Recordings

Responses recorded at the surface did not have sufficient SNR to analyze single sweeps. Thus, we used a subaveraging technique our group has used in human scalp recordings (White-Schwoch et al. 2015a; 2015b) and that was used in the human component of this study. The technique involves randomly computing, and then correlating, 2 subaverages. Each subaverage comprised 750 stimulus repetitions because the animal with the fewest recordings had a total of 1500 stimulus repetitions recorded at the surface. This procedure is performed 300 times, each time randomly selecting 2 different subaverages and calculating the mean of each of those correlations. Once again this procedure was applied separately over response regions corresponding to stimulus onset, the CV transition, and the vowel. This is referred to as the surface intertrial variability measure.

#### Human Component—Surface Recordings

A procedure identical to that used for animal surface recordings was used to compute the intertrial variability in human scalp recordings. In this component of the study, subaverages comprised 2000 trials.

#### Statistical Approach

Prior to statistical analyses we applied a number of transformations to the raw data values so that we could analyze them with the general linear model. Because the distribution of best frequencies was non-normal, we transformed it into a log scale. Intertrial variability was calculated using Pearson correlation

coefficients ( $r$ 's). The  $r$  distribution is not normal (it has a restricted range from 0 to 1) and so is not appropriate for statistical analysis under the general linear model. We therefore transformed these to Fisher  $z$  correlation values (Cohen et al. 2003); this increases the “spread” of the data, especially as  $r$  nears 1, and normalizes it. As with  $r$ , larger values of  $z$  indicate stronger correlations. Finally, we took the absolute value of our intertrial jitter calculations. The jitter between 2 trials could arbitrarily be positive or negative (a given trial could be earlier or later than the reference trial); we were interested in the magnitude of timing variability, not its direction.

The first set of analyses examined the extent to which intertrial variability and jitter vary across response time region (onset, CV transition, vowel), listening condition (quiet vs. noise), and their interactions, across animal ICc, animal surface, and human scalp recordings. We used mixed-effects modeling to accomplish this. Mixed-effects models are a variant of the general linear model that allow the simultaneous evaluation of fixed factors (e.g., stimulus time region and listening condition) and random factors (e.g., individual subjects). The advantage of mixed-effects models is that they allow us to analyze effects across a different number of levels for certain parameters; different animals had different numbers of ICc recordings (ranging from 5 to 10 recordings per condition). Additionally, we had up to 10 ICc recordings from each animal but only one surface recording per condition. The mixed-effects models allowed us to include subjects as a factor so that we could model recording site on a within-subject basis, and are reported similarly to conventional ANOVAs, with  $\eta^2$  as effect size. By convention,  $\eta^2 \geq 0.01$  is considered a “small” effect,  $\eta^2 \geq 0.06$  is considered a “medium” effect, and  $\eta^2 \geq 0.26$  is considered a “large” effect (Cohen 1988). PLFs for the ICc recordings were compared with a repeated-measures analysis of variance across the twenty frequency bins, covarying for the BF of each site.

We also conducted 2 multivariate, hierarchical regressions. Multivariate regression examines whether a set of independent variables, in combination, predict a single dependent variable. Hierarchical regression models the contributions of these variables in steps, and determines if specific measures explain additional variance after controlling for the contribution of one set of measures. The first regression tested the relationship between intertrial variability in the IC and jitter after controlling for tuning of the recording site and stimulus parameters. The second regression tested the relationship between intertrial variability at the scalp and language skills in children after controlling for demographic factors and intelligence.

## Results

### Animal Component

In total, 84 ICc recordings in response to [da] in quiet and in noise were collected across 10 animals. Tonotopic sites for each subject are summarized in Figure 2, with each dot representing one recording. The BF of that recording site is indicated on the ordinate; in some cases, several penetrations were made in an individual subject, resulting in multiple recordings from sites with an identical BF.

As expected, far-field recordings (animal surface and human surface) did not have sufficient SNR to analyze single trials. In contrast, near-field (animal ICc) recordings offered sufficient SNR to analyze single trials. The analysis approach therefore capitalized on this favorable SNR to evaluate trial-to-trial variability and jitter in neural coding. Several trials from an

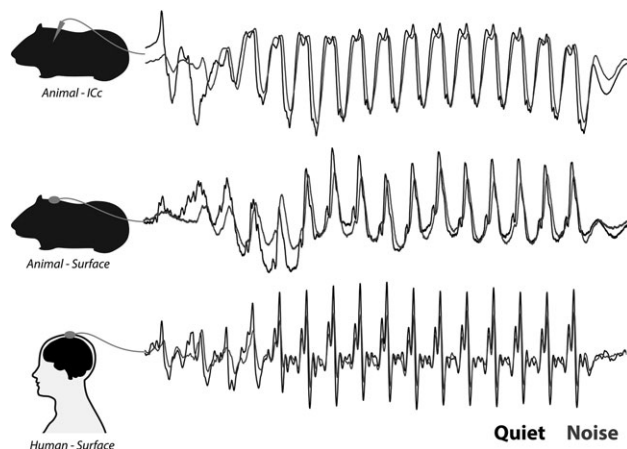


individual ICc recording are illustrated on Figure 1A, which illustrates that each trial in the near-field recordings is interpretable (contrast this to Fig. 1B, which shows an equivalent number of trials collected at the human scalp).

Figure 3 shows the grand average responses to [da] in quiet and noise, averaged across all 84 ICc recordings (top), all 10 animal surface recordings (middle), and all 50 human surface recordings (bottom). Broadly, responses from ICc, surface, and scalp were morphologically similar. All show a strong periodic component that corresponds to phase-locked activity in response to the fundamental frequency of the stimulus (100 Hz). When viewed as grand averages across all subjects, responses in humans show the most fine structure (the high-frequency components between each major peak in the response). The animal surface recordings resemble the human scalp recordings more strongly than the ICc recordings; this is expected because the ICc recording is a near-field response whereas the surfaces responses are far-field and more analogous to scalp recordings. Across all recording sites, averaged responses in noise tended to be smaller and later than responses in quiet.

### ICc Responses to Dynamic Speech Features in Quiet and Noise are Variable and Jittered

Responses from a representative ICc recording are illustrated in Figures 4 (speech in quiet) and 5 (speech in noise), broken down by time regions of the response; twenty randomly selected responses are overlaid with heat maps illustrating intertrial variability and jitter. As may be seen in Figure 4, responses are essentially similar across trials in the quiet condition. In Figure 4A, a negative deflection is apparent in each response at about 9 ms, which corresponds to the onset of the stimulus; heat maps are shown in Figure 4B,C, showing the low variability and jitter, respectively, across trials in response to the onset. In Figure 4D, the response to the CV transition is illustrated, and in each trial phase-locking to the fundamental frequency of the stimulus becomes apparent; heat maps show that the variability and jitter (Figs 4E,F, respectively) are lower for this response region than the preceding. Finally, this figure (Fig. 4G) illustrates the response to the vowel where a striking similarity



**Figure 3.** Grand average responses to the speech sound [da] are shown in quiet (black) and noise (gray) for (A) all ICc recordings ( $N = 84$ ); (B) all surface recordings from the animal model ( $N = 10$ ); and (C) all human scalp recordings ( $N = 50$ ).

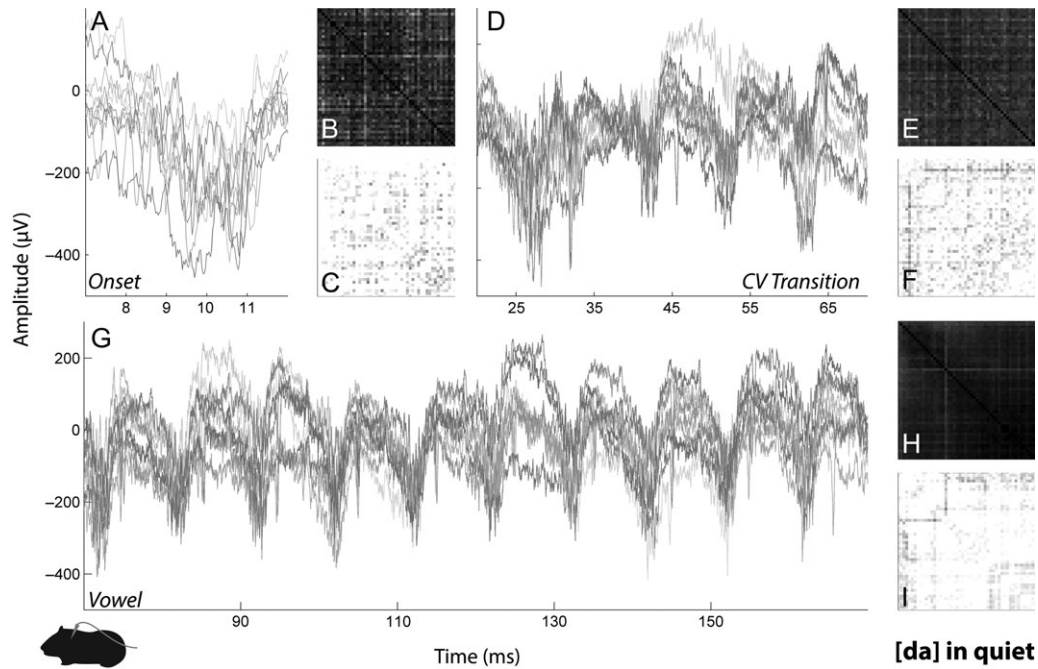
is observed across response trials, and variability and jitter reach their nadirs (Figs. 4H,I).

Contrast this to Figure 5, which is organized analogously and presents the same animal's response to speech in noise, recorded from the same site in ICc. In Figure 5A, it is extremely difficult to discern consistent onset response timing across trials; the onset response is essentially gone, and the heat maps show that the response is completely variable and jittered (Figs 5B,C). In Figure 5D, phase-locking begins to emerge but the responses are substantially more variable and jittered than the corresponding response in quiet (Fig. 4D). The heat maps also show that the response is somewhat less variable and jittered than the onset response in noise (Figs 5E,F). Finally, the response to the vowel in noise is shown in Figure 5G—phase-locked activity is apparent, however, these responses are still more variable and jittered (Figs 5H,I) than in quiet.

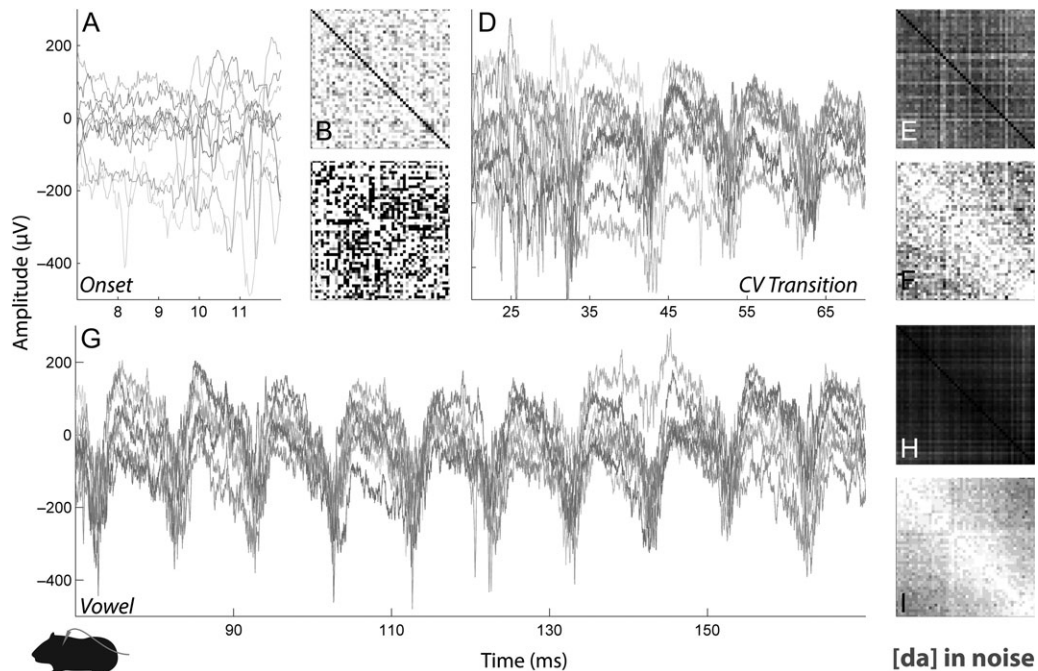
Similar observations were made for all animal ICc responses across tonotopic sites. In the sections that follow, these observations are tested statistically across the subject population. These results are illustrated in Figure 6A, with quiet responses in black and noise responses in red (recall these are Fisher's  $z$ -transformed correlation coefficients, so higher numbers are less variable across trials).

Responses to the vowel were less variable than responses to the consonant transition, both of which were less variable than responses to the onset burst (main effect of time region:  $F(2,21.815) = 438.212$ ,  $P < 0.001$ ,  $\eta^2 = 0.976$ ). When background noise was added, responses were more variable than they were in quiet (main effect of condition:  $F(1,9.251) = 137.242$ ,  $P < 0.001$ ,  $\eta^2 = 0.937$ ). This noise degradation was exacerbated in response to dynamic speech features, meaning that the responses to the onset and consonant transition were relatively more variable, compared with the vowel response, in noise than they were in quiet (condition  $\times$  time region interaction:  $F(2,19.214) = 56.560$ ,  $P < 0.001$ ,  $\eta^2 = 0.855$ ). Onset response variability in noise was particularly high.

Variability calculations were run over response time regions that differed in size: the onset analysis window was 10 ms long, the CV transition analysis window was 50 ms long, and the vowel analysis window was 110 ms long. To ensure that these results were not biased by differences in the size of the analysis window, a sliding window correlation analysis was performed on the responses. All possible pairs of trials were correlated in the sliding window analysis, which calculates the correlation between a pair of trials for 20 ms bins running across the response length. Results are illustrated as Fisher's  $z$  correlation values in Figure 7A. The pattern of results is consistent with those shown in Figure 6A, where responses in quiet are more consistent than responses in noise, responses to the vowel are more consistent than responses to the CV transition and onset, and noise dramatically increases variability in response to transient speech features (higher values on the ordinate reflect more consistent responses). Figure 7C illustrates the effect sizes (Cohen's  $d$ ) comparing variability in quiet versus noise across the response (each effect size reflects the magnitude of the difference in variability between corresponding 20-ms bins in quiet and noise). Consistent with the effects discussed above and illustrated in Figure 6A, noise has a larger effect on response variability in response to transient, rather than static, speech features. Remarkably, these patterns are observed in individual responses, which are illustrated in Figure 7B (quiet) and Figure 7D (noise)—in both cases, despite disparities across recordings in terms of the degree of response variability, the vast majority of responses show the patterns illustrated in



**Figure 4.** ICc responses to speech in quiet are illustrated from a representative animal subject; this is the same subject whose response in noise is illustrated in Figure 5. Shown are several single trials overlaid, intertrial variability, and intertrial jitter in response to the onset (A, B, C), the consonant–vowel transition (D, E, F), and the vowel (G, H, I). For each time region, several single sweeps are overlaid—each is shown as a gray trace (A, D, G). Next, a heat map illustrates the intertrial correlation between pairs of trials; each row and column represents a single trial and the heat map illustrates the correlation between those 2 trials, with white indicating a low correlation and black indicating a high correlation (B, E, and H). Finally, heat maps show the intertrial jitter between corresponding pairs of trials, with white representing zero jitter and black showing a greater timing difference between each pair (absolute value; C, F, and I).

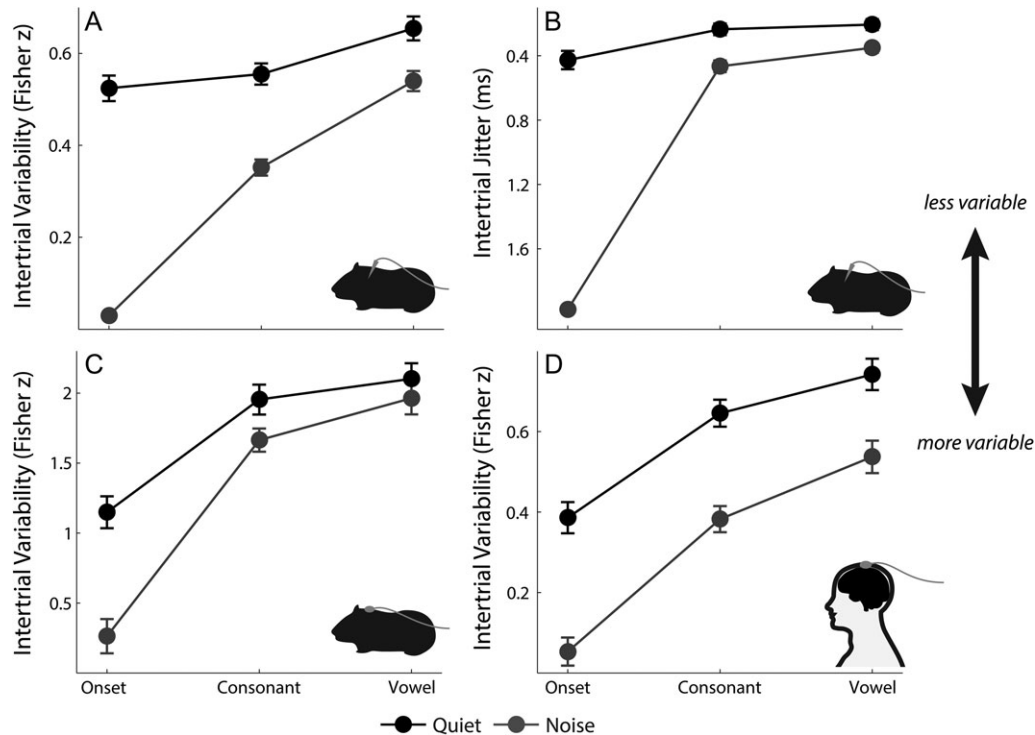


**Figure 5.** ICc responses to speech in noise are illustrated from a representative animal subject; this is the same subject whose response in quiet is illustrated in Figure 4. This figure is organized identically to Figure 4.

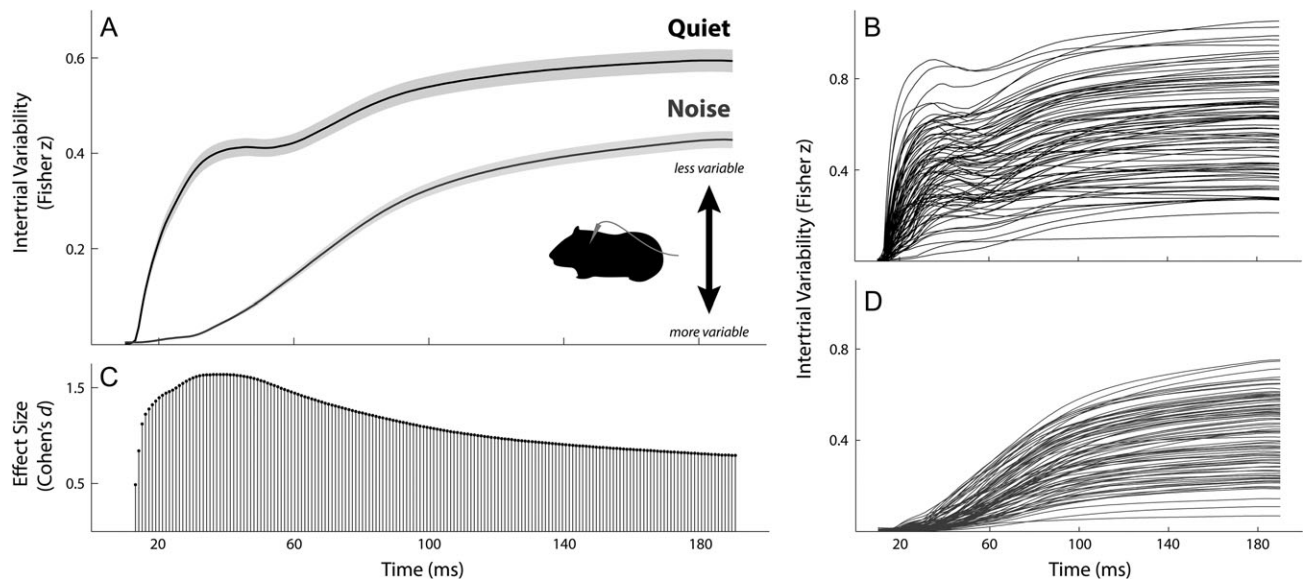
Figures 6A and 7A. Specifically, responses that tend to be the most consistent show greatest response stability for the vowel portion of the response, while the more variable responses show reduced response stability for the onset and CV transition features.

The intertrial jitter was computed as the mean of the absolute value of the lags between all pairs of trials, and followed an identical pattern of results as response variability. These results are illustrated in Figure 6B, with larger intertrial timing variability indicated by a larger degree of jitter (note that larger





**Figure 6.** Intertrial correlation and jitter are shown across time regions of the response (onset, consonant–vowel transition, and vowel) and conditions (quiet and noise) for the animal (A–C) and human (D) components of the study. (A) The intertrial correlation is shown for ICc recordings. (B) The intertrial jitter is shown from surface recordings. (C) The intertrial correlation is shown from surface recordings. (D) In humans, a similar pattern of intertrial correlations is observed in scalp recordings. Namely, responses in quiet (black) are more correlated across trials than responses in noise (gray), and responses to the vowel are more correlated than responses to the consonant–vowel transition, which are in turn more correlated than responses to the onset. The addition of background noise exacerbates the variability in response to the onset and consonant–vowel transition. This effect is statistically identical to that observed in the animal model (panels A, C). Illustrated are means with error bars  $\pm 1$  standard error of the mean.



**Figure 7.** To complement the correlation analyses over discrete time regions of the responses, a sliding-window analysis was conducted. Correlations were calculated between all possible pairs of trials for 20-ms windows running across the response. (A) Mean intertrial correlations are illustrated for responses in quiet (black) and noise (gray), with the shaded area showing  $\pm 1$  standard error of the mean. Higher values on the ordinate reflect stronger correlations, hence less variable responses. Each point on the abscissa is the midpoint of the window the correlation was conducted on (e.g., the point at time 20 ms is the correlation on a window from 10 to 30 ms). The pattern of intertrial correlations clearly reproduces the pattern illustrated in Figure 6A, with responses in quiet less variable than responses in noise, responses to the vowel less variable than responses to the onset or CV transition, and noise more adversely affecting response variability in response to the onset and CV transition than in response to quiet. This pattern of results was evident in all 84 recordings from each site across animals in quiet (B) and noise (D). Additionally, this pattern of results is seen when comparing effect sizes (Cohen's  $d$ ) between quiet and noise (C); each point shows the effect sizes of the difference in correlations between quiet and noise.

values are lower on this ordinate). Namely, the response to the onset was more jittered than the response to the consonant transition, which was in turn more jittered than the response to the vowel (main effect of time region:  $F(2,19.467) = 321.270$ ,  $P < 0.001$ ,  $\eta^2 = 0.971$ ). When background noise was added the responses also became more jittered (main effect of condition:  $F(1,9.148) = 166.935$ ,  $P < 0.001$ ,  $\eta^2 = 0.948$ ), but once again this effect was exacerbated for responses to the onset and consonant transition (condition  $\times$  time region interaction:  $F(2,18.849) = 49.659$ ,  $P < 0.001$ ,  $\eta^2 = 0.840$ ), with the most jitter seen in response to the onset in noise.

Like the variability calculations, jitter was computed over response time regions that differed in size (see above). To ensure these results were not biased by differences in the size of the analysis window, a sliding-window analysis was conducted. Specifically, the timing lag (absolute value) between all possible pairs of trials was calculated over a sliding window, which computes the timing difference between a pair of trials for 20 ms bins running across the response length. Results are illustrated in Figure 8A. The pattern of results is consistent with those shown in Figure 6B, where responses in quiet are less jittered than responses in noise, responses to the vowel are less jittered than responses to the CV transition and onset, and noise dramatically increases jitter in response to transient speech features (lower values on the ordinate reflect less jittered responses). Figure 8C illustrates the effect sizes (Cohen's  $d$ ) comparing jitter in quiet versus noise across the response (each effect size reflects the magnitude of the difference in jitter between corresponding 20-ms bins in quiet and noise). Consistent with the effects discussed above and illustrated in Figure 6B, noise has a larger effect on jitter in response to transient, rather than static speech features. These patterns are observed in individual responses, illustrated in Figure 8B (quiet) and Figure 8D (noise). In both cases, despite disparities across recordings in terms of the degree of response jitter, the vast majority show the pattern illustrated in Figures 6B and 8A.

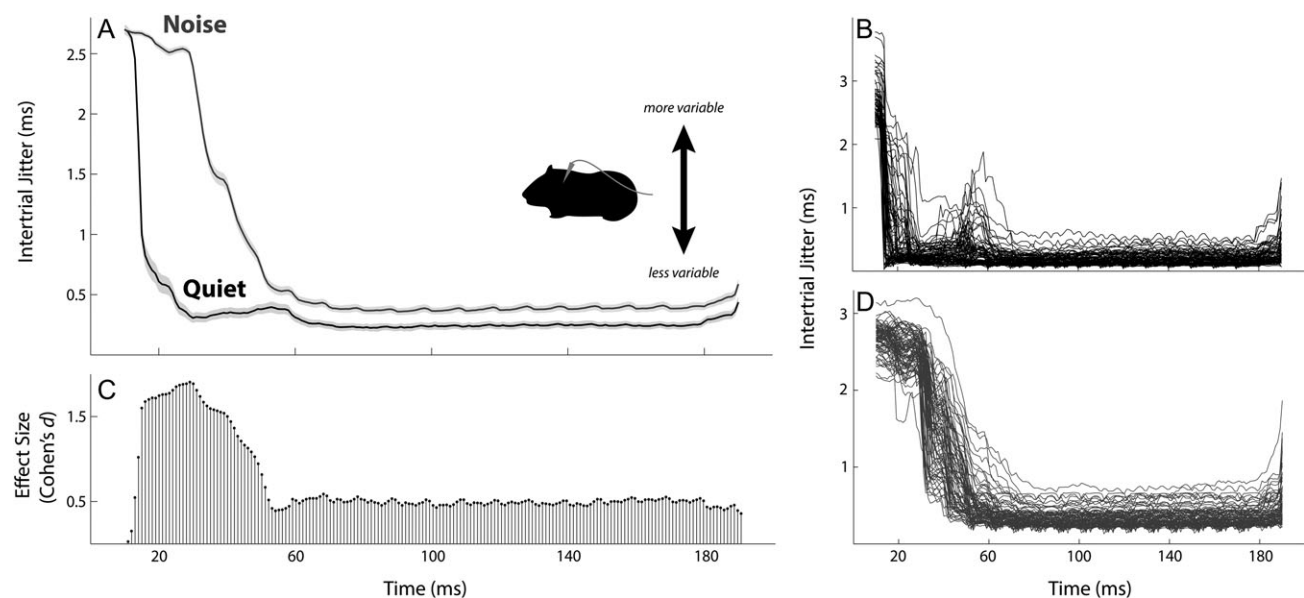
Most of the intertrial variability calculations for ICc recordings in the onset time region were extremely low—50% fell

below 0.1 (Fisher's  $z$ -transformed correlation coefficient; compared with 1.8% of CV transition responses and 0.01% of vowel responses). To ensure this skew did not drive the effects, the intertrial variability and jitter were reanalyzed, comparing only responses to the CV transition and vowel. Identical patterns were observed. The response to the transition was more variable than the response to the vowel (main effect of time region:  $F(1,11.183) = 194.319$ ,  $P < 0.001$ ,  $\eta^2 = 0.929$ ), the response in noise was more variable than the response in quiet (main effect of condition:  $F(1,10.164) = 133.826$ ,  $P < 0.001$ ,  $\eta^2 = 0.929$ ), and noise exacerbated the variability in response to the CV transition (condition  $\times$  time region interaction:  $F(1,14.124) = 33.071$ ,  $P < 0.001$ ,  $\eta^2 = 0.701$ ). In addition, the response to the transition was more jittered than the response to the vowel (main effect of time region:  $F(1,13.055) = 64.694$ ,  $P < 0.001$ ,  $\eta^2 = 0.832$ ), the response in noise was more jittered than the response in quiet (main effect of condition:  $F(1,9.239) = 86.865$ ,  $P < 0.001$ ,  $\eta^2 = 0.904$ ), and the jitter was exacerbated for the response to the transition in noise (condition  $\times$  time region interaction:  $F(1,14.280) = 11.744$ ,  $P = 0.004$ ,  $\eta^2 = 0.451$ ).

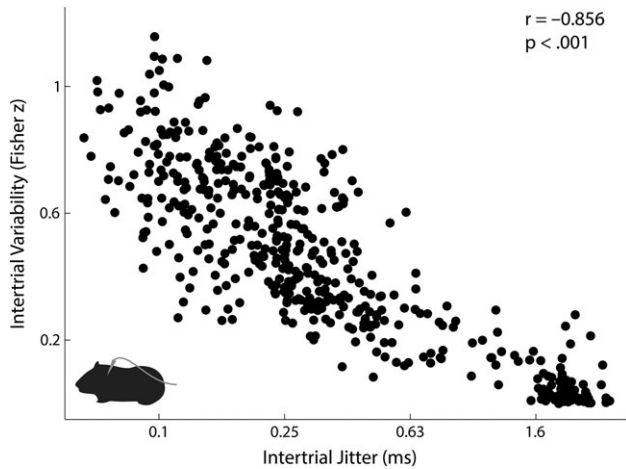
### Intertrial Variability Reflects Timing Jitter

ICc responses followed similar patterns in terms of how stimulus features affect representational variability and timing jitter. We next tested the hypothesis that variable responses in time (jittered) are also variable in morphology. Indeed, the 2 measures were strongly correlated across all recording sites, time regions, and conditions ( $r(502) = -0.856$ ,  $P < 0.001$ ). This is illustrated on Figure 9.

To ensure that this relation held irrespective of the tonotopy of the recording site, time region of the response, or condition, a multistep hierarchical regression was conducted to predict intertrial variability from the intertrial jitter. On the first step, the BF response of the site accounted for 16.1% variance in intertrial variability ( $F(1,502) = 96.004$ ,  $P < 0.001$ ). The addition of condition and time range on the second step accounted for an additional 44% of variance in intertrial variability



**Figure 8.** To complement the jitter analyses over discrete time regions of the responses, a sliding-window analysis was conducted. This figure is organized analogously to Figure 7, illustrating the intertrial jitter patterns show in Figure 6B. Note that the absolute value of the jitter was calculated, and here higher values on the ordinate reflect more jittered responses.



**Figure 9.** In ICc, intertrial correlations relate strongly to intertrial timing jitter: the more jittered a response across trials, the less correlated it is across trials. This holds across all recording sites, time regions of the response (onset, consonant–vowel transition, vowel) and conditions (quiet and noise). Each dot represents one “pair” of variability and jitter calculations, and this figure illustrates every calculation for the 10 subjects, 3 time regions of the response, and 2 conditions. Note that the abscissa is on a logarithmic scale to make it easier to see the relation to variability, but that the raw values (ms) are indicated at each tick mark.

**Table 1** In ICc, intertrial variability reflects intertrial timing jitter across tonotopic sites, response time regions, and conditions

Predictor	$\Delta R^2$	$\beta$
Step 1	0.161	
Best frequency		0.401**
Step 2	0.440	
Best frequency		0.401**
Time region		−0.477**
Condition		0.461**
Step 3	0.195	
Best frequency		0.254**
Time region		0.119**
Condition		−0.060*
Intertrial jitter		−0.712**
Total $R^2$	0.796	

Multiple regression results of factors predicting intertrial variability.

\* $P = 0.030$ , \*\* $P < 0.001$ .

( $F(2,500) = 275.609$ ,  $P < 0.001$ ). On the third step intertrial jitter was added, and this factor uniquely accounted for an additional 19.5% variance in intertrial variability ( $F(1,499) = 478.034$ ,  $P < 0.001$ ;  $\beta_{\text{jitter}} = -0.712$ ,  $t = -21.864$ ,  $P < 0.001$ ). Thus, irrespective of recording site, time region, or noise condition there is a consistent and robust relationship between timing jitter and representational variability (total  $R^2 = 0.796$ ,  $F(4,499) = 486.998$ ,  $P < 0.001$ ; see Table 1 for the full regression results).

In the multiple regression, the unstandardized beta coefficient ( $\beta$ ) of the intertrial jitter was  $-0.455$  (log scale; standard error, 0.021). The beta refers to the coefficient of this term in the regression equation, and shows the influence of a change in one unit of the independent variable (jitter) on the dependent variable (variability). For perspective, this means that across laminae, stimulus features, and condition, every 1 ms of timing jitter increases the response variability (i.e., decreases the correlation value) by 0.337 (Pearson’s  $r$ ).

## Variability of Extracellular Activity Across Frequencies Shapes Intertrial Variability and Jitter

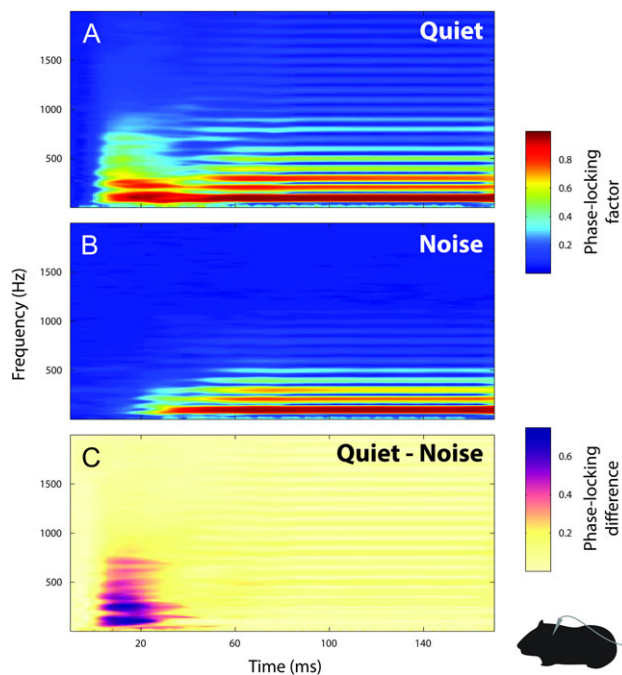
The preceding analyses of variability and jitter were conducted on responses in the time domain, meaning they considered extracellular activity across frequencies. This raises the question of whether certain frequencies in the response shape these patterns of variability. From a neurophysiological standpoint, if this variability was biased to a certain frequency band it could suggest that these patterns are driven by specific processes. For example, if variability patterns were restricted to neural activity occurring below 300 Hz it might suggest they are driven by the local field potentials, or the input to ICc. From an acoustic-phonetic standpoint, if variability patterns were restricted to specific frequencies it could indicate which stimulus features drive these effects. For example, if these variability patterns were restricted to activity around 700 Hz it might suggest they were driven by neural coding of the first formant (Fig. 1).

To tease apart these possibilities we did a phase-locking analysis. This analysis quantifies the consistency of the phase of the extracellular activity at specific time–frequency bins in the response. Thus, we can determine how consistently specific frequency bandwidths in the speech stimulus are coded for both the CV transition and the vowel. (Because the response to the onset burst is a transient we do not expect phase-locking to emerge). The phase consistency of the response is quantified with a “phase-locking factor” (PLF) that ranges from 0 (completely variable) to 1 (completely consistent). Previous work in humans show individuals with poor listening-in-noise skills, such as older adults, have diminished phase-locking in response to speech (Anderson et al. 2012; Ruggles et al. 2012). Additionally, work in humans shows this analysis often aligns with response variability in the time domain (Tierney and Kraus 2013; Woodruff Carr et al. 2015).

Grand-averaged response phase-locking is illustrated in Figure 10. These time–frequency plots use color to indicate the strength of phase-locking (ranging from blue for low PLFs to red for high PLFs). Figure 10A shows the response in quiet; the PLF plot resembles the spectrogram of the stimulus that is illustrated in Figure 1. The strongest phase-locking is observed at the fundamental frequency, and diminishes up to higher frequencies. Nevertheless, reliable phase-locking is observed out to very high frequencies—perhaps as high as 2000 Hz. This is consistent with our previous observations in speech-evoked ICc activity, which showed that some sites exhibit these high-frequency phase-locked responses (Warrier et al. 2011). Figure 10B shows the grand-average response in noise. Overall phase-locking is smaller, and absent at the highest frequencies. Additionally, phase-locking is substantially diminished in response to the CV transition, consistent with the aforementioned finding that response variability and jitter are higher for the CV transition in noise relative to quiet. This is further illustrated in Figure 10C, which is a difference plot of the top 2 panels (purple indicates time–frequency points where PLFs were higher in quiet than in noise). The strongest difference is observed in response to the CV transition, and smaller differences are seen in high frequencies in the vowel.

Figure 11 shows mean PLFs for the response to the CV Transition (Panel A) and vowel (Panel B) in quiet (black) and background noise (gray). Although Figure 10 shows a burst of energy corresponding to stimulus onset, we do not expect phase-locking in response to a transient burst and so restricted our statistical analyses to the transition and vowel regions. The





**Figure 10.** Phase-locking factor (PLFs) plots for responses to speech in quiet (A) and noise (B), averaged across all ICc recordings. The PLF is calculated for each time–frequency bin in the response and is illustrated on a colorscale, ranging from blue (0, complete phase variability) to red (1, complete phase consistency). The strongest phaselocking is observed in low frequencies, in response to the vowel, and in quiet. Background noise diminishes phase-locking, especially in response to the CV transition. Although PLFs decrease with ascending frequency, in quiet they are observed up to 2000 Hz. These observations are confirmed by subtracting PLFs in quiet and noise at corresponding time–frequency bins (C). In the difference plots, deep purple indicates points where PLFs were higher in quiet than in noise.

mean PLFs are plotted at each frequency, and the insets show the effect sizes of the PLF difference between quiet and noise. As is clear from the line plots and Figure 10, PLFs diminished as the frequency increased: phase-locking was strongest at 100 Hz and weakest at 2000 Hz (main effect of frequency:  $F(19,63) = 19.654, P < 0.001, \eta^2 = 0.195$ ). Overall, phase-locking was stronger in quiet than in noise (main effect of condition:  $F(1,81) = 10.256, P = 0.002, \eta^2 = 0.112$ ). Additionally, phase-locking was slightly stronger in response to the vowel than the CV transition (main effect of time region:  $F(1,81) = 4.050, P = 0.047, \eta^2 = 0.048$ ). As illustrated in Figure 11, although responses across frequency were more variable in noise and for the response to the CV transition, there were nuances to these effects. For example, low-frequency phase-locking was more degraded by background noise for the CV transition than for the vowel response; higher frequencies were equivalently affected by background noise (frequency  $\times$  condition  $\times$  time region interaction,  $F(19,63) = 3.202, P < 0.001, \eta^2 = 0.491$ ).

Thus, patterns of phase-locking across listening conditions and time regions of the response align with patterns of intertrial variability and intertrial jitter. But does this mean that variability across frequencies shapes variability and jitter patterns? We tested this hypothesis by correlating PLFs at individual frequencies (integers from 100 to 2000 Hz) with intertrial variability and jitter results, covarying for the BF of each ICc recording site. We performed 4 sets of correlations—responses to the CV transition and vowel for responses in quiet and noise. As reported in Table 2, we consistently observed that stronger

PLF's were associated with less variable responses (higher correlation coefficients) and less jittered responses. In quiet these correlations spanned the entire range of phase-locking (up to 2000 Hz). In background noise these correlations tapered off around 1000 Hz, but as shown in Figures 10B and 11 there is a steep drop in phase-locking in the higher frequencies; we think this floor effect accounts for the absence of correlations with variability and jitter.

Together, these results suggest that extracellular activity is variable and jittered across a wide frequency band. Specifically, variability of the phase of extracellular activity up to 2000 Hz relates to the intertrial variability and intertrial jitter recorded at that site. Additionally, patterns of response variability and jitter across time regions of the speech syllable and listening conditions are reproduced by the phase-locking analysis. Because our goal is to see if ICc activity aligns with population activity recorded at the scalp, we next tested response variability patterns recorded at the epidural surface in the guinea pig. We hypothesized that the pattern of response variability would be similar to the local activity patterns observed in ICc.

### Surface Responses to Dynamic Speech Features in Quiet and Noise are Variable

Grand average scalp-recorded responses of the guinea pig surface responses to the speech sound [da] in quiet and in noise are shown in Figure 3B. Animal surface recordings did not have single-trial resolution and so the intertrial jitter measure could not be calculated. At the surface recording, there was an identical pattern of results as a function of time region of the response and condition (Fig. 6C) as was observed in the ICc recordings.

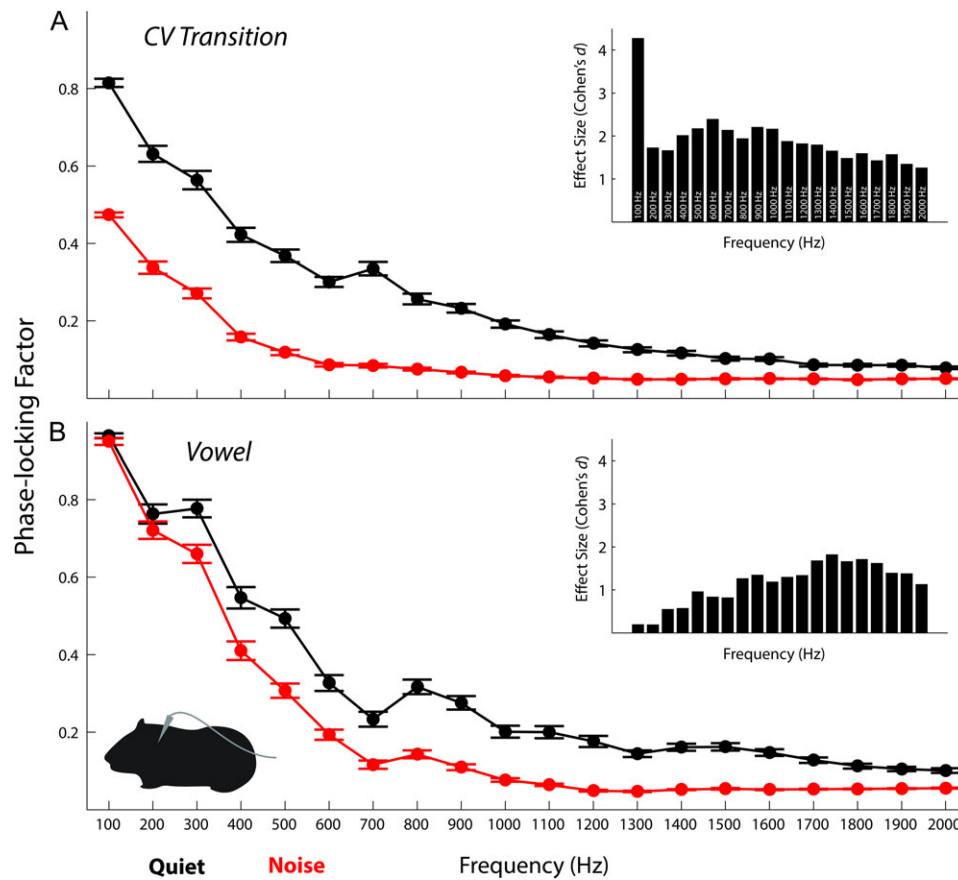
In particular, responses to the vowel were more consistent than responses to the CV transition, which were in turn less variable than responses to the onset (main effect of time region:  $F(2,18) = 124.594, P < 0.001, \eta^2 = 0.933$ ). In addition, response to speech in quiet were less variable than responses to speech in noise (main effect of condition:  $F(1,9) = 32.921, P < 0.001, \eta^2 = 0.785$ ). This noise degradation was exacerbated in response to transient and dynamic speech features, meaning that the responses to the onset and consonant transition were more variable in noise than they were in quiet, relative to the vowel (condition  $\times$  time region interaction:  $F(2,18) = 11.950, P < 0.001, \eta^2 = 0.570$ ).

### Human Component

Grand average scalp-recorded responses of the 50 children to the speech sound [da] in quiet and in noise are shown in Figure 3C. Again, human recordings did not have single-trial resolution and so intertrial jitter could not be calculated in this cohort.

### Human Scalp Responses to Dynamic Speech Features in Quiet and Noise are Variable

The human scalp responses showed an identical pattern of intertrial variability results as a function of time region of the response and condition as was observed in the animal recordings (both in ICc and surface recordings; Fig. 8). Responses to the vowel were less variable (higher intertrial correlation values) than response to the CV transition, which were in turn less variable than responses to the onset (main effect of time



**Figure 11.** Mean phase-locking factors (PLFs) are plotted at each frequency (integers from 100–2000 Hz) in response to the CV transition (A) and vowel (B) in quiet (black) and background noise (gray). These reinforce the observations from Figure 10, namely, that PLFs are lower in noise, lower for the response to the CV transition, and disproportionately lower for the response to the CV transition in noise. The insets show the effect sizes comparing PLFs at each frequency between quiet and noise. Each frequency is indicated by a bar in ascending order (top and bottom panels are organized identically). The strongest quiet-to-noise split is in response to the fundamental frequency (100 Hz) in the CV transition. Slightly larger effects are also seen around the first formant in both responses (~700–800 Hz). For the CV transition, the effect of noise slightly attenuates with increasing frequency, likely because PLFs are already lower to begin with. In contrast, the effect of noise increases with higher frequencies in response to the vowel. Error bars  $\pm 1$  standard error of the mean.

region:  $F(2,98) = 87.976$ ,  $P < 0.001$ ,  $\eta^2 = 0.642$ ). In addition, responses to speech in quiet were less variable than responses to speech in noise (main effect of condition:  $F(1,49) = 76.992$ ,  $P < 0.001$ ,  $\eta^2 = 0.611$ ). This noise degradation was exacerbated in response to transient and dynamic speech features, meaning that the responses to the onset and consonant transition were more variable in noise than they were in quiet, relative to the vowel (condition  $\times$  time region interaction:  $F(2,98) = 4.856$ ,  $P = 0.010$ ,  $\eta^2 = 0.090$ ).

### Comparisons Between Animal and Human Components

#### Patterns of Intertrial Correlations are Similar Between Animal ICc, Animal Surface, and Human Scalp Responses

The preceding results independently compared intertrial variability in responses recorded from animals (ICc and at the surface) and humans (recorded at the scalp). We next tested the hypothesis that these effects are maintained across species using a single mixed-effects model.

We found an identical pattern of results across the animal and human subjects.

Namely, across recording types (animal ICc, animal surface, and human scalp) responses to the vowel were less variable than responses to the CV transition, which were less variable

than responses to the onset ( $F(2,863) = 87.054$ ,  $P < 0.001$ ,  $\eta^2 = 0.170$ ). This effect was equivalent across guinea pigs and humans (no time region  $\times$  species interaction,  $F(2,863) = 0.087$ ,  $P = 0.916$ ,  $\eta^2 < 0.001$ ). In addition, responses in quiet were less variable than responses in noise (main effect of condition:  $F(1,836) = 107.971$ ,  $P < 0.001$ ,  $\eta^2 = 0.112$ ). This effect was equivalent across guinea pigs and humans (no condition  $\times$  species interaction:  $F(1,863) = 0.162$ ,  $P = 0.687$ ,  $\eta^2 < 0.001$ ). Finally, responses to the dynamic and transient speech features were rendered more variable by the addition of background noise than responses to the vowel (time region  $\times$  condition interaction:  $F(2,863) = 9.278$ ,  $P < 0.001$ ,  $\eta^2 = 0.021$ ). This effect was equivalent across guinea pigs and humans (no time region  $\times$  condition  $\times$  species interaction,  $F(2,863) = 2.925$ ,  $P = 0.054$ ,  $\eta^2 = 0.007$ ). These are illustrated in Figure 6D.

Recall that the ICc onset correlations were skewed towards zero. Therefore, these analyses were rerun only comparing responses to the vowel and CV transition. Essentially the same pattern of results emerged. Across species, responses to the vowel were less variable than responses to the CV transition (main effect of time region:  $F(1,575) = 14.213$ ,  $P < 0.001$ ,  $\eta^2 = 0.024$ ; no time region  $\times$  species interaction:  $F(1,575) = 0.129$ ,  $P = 0.720$ ,  $\eta^2 < 0.001$ ). Once again, responses in quiet were less variable than responses in noise (main effect of condition:

**Table 2** The phase-locking factor (PLF) calculated at each frequency was correlated to the intertrial variability and intertrial jitter corresponding to that recording (CV transition or vowel; quiet or noise). Partial correlation coefficients are reported, controlling for the tuned region of each recording site

Frequency (Hz)	Quiet				Noise			
	CV Transition		Vowel		CV Transition		Vowel	
	Variability	Jitter	Variability	Jitter	Variability	Jitter	Variability	Jitter
100	0.643***	-0.561***	0.381***	-0.638***	0.595***	-0.368***	0.384***	-0.705***
200	0.647***	-0.549***	0.454***	-0.451***	0.489***	-0.491***	0.602***	-0.480***
300	0.493***	-0.503***	0.340**	-0.380***	0.436***	-0.389***	0.372***	-0.328***
400	0.513***	-0.443***	0.533**	-0.405***	0.501***	-0.456***	0.386***	-0.167
500	0.397***	-0.413***	0.360***	-0.306**	0.317**	-0.369***	0.381***	-0.411***
600	0.387***	-0.347***	0.376***	-0.415***	0.314**	-0.383***	0.406***	-0.360***
700	0.352***	-0.348***	0.131	-0.333**	0.314**	-0.354***	0.360***	-0.347***
800	0.369***	-0.376***	0.126	-0.310**	0.297**	-0.428***	0.303**	-0.367***
900	0.386***	-0.387***	0.373***	-0.303**	0.066	-0.076	0.244*	-0.318**
1000	0.350***	-0.366***	0.455***	-0.488***	0.169	-0.365***	0.340**	-0.369***
1100	0.440***	-0.354***	0.309**	-0.439***	0.215	-0.190	0.176	-0.228*
1200	0.445***	-0.293**	0.391***	-0.430***	0.028	-0.139	0.179	-0.274*
1300	0.487***	-0.383***	0.340**	-0.378***	0.116	-0.079	0.324**	-0.370***
1400	0.263*	-0.277*	0.352***	-0.467***	0.129	-0.185	0.163	-0.387***
1500	0.408***	-0.307**	0.441***	-0.445***	-0.002	0.104	-0.049	0.016
1600	0.223*	-0.208	0.478***	-0.388***	0.149	-0.296**	0.082	-0.014
1700	0.318**	-0.249*	0.417***	-0.413***	0.087	0.043	0.065	-0.139
1800	0.420***	-0.280*	0.361***	-0.341**	-0.101	0.143	0.081	0.075
1900	0.195	-0.156	0.249*	-0.358***	0.081	-0.071	0.388	-0.205
2000	0.331**	-0.266*	0.330**	-0.365***	0.081	-0.109	0.197	-0.103

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P \leq 0.001$ .

$F(1,575) = 29.332$ ,  $P < 0.001$ ,  $\eta^2 = 0.049$ ); no condition  $\times$  species interaction: ( $F(1,575) = 0.884$ ,  $P = 0.347$ ,  $\eta^2 = 0.002$ ). However, the interaction between condition and time region was no longer statistically reliable (no time region  $\times$  condition interaction:  $F(1,575) = 1.092$ ,  $P = 0.296$ ,  $\eta^2 = 0.002$ ) but, importantly, this was the case for responses in guinea pigs and in humans, reinforcing the consistency of response patterns across species (no species  $\times$  time region  $\times$  condition interaction:  $F(1,575) = 0.063$ ,  $P = 0.802$ ,  $\eta^2 < 0.001$ ).

#### Despite an Overall Similarity, Distinct Recording Sites Exhibit Distinct Response Patterns

While the patterns of results are statistically similar, examination of Figure 6 also shows differences between recording sites. The clearest difference is between the near-field (ICc; Figs 6A,B) and far-field (animal and human surface; Figs 6C,D) sites. In the near-field sites, there is a much more dramatic effect of background noise on variability and jitter in response to the onset than at the surface. The overall pattern of results is the same (onset degradation is greater than CV transition degradation, which is greater than vowel degradation) but the extent to which noise increases variability in response to the onset is substantially greater in ICc. The surface recordings reflect dipoles from a much greater set of generators, including more peripheral brainstem structures (Melcher et al. 1996). As a population this system may have a more resilient onset coding in noise, and our previous work suggested that noise exerts a progressively larger degradation as information ascends the auditory pathway (Cunningham et al. 2002). In contrast, responses to periodic features such as the vowel were largely similar across near- and far-field responses. We speculate that responses at the surface/scalp reflect sustained phase-locking within IC (Rees and Møller 1983; Rees and Palmer 1988) that may explain why the patterns

of response variability for these periodic features are more similar between near-field and far-field recordings.

We also observe discrepancies between recording sites with respect to the magnitude of the correlations. The range of correlations in the animal surface recordings (Fig. 6C) is higher than in the ICc recordings or human scalp recordings (although, again, the overall pattern is similar). This may be attributable to differences in electrode placement. The electrode in the animal surface recordings was on the surface of the dura mater, whereas the electrode in the human recordings was on the scalp.

#### In Humans, Intertrial Variability Relates to Language Development

Previous work has found that children with more variable FFRs to speech have poorer literacy skills than their peers (Hornickel and Kraus 2013). White-Schwoch et al. (2015b) reported that intertrial variability in response to speech in noise, when taken in concert with additional response properties, was strongly predictive of prereading skills.

Here, we found a similar relationship. In a 2-step regression, age, sex, and nonverbal intelligence accounted for 21.8% of variance in phonological processing ( $F(3,46) = 4.265$ ,  $P = 0.010$ ). When the intertrial variability for the response in noise was added to the model, it accounted for an additional 16.5% of variance ( $F(1,45) = 12.029$ ,  $P = 0.001$ ; total  $R^2 = 0.383$ ,  $F(4,47) = 6.973$ ,  $P < 0.001$ ;  $\beta_{\text{correlation}} = 0.424$ ,  $t = 3.468$ ,  $P = 0.001$ ; Table 3). Thus, independent of demographic factors, we observe a systematic link between the variability of the response to speech in noise and early language development.

## Discussion

Here we show that timing jitter is a mechanism that contributes to the (un)reliability of neural coding in auditory



**Table 3.** In humans, intertrial variability predicts early language development.

Predictor	$\Delta R^2$	$\beta$
Step 1	0.218	
Sex <sup>a</sup>		-0.200
Age		0.370**
Nonverbal intelligence		0.151
Step 2	0.165	
Sex <sup>a</sup>		-0.297*
Age		0.416***
Nonverbal intelligence		0.079
Intertrial variability		0.424***
Total R <sup>2</sup>	0.383	

<sup>a</sup>Dummy-coded, female = 0.

\*P < 0.05; \*\*P < 0.01; \*\*\*P = 0.01.

midbrain. We tie this variability into variability of the scalp-recorded FFR in humans, a measure that has been linked to language abilities in children, suggesting a functional consequence of this variability for language development. Together, these results support the hypothesis that excessive variability blurs the neural coding of fine-grained speech features in noise, which may in turn compromise language development.

This is consistent with the idea that variability in neural coding compromises the integrity of that coding (Faisal et al. 2008), and the idea that coding subtle temporal cues in sound, such as interaural time differences or rapidly changing consonants, depends on consistent neural activity (Carr and Konishi 1990; Engineer et al. 2008). This reliability becomes especially important at the population level, where it has been argued that synchronized neural activity between units supports visual object binding (Singer 1999), sound representation in auditory cortex (deCharms and Merzenich 1996), visual-motor synchrony (Lee et al. 2016), and learning (Hebb 1949; Bi and Poo 2001). Similarly, variability is thought to contribute to poor evoked responses in human neurophysiological techniques that operate on a slow timescale, such as cortical evoked potentials (Arieli et al. 1996) and, on an even slower timescale, functional magnetic resonance imaging signals (Fox et al. 2006, 2007). Less is known about the sources of variability in the speech-evoked FFR, but auditory brainstem response studies suggest that jitter is a factor (Starr et al. 1996, 2003). We show that this idea extends to the neural coding of speech vis-à-vis the FFR, and provide a framework to understand variability in the human FFR through the lens of an animal model.

In particular, we show that speech-evoked activity in ICc is more variable in response to transient and dynamic sound elements (onset and CV transition) than in response to static cues (vowel). The addition of background noise increases this variability, particularly in response to transient and dynamic cues. Across all laminae and stimulus features that were tested, this intertrial variability related to the timing jitter between responses, suggesting that variability in the timing of neural coding constrains the precise neural coding of acoustic cues. These patterns of variability are similar across a wide range of response frequencies. In addition, patterns of response variability across speech elements and conditions were mirrored between near-field ICc and human scalp recordings; in the latter, intertrial variability relates to language development. Taken together with evidence that listeners with poor auditory processing exhibit variable auditory-neurophysiological responses to speech, these results suggest that timing jitter at

the level of auditory midbrain contributes to these listening difficulties.

### Timing Jitter Emerges as Variable Scalp-Recorded Responses

A major finding of this report is that effects of time region (onset, CV transition, vowel) and background noise are equivalent across tonotopic sites in ICc, and that the relationship between intertrial variability and intertrial timing jitter holds across recording sites. Responses were elicited at high sound levels, well above the threshold of each recording site, which likely contribute to the weak effects of tonotopy in our data. Moreover, whole-cell recordings from the awake bat IC suggest that many IC cells' synaptic inputs are tuned over at least 2 octaves (Xie et al. 2007). Thus, insofar as cellular response properties in the bat midbrain generalize to other species, when considering responses to broadband, spectrotemporally rich stimuli such as speech presented at high sound intensities, there may be relatively weak effects of tuning. That said, we caution that our dataset represents only about the lower third of tuned laminae in the guinea pig ICc (Malmierca et al. 1995; see also Schreiner and Langner 1997). Additionally, cochlear tuning may be sharper in primates (Nelson et al. 2009; Nelson and Young 2010), and in humans there are complex interactions between frequency and intensity (Plack and Oxenham 1998; Shera et al. 2002), tempering generalization of our results from the animal model to humans (but see Ruggero and Temchin, 2005). We also note that midbrain synaptic tuning can be broader than spiking output (Geis and Borst 2009; Geis et al. 2011). Future work can explore this issue with respect to variability across IC laminae more thoroughly by deriving input-output functions at multiple sound levels.

Our ICc recordings may reflect both presynaptic and postsynaptic activity across multiple units (i.e., both local field potentials and spikes) because we did not low or high-pass filter the responses, and spikes can contribute to high-frequency components of extracellular potentials (Buzsáki et al. 2012). There is a diversity of cell types within ICc (Oliver 2005). Our multiunit recordings may reflect this diversity. However, recordings from the central nucleus and dorsal cortex of IC suggest that response properties may not be correlated to cell morphology (Tan and Borst 2007; Tan et al. 2007; although there are exceptions, Geis and Borst 2013). In fact, neighboring cells in the central nucleus that exhibit similar tuning can differ widely in their evoked response properties (Seshagiri and Delgutte 2007). Thus, while we hypothesize that the neural activity patterns we show would be evident across multiple IC nuclei, this remains an open question.

Nevertheless, we suggest that the similarity of our effects across recordings reinforces the idea that human scalp recordings (that are by nature population responses) reflect similar phenomena, and may indeed index timing jitter (cf. Lin et al. 2015). There is reason to believe this phenomenon extends beyond the auditory system: some individuals with dyslexia have poor visual perception in noise (Sperling et al. 2005), and individuals with autism exhibit variable evoked potentials across sensory modalities (Dinstein et al. 2012). Jittered timing in sensory systems may underlie these phenomena (Churchland et al. 2010) and, perhaps, impose constraints on language development.

The ICc results present a framework to contextualize observations made in human listeners. It was estimated that 1 ms of timing jitter decreases the intertrial correlation by

approximately 0.3 (Pearson's  $r$ ). This facilitates new understanding of work in humans that establishes group differences with regards to intertrial variability. For example, [Anderson et al. \(2012\)](#) reported that older adults with normal hearing had intertrial correlations 0.15 lower than young adults in response to speech in quiet. Based on our animal model, this corresponds to roughly 0.5 ms of timing jitter in ICc. [Hornickel and Kraus \(2013\)](#) reported that children with high versus low reading achievement had a similar split in terms of their FFR intertrial correlations. In an intervention study, [Hornickel et al. \(2012\)](#) investigated the use of classroom assistive listening devices in children with dyslexia; they reported gains in intertrial correlations of up 0.15, which again suggests a decrease in timing jitter of about 0.5 ms. Future work can use techniques such as computational modeling to estimate the point at which the jitter becomes sufficiently excessive to ablate an averaged scalp-recorded potential ([Starr et al. 2003](#)).

### Stimulus-Driven Factors: Dynamic Speech Features in Noise

Neural response variability and timing jitter differed markedly in different time regions of the response. Responses to the vowel were most stable, followed by responses to the consonant transition, followed by responses to the consonant onset. These effects of stimulus features became more pronounced with the addition of background noise. These results align with perceptual and neurophysiological studies that document the acoustic vulnerability of transient speech features, especially in noise.

For example, perceptual studies show that children and adults have more trouble recognizing consonants in noise than they do vowels ([Miller and Nicely 1955](#); [Cutler et al. 2004](#); [Nishi et al. 2010](#)). Similar observations have been made in neurophysiological investigations ([Cunningham et al. 2002](#); [White-Schwoch, et al. 2015a](#)); our results are consistent with the hypothesis that noise disrupts the neural coding of these transient cues in the subcortical auditory system. As compared with consonants, vowels are periodic, of longer duration, and have a higher amplitude, making them easier to perceive and lock onto, and conferring a higher SNR ratio when background noise is added. These differences are evident in [Figure 1](#), where the [da] stimulus is illustrated. The vowel is the periodic segment of the stimulus, which is higher in peak intensity than the onset or CV transition; thus, when background noise was added, the vowel segment of the stimulus had a higher SNR.

Together, these acoustic factors likely contribute to the effects we observed, and, crucially, we show that midbrain coding is more variable when the acoustic input is compromised. This variability may blur the representation of these features in noise, undermining their intelligibility. It should, however, be noted that we only employed a single instance of the sound [da]; a next step for this work is to determine the generality of our findings across other of speech tokens and listening conditions (such as different types of background noise), in addition to more complex syllables and words. More complex stimuli could also disentangle onset coding from CV transition and vowel coding (such as comparing [da] to [ada]; cf. [Cunningham et al. 2002](#); [Cutler et al. 2004](#)). Noteworthy is that previous findings from rat IC and auditory cortex suggest a shared spike-timing code across consonants, including voiced and unvoiced stops, liquids, fricatives, nasals, and glides ([Engineer et al. 2008](#); [Ranasinghe et al. 2013](#)). In contrast, recordings from rat auditory cortex suggest that spike count (not variability) codes

vowel sounds ([Perez et al. 2013](#)). It is plausible that background noise disrupts the precision of timing information while preserving (if not increasing) overall spike count, thereby leaving vowel coding relatively unaffected. This hypothesis can be tested with respect to variability and jitter in IC in future work employing a more diverse stimulus set.

### Local Factors: Inhibitory and Excitatory Neurotransmission

One hypothesis to explain central processing bottlenecks that constrain auditory processing comes from the literature on auditory aging. Rodent models document extensive and profound declines in inhibitory neurotransmitter receptors (both  $\gamma$ -Aminobutyric acid [GABA] and glycine receptors) throughout the auditory neuraxis, including in IC ([Milbrandt et al. 1997](#); [Walton et al. 1998](#); [Tadros et al. 2007](#)). In the aged rhesus macaque, [Engle et al. \(2014\)](#) document increases in parvalbumin (PV) in ICc. PV is a calcium-binding protein that is expressed by GABAergic and glycinergic neurons; this up-regulation may compensate for the overall decrease in inhibitory activity in IC. [Casparly and colleagues \(2008\)](#) argue that age-related loss of GABAergic receptors, in particular, may be responsible for many of the speech understanding difficulties older adults experience. There is also evidence that GABAergic input is necessary to encode dynamic frequency content, and that it mediates the variability observed in first-spike latency in IC ([Park and Pollak 1993](#)). In addition, auditory training in a rat model of aging causes activity-driven increases in inhibitory neurotransmitter function commensurate with more consistent spike timing in auditory cortex ([de Villers-Sidani et al. 2010](#)).

The timing jitter we document in response to fast-changing sounds in noise is evocative of evidence in humans that suggest difficulty understanding fast-changing sounds, especially in noisy environments ([Gordon-Salant and Fitzgibbons 1993](#); [Pichora-Fuller et al. 2007](#); [White-Schwoch et al. 2015a](#)). We speculate that in disordered systems (including older adults and children with language impairment) excessive jitter may be due to a loss, or abnormal development, of inhibition.

Less is known about neuropharmacological expression in the central auditory systems of animal models of language impairment. [Wright and Zecker \(2004\)](#) propose a model whereby maturational delay in central processing underlies developmental disorders of auditory processing and language. [Sanchez et al. \(2015\)](#) suggest the number and/or distribution of N-methyl-D-aspartate receptors (NMDA-Rs) mediates spike-timing variability in brainstem and midbrain, and note that application of an NMDA-Rs agonist reduces the number of action potentials in response to sound onsets ([Sanchez et al. 2007](#)). Glutamate, an excitatory neurotransmitter, binds to NMDA-Rs. NMDA-Rs are down-regulated maturationally in IC, and eventually fast-latency subunits proliferate. Maturational delays in this process may lead to abnormal neural firing in response to speech, which we speculate affects language development. Thus, one possibility is that poor processing due to language impairment may be attributed more to problems of excitation, whereas processing bottlenecks due to aging may be rooted in aberrant inhibitory function.

Perhaps more likely a balance of excitation and inhibition is necessary for consistent neural firing ([Wehr and Zador 2003](#)), especially in response to fast-changing sounds in noise. This view is consistent with evidence for abnormal concentrations of both GABA and glutamate in children with

neurodevelopmental disorders (Edden et al. 2012; Pugh et al. 2014; Braat and Kooy 2015).

### Utility of the Animal Model

Response patterns measured in IC, at the cortical surface, and in humans were broadly similar as a function of stimulus features (consonant vs. vowel and quiet vs. noise). To our knowledge, this is the first direct comparison of the speech-evoked FFR in humans and an animal model using identical stimuli. While there were differences in response properties across the three sites (Fig. 3), it is important to emphasize the similarities because they hint at the biological origins of the FFR: our results are consistent with the hypothesis that the FFR reflects IC activity (Chandrasekaran and Kraus 2010). This was observed when near-field IC activity was compared with surface recordings in the animal model that were filtered for the FFR spectrum, and when the two were compared with recordings in humans. It should be noted, however, that this evidence is correlational, and future work employing techniques such as cooling and lesioning will help explore this hypothesis more thoroughly.

### Listening Experience and Language Impairment

The auditory system is highly interconnected (Kral and Eggermont 2007) and throughout life the inferior colliculi are tuned bidirectionally from modulatory ascending and descending input. The descending fibers are thought to mediate functional remodeling in certain tasks (Bajo et al. 2010). Our view is that auditory learning is the product of coordinated input from cognitive, sensory, and reward networks, and we are therefore motivated by a conceptual framework whereby the FFR reflects an individual's life in sound—for better or for worse—and success in everyday speech communication (Kraus and White-Schwoch 2015; White-Schwoch and Kraus forthcoming). In humans, auditory training has been associated with increased intertrial stability of the FFR. This includes music training, second language experience, and the use of assistive listening devices (reviewed in Kraus and Nicol 2014; Kraus and White-Schwoch forthcoming). We now attribute these gains to decreases in trial-by-trial timing variability of neural activity patterns.

It has long been argued that imprecise timing in the central auditory system is a major cause of language impairment (Tallal and Piercy 1973; Nagarajan et al. 1999; Ahissar et al. 2000; Benasich and Tallal 2002). A challenge to this view, however, has been the difficulty of directly comparing neurophysiological activity in humans with language disorders and animal models. The approach presented here may offer an avenue to bridge this gap. For example, Centanni and colleagues (2014a, 2014b) found timing variability in auditory cortex in response to speech sounds—a variability evocative of that we document in ICc, and what has been demonstrated in humans with reading impairment using scalp recordings (Hornickel and Kraus 2013; White-Schwoch et al. 2015b). Some cases of dyslexia can be attributed to mutations in the gene KIAA0319 (Cope et al. 2005; Galaburda et al. 2006), and Centanni et al. used RNA interference to reduce expression of the rat homolog *Kiaa0319*. This gene is noteworthy because it is expressed in both brainstem and cortex, and its knockdown is associated with abnormalities in axonal migration (Platt et al. 2013) and dendritic morphology and orientation (Peschansky et al. 2009). Given the similarity of FFR activity patterns between the animal model and humans,

studies of language impairment could compare IC activity and the FFR in animal models such as the aforementioned *Kiaa0319* knockdown rat directly to humans using genetic markers.

### Limitations and Future Directions

It is important to point out some technical differences between the animal and human recordings. The animals were anesthetized for the recording session, and the anesthetic may have affected response properties; ketamine is an NMDA-R antagonist and so may have contributed to variability in first-spike latency. However, Ter-Mikaelian et al. (2007) have shown that isoflurane, also an NMDA-R antagonist, minimally affects timing variability in IC. In addition, there were differences in the stimulus delivery: there was a slightly faster presentation rate for the animal component, and the ear inserts used in the animal component had a broader frequency response. Yet despite these differences a remarkable similarity was established between responses recorded from ICc in guinea pigs and the scalp in humans.

We note that both the animal and human subjects represented relatively homogenous cohorts, although individual differences in human listeners were linked to language development. Still, future work should evaluate intertrial timing jitter as a function of listening experience and environment. Although previous work suggests that a modified milieu and/or explicit training alters response properties (Engineer et al. 2004; Centanni et al. 2014b), the link is missing from the current study. An additional open question is the relationship between timing jitter in IC and behavior in awake and functioning animals. These questions represent exciting avenues for future work that can further our understanding of phenomena observed in human listeners.

Finally, an important consideration is that any interspecies comparison is inherently correlational. Moreover, many factors contribute to language development, and our results indicate that scalp-recorded FFR variability accounted for 16.5% of unique variance in phonological skills. Additionally, in the animal model, intertrial jitter accounted for 19.5% of variance in the evoked response morphology after removing the influence of stimulus-related factors and the tuning of the recording site. We also note that stimulus regions and background noise strongly affected both variability measures (correlation and jitter). When not controlling for stimulus factors, jitter and variability share 73.2% variance; additionally, both jitter and variability relate to phase-locking across frequencies. Therefore, results from human and animal data reflect a chain of moderately correlated phenomena, and FFR variability viewed at the scalp is likely also influenced by factors other than jitter. Future work should consider how these factors, in combination with many more, emerge in scalp-recorded potentials, and the developmental impact of variability on language development.

### Conclusion

We show that timing variability in auditory midbrain constrains the neural coding of speech, and that transient and dynamic speech cues in noise are especially susceptible to this jitter. In turn, we show a consistent and profound relationship between timing jitter and the trial-to-trial stability of neural coding: less-jittered responses are morphologically more similar across trials. In humans, the degree of intertrial variability (and, presumably, jitter) relates to early language skills.



Together, this report defines timing jitter as a mechanism contributing to the phenomenon of variable auditory processing documented in humans with listening difficulties, including children with language impairment and older adults.

## Funding

Supported by the National Institutes of Health (R01 DC01510 & R01 HD069414) and the Knowles Hearing Center.

## Notes

We thank Jason Tait Sanchez for his input on this work and comments on an earlier draft of the article. We also thank members of the Auditory Neuroscience Laboratory for their assistance with data collection and Jennifer Krizman for her input. *Conflict of Interest:* N.K. is Chief Scientific Officer of Synaural, a company working to develop a user-friendly measure of auditory processing.

## References

- Abrams DA, Lynch CJ, Cheng KM, Phillips J, Supekar K, Ryali S, Uddin LQ, Menon V. 2013. Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proc Natl Acad Sci.* 110:12060–12065.
- Ahissar M, Protopapas A, Reid M, Merzenich MM. 2000. Auditory processing parallels reading abilities in adults. *Proc Natl Acad Sci.* 97:6832–6837.
- Alcántara JI, Weisblatt EJ, Moore BC, Bolton PF. 2004. Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *J Child Psychol Psychiatry.* 45:1107–1114.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N. 2012. Aging affects neural precision of speech encoding. *J Neurosci.* 32:14156–14164.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N. 2015. Development of subcortical speech representation in human infants. *J Acoust Soc Am.* 137:3346–3355.
- Arieli A, Sterkin A, Grinvald A, Aertsen A. 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science.* 273:1868–1871.
- Bajo VM, Nodal FR, Moore DR, King AJ. 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci.* 13:253–260.
- Benasich AA, Tallal P. 2002. Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res.* 136:31–49.
- Bi G, Poo M. 2001. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu Rev Neurosci.* 24:139–166.
- Bishop DV. 1997. *Uncommon understanding* (Classic edition): development and disorders of language comprehension in children. New York, NY: Psychology Press.
- Bishop DV, Adams C. 1990. A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *J Child Psychol Psychiatry.* 31:1027–1050.
- Braat S, Kooy RF. 2015. The GABA<sub>A</sub> receptor as a therapeutic target for neurodevelopmental disorders. *Neuron.* 86:1119–1130.
- Bradlow AR, Kraus N, Hayes E. 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res.* 46:80–97.
- Buzsáki G, Anastassiou CA, Koch C. 2012. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci.* 13:407–420.
- Cai R, Caspary D. 2015. GABAergic inhibition shapes SAM responses in rat auditory thalamus. *Neuroscience.* 299:146–155.
- Carr C, Konishi M. 1990. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J Neurosci.* 10:3227–3246.
- Caspary DM, Hughes LF, Ling LL. 2013. Age-related GABA<sub>A</sub> receptor changes in rat auditory cortex. *Neurobiol Aging.* 34:1486–1496.
- Caspary DM, Ling L, Turner JG, Hughes LF. 2008. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J Exp Biol.* 211:1781–1791.
- Caspary DM, Milbrandt JC, Helfert RH. 1995. Central auditory aging: GABA changes in the inferior colliculus. *Exp Gerontol.* 30:349–360.
- Centanni TM, Booker A, Sloan A, Chen F, Maher B, Carraway R, Khodaparast N, Rennaker R, LoTurco J, Kilgard M. 2014a. Knockdown of the dyslexia-associated gene *Kiaa0319* impairs temporal responses to speech stimuli in rat primary auditory cortex. *Cereb Cortex.* 24:1753–1766.
- Centanni TM, Chen F, Booker AM, Engineer CT, Sloan AM, Rennaker RL, LoTurco JJ, Kilgard MP. 2014b. Speech sound processing deficits and training-induced neural plasticity in rats with dyslexia gene knockdown. *PLoS One.* 9:e98439.
- Chandrasekaran B, Kraus N. 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology.* 47:236–246.
- Churchland MM, Byron MY, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, et al. 2010. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci.* 13:369–378.
- Cohen J. 1988. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen J, Cohen P, West SG, Aiken LS. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences.* 3rd ed. Mahwah, NJ: Erlbaum.
- Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, Owen MJ, O'Donovan MC, Williams J. 2005. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet.* 76:581–591.
- Cunningham J, Nicol T, King C, Zecker SG, Kraus N. 2002. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res.* 169:97–111.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol.* 112:758–767.
- Cunningham J, Nicol T, Zecker S, Kraus N. 2000. Speech-evoked neurophysiologic responses in children with learning problems: development and behavioral correlates of perception. *Ear Hear.* 21:554–568.
- Cutler A, Weber A, Smits R, Cooper N. 2004. Patterns of English phoneme confusions by native and non-native listeners. *J Acoust Soc Am.* 116:3668–3678.
- deCharms RC, Merzenich MM. 1996. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature.* 381:13.
- de Villers-Sidani E, Alzghoul L, Zhou X, Simpson KL, Lin R, Merzenich MM. 2010. Recovery of functional and structural

- age-related changes in the rat primary auditory cortex with operant training. *Proc Natl Acad Sci.* 107:13900–13905.
- Dinstein I, Heeger DJ, Lorenzi L, Minshew NJ, Malach R, Behrmann M. 2012. Unreliable evoked responses in autism. *Neuron.* 75:981–991.
- Dubno JR, Dirks DD, Morgan DE. 1984. Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am.* 76:87–96.
- Edden RA, Crocetti D, Zhu H, Gilbert DL, Mostofsky SH. 2012. Reduced GABA concentration in attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry.* 69:750–753.
- Engineer C, Centanni T, Im K, Borland M, Moreno N, Carraway R, Wilson L, Kilgard M. 2014. Degraded auditory processing in a rat model of autism limits the speech representation in non-primary auditory cortex. *Dev Neurobiol.* 74:972–986.
- Engineer CT, Perez CA, Chen YTH, Carraway RS, Reed AC, Shetake JA, Jakkamsetti V, Chang KQ, Kilgard MP. 2008. Cortical activity patterns predict speech discrimination ability. *Nat Neurosci.* 11:603–608.
- Engineer CT, Rahebi KC, Borland MS, Buell EP, Centanni TM, Fink MK, Im KW, Wilson LG, Kilgard MP. 2015. Degraded neural and behavioral processing of speech sounds in a rat model of Rett syndrome. *Neurobiol Dis.* 83:26–34.
- Engineer ND, Percaccio CR, Pandya PK, Moucha R, Rathbun DL, Kilgard MP. 2004. Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *J Neurophysiol.* 92:73–82.
- Engle JR, Gray DT, Turner H, Udell JB, Recanzone GH. 2014. Age-related neurochemical changes in the rhesus macaque inferior colliculus. *Front Aging Neurosci.* 6:73.
- Faisal AA, Selen LP, Wolpert DM. 2008. Noise in the nervous system. *Nat Rev Neurosci.* 9:292–303.
- Fay RR. 1988. Hearing in vertebrates: a psychophysics databook. IL: Hill-Fay Associates Winnetka.
- Fox MD, Snyder AZ, Vincent JL, Raichle ME. 2007. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron.* 56:171–184.
- Fox MD, Snyder AZ, Zacks JM, Raichle ME. 2006. Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nat Neurosci.* 9:23–25.
- Füllgrabe C, Moore BC, Stone MA. 2015. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front Aging Neurosci.* 6:347.
- Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. 2006. From genes to behavior in developmental dyslexia. *Nat Neurosci.* 9:1213–1217.
- Geis H, Borst JGG. 2013. Large GABAergic neurons form a distinct subclass within the mouse dorsal cortex of the inferior colliculus with respect to intrinsic properties, synaptic inputs, sound responses, and projections. *J Comp Neurol.* 521:189–202.
- Geis H-R, Borst JGG. 2009. Intracellular responses of neurons in the mouse inferior colliculus to sinusoidal amplitude-modulated tones. *J Neurophysiol.* 101:2002–2016.
- Geis H, Rüdiger A, van der Heijden M, Borst JGG. 2011. Subcortical input heterogeneity in the mouse inferior colliculus. *J Physiol.* 589:3955–3967.
- Gervais H, Belin P, Boddarta N, Leboyer M, Coez A, Sfaello I, Barthélémy C, Brunelle F, Samson Y, Zilbovicius M. 2004. Abnormal cortical voice processing in autism. *Nat Neurosci.* 7:801–802.
- Gordon-Salant S. 2014. Aging, hearing loss, and speech recognition: stop shouting, I can't understand you. In: Popper AN, Fay RR, editors. *Perspectives on auditory research. Springer handbook of auditory research.* Springer. p. 211–228.
- Gordon-Salant S, Fitzgibbons PJ. 1993. Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Lang Hear Res.* 36:1276.
- Gordon-Salant S, Yeni-Komshian GH, Fitzgibbons PJ. 2010. Recognition of accented English in quiet and noise by younger and older listeners. *J Acoust Soc Am.* 128:3152–3160.
- Goswami U. 2014. Sensory theories of developmental dyslexia: three challenges for research. *Nat Rev Neurosci.* 16:43–54.
- Hebb DO. 1949. *The organization of behavior: a neuropsychological theory.* New York, NY: Wiley.
- Hornickel J, Kraus N. 2013. Unstable representation of sound: A biological marker of dyslexia. *J Neurosci.* 33:3500–3504.
- Hornickel J, Zecker SG, Bradlow AR, Kraus N. 2012. Assistive listening devices drive neuroplasticity in children with dyslexia. *Proc Natl Acad Sci.* 109:16731–16736.
- Johnson KL, Nicol T, Zecker SG, Kraus N. 2008. Developmental plasticity in the human auditory brainstem. *J Neurosci.* 28:4000.
- Kral A, Eggermont JJ. 2007. What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res Rev.* 56:259–269.
- Kraus N, Bradlow AR, Cheatham MA, Cunningham J, King CD, Koch DB, Nicol TG, McGee TJ, Stein L, Wright BA. 2000. Consequences of neural asynchrony: a case of auditory neuropathy. *J Assoc Res Otolaryngol.* 1:33–45.
- Kraus N, McGee T, Carrell T, King C, Littman T, Nicol T. 1994. Discrimination of speech-like contrasts in the auditory thalamus and cortex. *J Acoust Soc Am.* 96:2758–2768.
- Kraus N, Nicol T. 2014. *The cognitive auditory system.* In: Fay R, Popper A, editors. *Perspectives on auditory research. Springer handbook of auditory research.* Heidelberg: Springer-Verlag. p. 299–319.
- Kraus N, White-Schwoch T. forthcoming. Neurobiology of everyday communication: what have we learned from music? *The Neuroscientist.* doi:10.1177/1073858416653593
- Kraus N, White-Schwoch T. 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn Sci.* 19:642–654.
- Kuhl PK, Coffey-Corina S, Padden D, Dawson G. 2005. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Dev Sci.* 8:F1–F12.
- Lee J, Joshua M, Medina JF, Lisberger SG. 2016. Signal, noise, and variation in neural and sensory-motor latency. *Neuron.* 90:165–176.
- Lin I-C, Okun M, Carandini M, Harris KD. 2015. The nature of shared cortical variability. *Neuron.* 87:644–656.
- Liu L-F, Palmer AR, Wallace MN. 2006. Phase-locked responses to pure tones in the inferior colliculus. *J Neurophysiol.* 95:1926–1935.
- Mainen ZF, Sejnowski TJ. 1995. Reliability of spike timing in neocortical neurons. *Science.* 268:1503–1506.
- Malmierca MS, Rees A, Le Beau FE, Bjaalie JG. 1995. Laminar organization of frequency-defined local axons within and between the inferior colliculi of the guinea pig. *J Comp Neurol.* 357:124–144.
- McGee T, Kraus N, King C, Nicol T, Carrell TD. 1996. Acoustic elements of speechlike stimuli are reflected in surface recorded responses over the guinea pig temporal lobe. *J Acoust Soc Am.* 99:3606–3614.
- Melcher JR, Guinan JJ, Knudson IM, Kiang NY. 1996. Generators of the brainstem auditory evoked potential in cat. II.

- Correlating lesion sites with waveform changes. *Hear Res.* 93:28–51.
- Messaoud-Galusi S, Hazan V, Rosen S. 2011. Investigating speech perception in children with dyslexia: is there evidence of a consistent deficit in individuals? *J Speech Lang Hear Res.* 54:1682–1701.
- Milbrandt JC, Hunter C, Caspary DM. 1997. Alterations of GABA<sub>A</sub> receptor subunit mRNA levels in the aging Fischer 344 rat inferior colliculus. *J Comp Neurol.* 379:455–465.
- Miller GA, Nicely PE. 1955. An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am.* 27:338–352.
- Moore DR, Ferguson MA, Edmondson-Jones AM, Ratib S, Riley A. 2010. Nature of auditory processing disorder in children. *Pediatrics.* 126:e382–e390.
- Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM. 1999. Cortical auditory signal processing in poor readers. *Proc Natl Acad Sci.* 96:6483–6488.
- Nelson PC, Smith ZM, Young ED. 2009. Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. *J Neurosci.* 29:2553–2562.
- Nelson PC, Young ED. 2010. Neural correlates of context-dependent perceptual enhancement in the inferior colliculus. *J Neurosci.* 30:6577–6587.
- Nishi K, Lewis DE, Hoover BM, Choi S, Stelmachowicz PG. 2010. Children's recognition of American English consonants in noise. *J Acoust Soc Am.* 127:3177–3188.
- Oliver DL. 2005. Neuronal organization in the inferior colliculus. In: *The inferior colliculus*. New York, NY: Springer. p. 69–114.
- Park TJ, Pollak GD. 1993. GABA shapes a topographic organization of response latency in the mustache bat's inferior colliculus. *J Neurosci.* 13:5172–5187.
- Parthasarathy A, Bartlett E. 2011. Age-related auditory deficits in temporal processing in F-344 rats. *Neuroscience.* 192:619–630.
- Perez CA, Engineer CT, Jakkamsetti V, Carraway RS, Perry MS, Kilgard MP. 2013. Different timescales for the neural coding of consonant and vowel sounds. *Cereb Cortex.* 23:670–683.
- Peschansky VJ, Burbridge TJ, Volz AJ, Fiondella C, Wissner-Gross Z, Galaburda AM, Turco JLL, Rosen GD. 2009. The effect of variation in expression of the candidate dyslexia susceptibility gene homolog *Kiaa0319* on neuronal migration and dendritic morphology in the rat. *Cereb Cortex.* 20:884–897.
- Pichora-Fuller MK, Schneider BA, Daneman M. 1995. How young and old adults listen to and remember speech in noise. *J Acoust Soc Am.* 97:593.
- Pichora-Fuller MK, Schneider BA, MacDonald E, Pass HE, Brown S. 2007. Temporal jitter disrupts speech intelligibility: a simulation of auditory aging. *Hear Res.* 223:114–121.
- Plack CJ, Barker D, Prendergast G. 2014. Perceptual consequences of “Hidden” hearing loss. *Trends Hear.* 18:1–11.
- Plack CJ, Oxenham AJ. 1998. Basilar-membrane nonlinearity and the growth of forward masking. *J Acoust Soc Am.* 103:1598–1608.
- Platt M, Adler W, Mehlhorn A, Johnson G, Wright K, Choi R, Tsang W, Poon M, Yeung S, Wayne M, et al. 2013. Embryonic disruption of the candidate dyslexia susceptibility gene homolog *Kiaa0319*-like results in neuronal migration disorders. *Neuroscience.* 248:585–593.
- Presacco A, Jenkins K, Liberman R, Anderson S. 2015. Effects of aging on dynamic and static encoding of speech. *Ear Hear.* 36:352–363.
- Pugh KR, Frost SJ, Rothman DL, Hoeft F, Del Tufo SN, Mason GF, Molfese PJ, Mencl WE, Grigorenko EL, Landi N, et al. 2014. Glutamate and choline levels predict individual differences in reading ability in emergent readers. *J Neurosci.* 34:4082–4089.
- Ranasinghe KG, Vrana WA, Matney CJ, Kilgard MP. 2013. Increasing diversity of neural responses to speech sounds across the central auditory pathway. *Neuroscience.* 252:80–97.
- Rees A, Møller AR. 1983. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear Res.* 10:301–330.
- Rees A, Palmer AR. 1988. Rate-intensity functions and their modification by broadband noise for neurons in the guinea pig inferior colliculus. *J Acoust Soc Am.* 83:1488–1498.
- Richardson BD, Ling LL, Uteshev VV, Caspary DM. 2013. Reduced GABA<sub>A</sub> receptor-mediated tonic inhibition in aged rat auditory thalamus. *J Neurosci.* 33:1218–1227.
- Rosen S. 2003. Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything? *J Phon.* 31:509–527.
- Ruggero MA, Temchin AN. 2005. Unexceptional sharpness of frequency tuning in the human cochlea. *Proc Natl Acad Sci USA.* 102:18614–18649.
- Ruggles D, Bharadwaj H, Shinn-Cunningham BG. 2012. Why middle-aged listeners have trouble hearing in everyday settings. *Curr Biol.* 1417–1422.
- Russo N, Skoe E, Trommer B, Nicol T, Zecker SG, Bradlow AR, Kraus N. 2008. Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clin Neurophysiol.* 119:1720–1731.
- Russo N, Zecker S, Trommer B, Chen J, Kraus N. 2009. Effects of background noise on cortical encoding of speech in autism spectrum disorders. *J Autism Dev Disord.* 39:1185–1196.
- Sanchez J, Ghelani S, Otto-Meyer S. 2015. From development to disease: diverse functions of NMDA-type glutamate receptors in the lower auditory pathway. *Neuroscience.* 285:248–259.
- Sanchez JT, Gans D, Wenstrup JJ. 2007. Contribution of NMDA and AMPA receptors to temporal patterning of auditory responses in the inferior colliculus. *J Neurosci.* 27:1954–1963.
- Schoof T, Rosen S. 2014. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Front Aging Neurosci.* 6:307.
- Schreiner CE, Langner G. 1997. Laminar fine structure of frequency organization in auditory midbrain. *Nature.* 388:383–386.
- Sergeyenko Y, Lall K, Liberman MC, Kujawa SG. 2013. Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J Neurosci.* 33:13686–13694.
- Seshagiri CV, Delgutte B. 2007. Response properties of neighboring neurons in the auditory midbrain for pure-tone stimulation: a tetrode study. *J Neurophysiol.* 98:2058–2073.
- Shera CA, Guinan JJ, Oxenham AJ. 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci.* 99:3318–3323.
- Singer W. 1999. Neuronal synchrony: a versatile code for the definition of relations? *Neuron.* 24:49–65.
- Skoe E, Krizman J, Anderson S, Kraus N. 2015. Stability and plasticity of auditory brainstem function across the lifespan. *Cereb Cortex.* 25:1415–1426.
- Sperling AJ, Lu Z-L, Manis FR, Seidenberg MS. 2005. Deficits in perceptual noise exclusion in developmental dyslexia. *Nat Neurosci.* 8:862–863.

- Starr A, Michalewski HJ, Zeng F-G, Fujikawa-Brooks S, Linthicum F, Kim CS, Winnier D, Keats B. 2003. Pathology and physiology of auditory neuropathy with a novel mutation in the MPZ gene (Tyr145→ Ser). *Brain*. 126:1604–1619.
- Starr A, Picton TW, Sininger Y, Hood LJ, Berlin CI. 1996. Auditory neuropathy. *Brain*. 119:741–753.
- Syka J, Popelář J, Kvašňák E, Astl J. 2000. Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. *Exp Brain Res*. 133:254–266.
- Tadros SF, D'Souza M, Zettel ML, Zhu X, Waxmonsky NC, Frisina RD. 2007. Glutamate-related gene expression changes with age in the mouse auditory midbrain. *Brain Res*. 1127:1–9.
- Tallal P, Piercy M. 1973. Defects of non-verbal auditory perception in children with developmental aphasia. *Nature*. 241:468–469.
- Tan M, Borst JGG. 2007. Comparison of responses of neurons in the mouse inferior colliculus to current injections, tones of different durations, and sinusoidal amplitude-modulated tones. *J Neurophysiol*. 98:454–466.
- Tan M, Theeuwes HP, Feenstra L, Borst JGG. 2007. Membrane properties and firing patterns of inferior colliculus neurons: an in vivo patch-clamp study in rodents. *J Neurophysiol*. 98:443–453.
- Ter-Mikaelian M, Sanes DH, Semple MN. 2007. Transformation of temporal properties between auditory midbrain and cortex in the awake Mongolian gerbil. *J Neurosci*. 27:6091–6102.
- Tierney AT, Kraus N. 2013. The ability to move to a beat is linked to the consistency of neural responses to sound. *J Neurosci*. 33:14981–14988.
- Van Engen KJ, Bradlow AR. 2007. Sentence recognition in native-and foreign-language multi-talker background noise. *J Acoust Soc Am*. 121:519.
- Walton JP, Frisina RD, O'Neill WE. 1998. Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse. *J Neurosci*. 18:2764–2776.
- Warrier CM, Abrams DA, Nicol TG, Kraus N. 2011. Inferior colliculus contributions to phase encoding of stop consonants in an animal model. *Hear Res*. 282:108–118.
- Wehr M, Zador AM. 2003. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*. 426:442–446.
- White-Schwoch T, Davies EC, Thompson EC, Woodruff Carr K, Nicol T, Bradlow AR, Kraus N. 2015a. Auditory-neurophysiological responses to speech during early childhood: effects of background noise. *Hear Res*. 328:34–47.
- White-Schwoch T, Kraus N. forthcoming. The Janus face of auditory learning: how everyday experience shapes communication. In: Kraus N, Anderson S, White-Schwoch T, Popper AN, Fay RR, editors. *The frequency-following response: a window to human communication*. Springer handbook of auditory research. Berlin, Germany: Springer.
- White-Schwoch T, Woodruff Carr K, Thompson EC, Anderson S, Nicol T, Bradlow AR, Zecker SG, Kraus N. 2015b. Auditory processing in noise: A preschool biomarker for literacy. *PLOS Biol*. 13:e1002196.
- Woodruff Carr K, Tierney A, White-Schwoch T, Kraus N. 2015. Intertrial auditory neural stability supports beat synchronization in preschoolers. *Dev Cogn Neurosci*. 17:76–82.
- Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. 1997. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*. 387:176–178.
- Wright BA, Zecker SG. 2004. Learning problems, delayed development, and puberty. *Proc Natl Acad Sci USA*. 101:9942–9946.
- Xie R, Gittelman JX, Pollak GD. 2007. Rethinking tuning: in vivo whole-cell recordings of the inferior colliculus in awake bats. *J Neurosci*. 27:9469–9481.
- Zeng F-G, Oba S, Garde S, Sininger Y, Starr A. 1999. Temporal and speech processing deficits in auditory neuropathy. *Neuroreport*. 10:3429–3435.
- Ziegler JC, Pech-Georgel C, George F, Alario F-X, Lorenzi C. 2005. Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci*. 102:14110–14115.