Binaural unmasking with temporal envelope and fine structure in listeners with cochlear implants[a)]

Ann E. Todd[b)]
Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705, USA

Matthew J. Goupell
Department of Hearing and Speech Sciences, University of Maryland at College Park, College Park, Maryland 20742, USA

Ruth Y. Litovsky
Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705, USA

For normal-hearing (NH) listeners, interaural information in both temporal envelope and temporal fine structure contribute to binaural unmasking of target signals in background noise; however, in many conditions low-frequency interaural information in temporal fine structure produces greater binaural unmasking. For bilateral cochlear-implant (CI) listeners, interaural information in temporal envelope contributes to binaural unmasking; however, the effect of encoding temporal fine structure information in electrical pulse timing (PT) is not fully understood. In this study, diotic and dichotic signal detection thresholds were measured in CI listeners using bilaterally synchronized single-electrode stimulation for conditions in which the temporal envelope was presented without temporal fine structure encoded (constant-rate pulses) or with temporal fine structure encoded (pulses timed to peaks of the temporal fine structure). CI listeners showed greater binaural unmasking at 125 pps with temporal fine structure encoded than without. There was no significant effect of encoding temporal fine structure at 250 pps. A similar pattern of performance was shown by NH listeners presented with acoustic pulse trains designed to simulate CI stimulation. The results suggest a trade-off across low rates between interaural information obtained from temporal envelope and that obtained from temporal fine structure encoded in PT. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5102158

[MD]   Pages: 2982–2993

I. INTRODUCTION

In normal-hearing (NH) listeners, signal detection and speech understanding in noise improve when the target signal and noise have different interaural time differences (ITDs) or interaural level differences (ILDs) compared to when target and noise have the same interaural parameters (Schubert, 1956; Egan, 1965; Carhart *et al.*, 1967). This phenomenon, referred to as binaural unmasking, has been found to occur when there is a reduction in the correlation of the sounds at the two ears. Reduced interaural correlation or "decorrelation" occurs when sounds with different interaural parameters overlap in time and frequency (Durlach *et al.*, 1986; Bernstein and Trahiotis, 1992). In these situations, interaural decorrelation occurs in both the temporal fine structure and temporal envelope of the sounds (Bernstein, 1991; van de Par and Kohlrausch, 1995). At frequencies above approximately 1400 Hz, NH listeners are sensitive to interaural differences in temporal envelope but not temporal fine structure (Klumpp and Eady, 1956; Henning,

1974; Brughera *et al.*, 2013). At lower frequencies, NH listeners are sensitive to interaural differences in both temporal fine structure and temporal envelope, but interaural differences in temporal fine structure appear to dominate perception (Bernstein and Trahiotis, 1985; van der Heijden and Joris, 2010). Thus, findings that NH listeners show greater binaural unmasking at frequencies below approximately 1400 Hz suggest that access to interaural decorrelation caused by temporal fine structure results in greater binaural unmasking than access to interaural decorrelation caused by temporal envelope alone (Schubert and Schultz, 1962; Durlach, 1964; Eddins and Barber, 1998).

Listeners with bilateral cochlear implants (CIs) have shown binaural unmasking for tone detection when research processors have been used to synchronize the timing of stimuli bilaterally and a limited number of electrodes have been stimulated (e.g., Long *et al.*, 2006; Lu *et al.*, 2011; Goupell and Litovsky, 2015). In these studies, auditory information was presented in the electrical temporal envelope, indicating that bilateral CI listeners can use interaural decorrelation resulting from the temporal envelopes for binaural unmasking. However, in these studies, no information was conveyed in the electrical temporal fine structure (i.e., the timing of the electrical pulses) because the carrier pulses were constant rate. In the present study, we investigated the effect of encoding acoustic temporal fine structure in electrical pulse

timing (PT) on binaural unmasking. Ideally, encoding temporal fine structure information for bilateral CI listeners could improve their binaural unmasking as it does for NH listeners.

Because interaural decorrelation caused by temporal fine structure can be thought of as time-varying ITDs, the highest pulse rate at which listeners can detect interaural decorrelation caused by PT may be similar to the highest pulse rate at which they can discriminate static ITDs in PT. Studies examining static ITD discrimination with constant-rate pulse trains in bilateral CI listeners have shown that good discrimination performance is limited to lower pulse rates in most cases. The highest pulse rate at which listeners are able to discriminate static ITDs varies with listener and stimulation site, but it is typically less than 300 pulses per second (pps; van Hoesel and Tyler, 2003; Majdak et al., 2006; van Hoesel et al., 2009; Ihlefeld et al., 2015). This rate is well below the 1400-Hz limit found with acoustic temporal fine structure for young NH listeners (Brughera et al., 2013). Thus, encoding temporal fine structure at higher pulse rates is likely to be ineffective at enhancing binaural unmasking, and the use of lower pulse rates may be necessary to make interaural information in PT available.

Although the use of lower pulse rates may render interaural decorrelation caused by PT more accessible, a negative consequence might be that temporal envelope information is rendered less accessible. Goupell and Litovsky (2015) found that listeners with CIs showed poorer interaural decorrelation discrimination at a rate of 100 pps compared to 1000 pps when information was presented solely in the electrical temporal envelope. Poorer performance with lower pulse rates may be due to lower pulse rates providing a sparser representation of temporal envelope information. Lower pulse rates also result in smaller dynamic ranges and changes to loudness growth functions (Fu, 2005; Galvin and Fu, 2009), which might impact listeners' performance. Additionally, poorer performance with lower pulse rates may be due to the irrelevant interaural information in PT (e.g., 0-μs ITD) when constant-rate pulses are used. This irrelevant interaural information in PT is more salient at lower rates and could potentially interfere with listeners' processing of the interaural temporal envelope information (see Bernstein and Trahiotis, 1995).

The purpose of this study was to examine the effect of encoding temporal fine structure in electrical PT (i.e., encoding time-varying ITDs) on binaural unmasking for signal detection. We hypothesized that a trade-off exists between interaural information received from PT and temporal envelope as a function of pulse rate. It was expected that temporal fine structure information presented in PT would produce greater binaural unmasking at rates lower than about 300 pps because of ITD rate limitations (van Hoesel et al., 2009), but that temporal envelope information would produce greater binaural unmasking at higher rates (Goupell and Litovsky, 2015). Performance of NH listeners with acoustic pulsatile stimuli was also examined as a simulation of single-electrode CI stimulation.

II. EXPERIMENT 1: CI PARTICIPANTS

In experiment 1, the effect of presenting temporal fine structure information in PT on binaural unmasking was examined in CI users using low pulse rates (125 and 250 pps). Two main conditions were compared, referred to as the pulse amplitude (PA) and PA + PT conditions. In the PA condition, temporal envelope information was presented in the amplitudes of constant-rate pulses (i.e., in the electrical temporal envelope). In the PA + PT condition, temporal envelope information was presented in the PAs and temporal fine structure information was presented in the PT. Rates up to 250 pps were examined because preliminary data from four CI participants (including participants IAJ, IBN, and ICB; see Table I) using higher pulse rates at 500 and 750 pps suggested no benefit from encoding temporal fine structure in PT on binaural unmasking.

A. Participants

Seven adults with bilateral CIs participated in experiment 1. Table I shows the characteristics of the participants with CIs. All participants had been tested previously and had ITD just noticeable differences (JNDs) that were <400 μs when presented with 100-pps constant-amplitude pulse trains at the single electrode pair listed in Table I. The ITD JNDs in Table I are the ITD of the stimulus (at the JND) multiplied by two, because the ITD discrimination task involved two intervals that were equal in magnitude but opposite in direction (left-right or right-left). Four participants had adult onset of hearing loss. The other three participants (ICB, IAJ, and IBN) had an earlier onset of hearing loss. Specifically, participant ICB had onset of hearing loss at 9 years of age, and participants IAJ and IBN reported being fit with hearing aids at 5 and 4 years of age, respectively. All participants had at least two years of experience using bilateral CIs at the time of testing.

B. Stimuli

Electrical stimuli were presented via the Nucleus Implant Communicator and L34 processors (NIC2 software; Cochlear Ltd., Sydney, Australia), which allowed control of the interaural timing of pulses. All electrical stimuli were presented using one active electrode in each ear, in monopolar 1 + 2 mode. Electrical pulse trains consisted of biphasic cathodic-first pulses with 25-μs phase durations and 8-μs interphase gaps. Stimuli used for calculating threshold and maximum comfort levels were 400-ms constant-amplitude electrical pulse trains presented at a constant rate, which corresponded to the average pulse rate of the stimuli to be used with that loudness map.

Electrical pulse train stimuli used for examining diotic and dichotic signal detection thresholds were based on unprocessed digital waveforms generated at a sampling frequency of 14 286 Hz. Unprocessed waveforms consisted of 400-ms Gaussian noise with a bandwidth of 50 Hz generated in the frequency domain. Center frequencies of the noise were 125, 250, or 1000 Hz. A 300-ms tone generated in the time domain at the center frequency of the noise served as the target. The onset of the tone began 50 ms after the onset of the noise. A Hanning window with 50-ms onset and offset ramps was applied separately to the noise and the tone. The noise was interaurally correlated or in-phase (No), and the tone was interaurally in-phase (So) or interaurally out-of-phase (Sπ). The

TABLE I. Participant characteristics of experiment 1.

Participant	Age (yr)	CI experience (yr)	Bilateral experience (yr)	Left internal device[a]	Right internal device[a]	Electrode pair (left, right)	ITD just noticeable difference (JND) (μs) at 100 pps
IAJ	69	17	11	CI24M	CI24R(CS)	L16, R19	352
IBF	62	8	6	Freedom Contour Advance	Freedom Contour Advance	L4, R6	46
IBK	73	10	4	CI24R (CS)	Freedom Contour Advance	L6, R6	115
IBN	67	13	4	Freedom Contour Advance	CI24R(CS)	L12, R16	382
IBY	50	6	2	Freedom Contour Advance	CI512	L4, R7	193
ICB	64	12	9	Freedom Contour Advance	CI24R(CA)	L4, R4	398
ICI	55	5	4	Freedom Contour Advance	Freedom Contour Advance	L4, R6	282

[a]Devices were produced by Cochlear Ltd., Sydney, Australia.

in-phase noise combined with the in-phase and out-of-phase tones are referred to as NoSo and NoSπ, respectively. The stimuli were pre-generated to reduce testing time. Twenty-five different noise tokens were created and used in each of the stimulus conditions for each signal-to-noise ratio (SNR). The SNR of the tone in noise varied between +20 and –30 dB in 2-dB steps. The temporal envelope and temporal fine structure of the stimuli were calculated using the Hilbert transform. The temporal envelopes of the combined tone and noise signal were normalized such that the average amplitude was 0.4 units on an arbitrary scale, which resulted in peak amplitudes of approximately 1. For the PA condition, the temporal envelopes were resampled at constant rates corresponding to the center frequency of the noise band. For the PA + PT condition, temporal envelopes were resampled at times corresponding to zero crossings in the Hilbert phase during times at which the digital waveform had a positive local maximum (i.e., the peaks in the temporal fine structure). Temporal envelopes were compressed between each CI participant's threshold and maximum comfort levels using the function described by Long et al. (2006). Levels 30 dB below an amplitude of 1 unit were set to zero, as was done in Long et al. (2006), which corresponds to the dropping of low-amplitude information similar to what occurs in a Cochlear-brand sound processing strategy. In the PA condition, the compressed envelope was used to amplitude modulate an electrical pulse train running at a rate matching the envelope sampling rate. Pulses in the PA condition had the same periodic timing in left and right channels (i.e., the pulses had 0-μs ITD). For the PA + PT condition, the compressed envelope was used to amplitude modulate a train of aperiodic pulses timed to the zero crossings in the Hilbert phase for the positive portion of the stimulus waveform. Figure 1 shows an example of an NoSπ stimulus in the PA and PA + PT conditions. Pulse trains in the PA + PT condition had an average rate that was approximately the center frequency of the unprocessed waveforms. Pulse rates in the PA + PT condition will be referred to by the rate matching the center frequency (e.g., 125 pps). Both the PA and PA + PT conditions were presented at the rates of 125 and 250 pps. Only the PA condition was examined at 1000 pps as a reference of performance with only temporal envelope information at a rate similar to what is used with many clinical CI processors.

C. Procedure

For each participant, one pair of interaurally pitch-matched electrodes was chosen from a set of three pairs that

had been identified in a previous experiment. For that previous experiment, two steps had been taken, using the interaural pitch-matching method described by Kan et al. (2013). First, all even-numbered electrodes on the left and right sides were stimulated individually in random order with ten repetitions per electrode, and the participant rated the pitch of the electrode on an arbitrary scale. Second, an interaural pitch comparison task was conducted in which the participant judged the pitch of each of six electrodes on the right relative to the pitch of one electrode on the left with 20 repetitions per comparison. The electrode pair chosen for the current study was the one out of the three pitch-matched pairs that showed the best ITD JNDs at 100 pps.

For each electrode of the pair, thresholds and maximum comfort levels were found using constant-amplitude pulse trains at 125, 250, and 1000 pps. Thresholds were the lowest levels that produced reliable detection responses when levels were increased from below. Maximum comfort levels were defined for the participants as the highest level still in the comfortable range. The thresholds and maximum comfort levels were used in the compression functions of the stimuli at the corresponding rates.

Trials for measuring NoSo and NoSπ signal detection thresholds consisted of a three-interval two-alternative forced-choice task. The inter-stimulus interval was 300 ms. On any trial, either the second or third interval contained the target signal embedded in the noise (NoSo or NoSπ), while the other intervals consisted of only diotic noise. For each trial, three noise tokens were randomly selected for presentation without replacement. The SNR was varied using two-down one-up
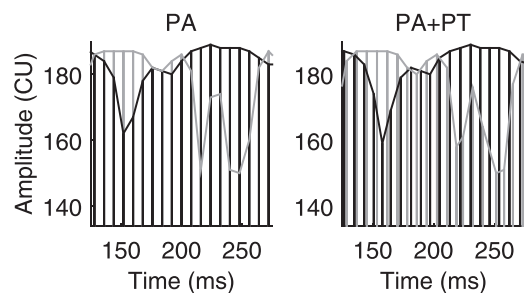

FIG. 1. Representations of an NoSπ electrical stimulus at 125 pps in the PA (left) and PA + PT (right) conditions at 0-dB SNR. Left and right channels are shown in gray and black, respectively. The temporal envelope of each channel is traced with a line for clarity. The units of the y axis are clinical current units (CU). The stimulus was compressed between 134 and 190 CU (arbitrarily chosen) for the figure. Only the middle 150 ms of the stimulus is shown.

adaptive tracks. Tracks started at +20-dB SNR and varied in steps of 8 dB until the second reversal, 4 dB until the fourth reversal, and then 2 dB. Tracks ended after ten reversals in total. NoSo and NoSπ signal detection thresholds were calculated as the average SNR of the last six reversals. Three to five tracks were collected per condition. Adaptive tracks were presented in blocks of NoSo and NoSπ, which were alternated. Each block consisted of one track per condition (within the specified phase condition) with conditions presented in a newly randomized order across blocks. Correct answer feedback was always provided to support participants in attending to the correct cues. Participants were told to listen for the interval that was different, which would likely sound smoother for the NoSo conditions and have an additional perceived cue of diffuseness or movement within the head for the NoSπ conditions.

Prior to testing, participants were familiarized with the PA stimuli at 1000 pps and the PA + PT stimuli at 125 pps using the three-interval two-alternative forced-choice task. Initially for the familiarization, NoSo stimuli were presented at +12-dB SNR and NoSπ stimuli at 0-dB SNR. Subsequently for the familiarization, participants completed one or two adaptive tracks for each of the two conditions for both the NoSo and NoSπ conditions.

Thresholds were fit to linear mixed-effects models, which had random intercepts for participants. Likelihood ratio tests (reported with the symbol X^2) were performed by comparing models with and without the variable of interest. The p-values were calculated by comparing the test statistics to distributions calculated using parametric bootstrap analyses (1000 iterations) in which data were simulated under the model without the variable of interest (the null hypothesis). This approach was used to provide a better estimate of the null distribution than provided by the X^2 distribution. The within-participant factors of phase (NoSo and NoSπ), pulse rate, and stimulus type (PA and PA + PT) were examined for significance.

Post hoc pairwise t-tests (two-tailed) were carried out following significant likelihood ratio tests. *Post hoc* tests included a Holm correction to the p-values based on the number of tests carried out for each factor at a particular level of the other factors. Only significant *post hoc* tests are reported. For three adaptive tracks (two for participant IBF and one for participant IBN) in the PA condition at 125 pps, participants had enough incorrect responses at +20-dB SNR (the maximum SNR) that a threshold could not be calculated accurately. In these cases, thresholds were recorded as +24 dB. The maximum SNR tested was chosen to be +20 dB because stimulus analyses have suggested that

performance was unlikely to improve with SNRs higher than +12 dB (Goupell and Litovsky, 2015).

D. Results and discussion

Figure 2 shows NoSo and NoSπ thresholds for the PA and PA + PT conditions. Panels from left to right show thresholds at 125, 250, and 1000 pps. First, thresholds at 125 and 250 pps are described because both the PA and PA + PT conditions were tested at these rates. The mean NoSo threshold was 3.4 dB [standard deviation (SD) = 1.8 dB] and the mean NoSπ threshold was −5.4 dB (SD = 5.6 dB). The effect of phase was significant ($X_1^2 = 101.88$, $p < 0.001$). The binaural masking level difference was 8.8 dB.

There was a greater difference between NoSo and NoSπ thresholds in the PA + PT condition compared to the PA condition. In other words, there was a larger binaural masking level difference in the PA + PT condition. The effect of type ($X_1^2 = 8.25$, $p = 0.002$) and the phase × type interaction ($X_1^2 = 5.99$, $p = 0.014$) were significant. *Post hoc* tests showed that NoSπ thresholds were lower than NoSo thresholds for both the PA ($t_{221} = 6.48$, $p < 0.0001$) and the PA + PT conditions ($t_{221} = 9.85$, $p < 0.0001$), indicating that participants showed binaural unmasking for both the PA and PA + PT conditions. Additionally, NoSπ PA + PT thresholds were lower than NoSπ PA thresholds ($t_{221} = 3.78$, $p = 0.00039$), suggesting that the participants were sensitive to the interaural differences caused by temporal fine structure in the PA + PT condition.

There was a greater difference between PA and PA + PT thresholds at 125 pps than at 250 pps. The effect of rate was not significant ($X_1^2 = 2.26$, $p = 0.14$), but the rate × type interaction was significant ($X_1^2 = 7.38$, $p = 0.006$). *Post hoc* tests showed PA + PT thresholds were lower than PA thresholds at 125 pps ($t_{221} = 3.91$, $p = 0.00024$), indicating that participants were sensitive to the interaural information caused by temporal fine structure at 125 pps. Additionally, thresholds at 250 pps were lower than at 125 pps in the PA condition ($t_{221} = 2.95$, $p = 0.0068$). This suggests that participants could not use the temporal envelope information as effectively at 125 pps as at 250 pps in the PA condition. From Fig. 2, these findings appear to be due to the relatively higher NoSπ PA thresholds at 125 pps. The phase × rate interaction ($X_1^2 = 0.23$, $p = 0.63$) and the phase × rate × type interaction ($X_1^2 = 1.55$, $p = 0.28$) were not significant.

Despite listeners showing sensitivity to temporal fine structure information, there was no evidence of better performance with the PA + PT stimuli at low rates compared to the
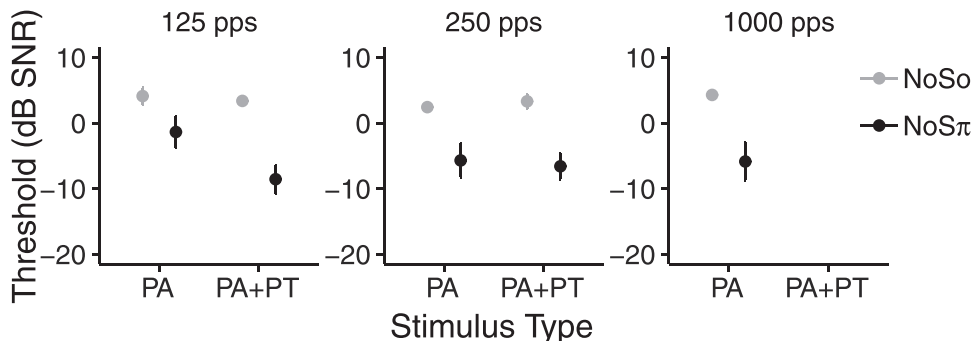

FIG. 2. NoSo (gray) and NoSπ (black) signal detection thresholds (dB SNR) in the PA and PA + PT conditions averaged across CI participants. Left, middle, and right panels show thresholds for 125 pps, 250 pps, and 1000 pps, respectively. Data were not collected for the 1000-pps PA + PT condition. Error bars show ±1 standard error of the mean.

J. Acoust. Soc. Am. **145** (5), May 2019

Todd *et al.*   2985

PA stimuli at higher rates. To examine this, the PA + PT thresholds (at 125 and 250 pps) were combined with the PA thresholds at 1000 pps. For this data set, the effect of rate (125, 250, 1000 pps) was not significant ($X_2^2 = 2.56$, $p = 0.29$), and the phase × rate interaction was not significant ($X_2^2 = 0.42$, $p = 0.81$).

We analyzed the PA data alone (125, 250, and 1000 pps) to examine the effect of rate up to 1000 pps. The effect of rate was significant ($X_2^2 = 8.97$, $p = 0.011$). *Post hoc* tests showed that thresholds were lower at 250 compared to 125 pps ($t_{164} = 2.79$, $p = 0.018$) and at 1000 pps compared to 125 pps ($t_{164} = 2.33$, $p = 0.042$). The phase × rate interaction was not significant ($X_2^2 = 4.1675$, $p = 0.12$).

Overall, the difference between the PA and PA + PT conditions seems to be due to higher PA thresholds compared to PA + PT thresholds in the NoSπ 125-pps condition (Fig. 2). This would suggest that at the lowest rate (125 pps), listeners had difficulty using interaural temporal envelope information with constant-rate pulses. This may have been due to poorer temporal envelope representation at the lowest rate, which presumably was compensated for in the PA + PT condition by the interaural temporal fine structure information in the PT. Another explanation is that the higher thresholds in the NoSπ PA condition at 125 pps may have been due to a type of binaural interference from the 0-μs ITD with the constant-rate pulses, which would have been more salient at the lowest rate. At the lowest rate, listeners had better NoSπ signal detection thresholds when temporal fine structure information was encoded in PT. However, the encoding of temporal fine structure in PT did not improve performance beyond that which was obtained at higher pulse rates.

III. EXPERIMENT 2: NH PARTICIPANTS

Experiment 2 examined whether the pattern of performance of the CI users in experiment 1 is similar to the pattern of performance of NH listeners listening to acoustic pulsatile stimuli. Acoustic pulses were used to simulate electrical pulses. As with the CI users, we examined NH-listener performance with and without temporal fine structure information encoded in the pulses. The terms PA and PA + PT are used for the NH listeners as they were for the CI users. That is, in the PA condition, temporal envelope information was encoded in the amplitudes of constant-rate acoustic pulses, and in the PA + PT condition, temporal envelope information was encoded in the PAs, and temporal fine structure information was encoded the PT. In addition to examining performance in the PA and PA + PT conditions as was done in experiment 1, performance in a PT condition was examined. The PT condition was similar to the PA + PT condition except that temporal envelope information encoded in the PAs was made consistent across the ears. This was done to evaluate listeners' use of interaural information in the PT in the absence of useful interaural information in the PAs. The effect of noise bandwidth was also examined at the highest pulse rate (500 pps).

A. Participants

Eight NH adults participated in experiment 2. The NH participants were between 19 and 32 years of age (mean = 23.3 years, SD = 5 years). All NH participants could detect a 25 dB hearing level (HL) tone at octave intervals from 500 to 8000 Hz in each ear.

B. Stimuli and procedure

Acoustic stimuli were presented to NH participants using a Tucker Davis System 3 (Alachua, FL) and ER-2 insert earphones (Etymotic Research, Inc., Elk Grove Village, IL). NH participants were tested in a double-walled sound attenuating chamber (IAC Acoustics, North Aurora, IL). Trains of Gaussian-shaped pulses applied to a tonal carrier were created to simulate the electrical pulse trains presented to the CI participants (Lu *et al.*, 2007; Goupell *et al.*, 2010; Goupell, 2012). Figure 3 shows the waveform and spectrum of an example stimulus in the PA condition at 125 pps. Pulse trains had a center frequency of 9.2 kHz. A high center frequency was used to reduce the effect of auditory filtering on the modulation depth of the pulses. The equivalent rectangular bandwidth of the pulses was 2500 Hz, which was used to simulate spread of excitation with single-electrode stimulation. The equivalent rectangular duration of the pulses was 0.4 ms. The pulses were temporally brief and therefore had a 100% modulation depth for all pulse rates, even at 500 pps. The pulse train stimuli were based on unprocessed waveforms that were like those created for the CI participants. Unprocessed waveforms and pulse trains were generated at a sampling rate of 50 kHz. Noise samples were created with 125 -, 250 -, and 500-Hz center frequencies and 50-Hz bandwidths. In addition to the 50-Hz bandwidth noise, noise samples were created with a 125-Hz bandwidth at the 500-Hz center frequency. The amplitude and timing of pulses were determined in the manner used for the stimuli of the CI participants except that envelopes were left uncompressed. Pulse trains were used to amplitude modulate a tone at 9.2 kHz. Stimuli were presented at a 65 dB sound pressure level (SPL). Interaurally uncorrelated pink noise was played continuously from 0 to 20 kHz at 60 dB SPL (spectrum level at 500 Hz = 39 dB SPL) to mask any potential combination tones, which could provide the NH listeners with an advantage over the CI users.
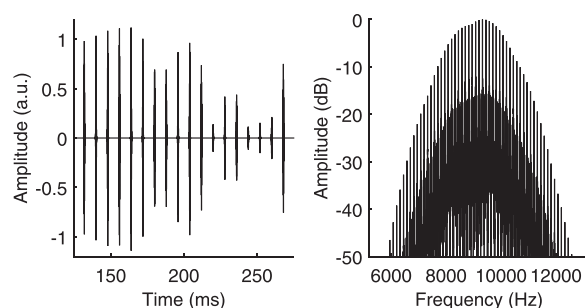
FIG. 3. The waveform (left) and spectrum (right) of an acoustic stimulus with a 125-pps pulse rate in the PA condition. Only the left channel is shown. The waveform shows the middle 150 ms of the stimulus.
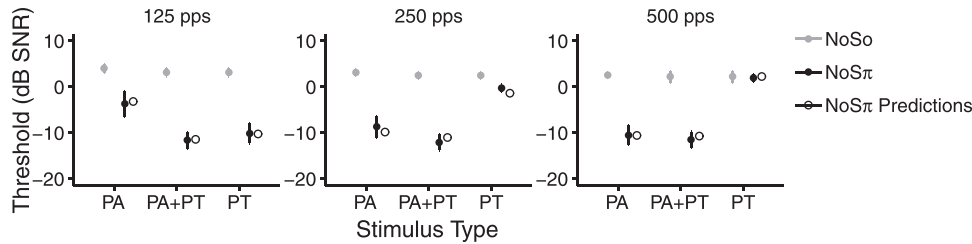
FIG. 4. NoSo (gray) and NoSπ (black) signal detection thresholds (dB SNR) in the PA, PA + PT, and PT conditions with the 50-Hz bandwidth stimuli, averaged across NH participants. Left, middle, and right panels show thresholds for 125 pps, 250 pps, and 500 pps, respectively. NoSo PA + PT and NoSo PT thresholds are the same data. Error bars show ±1 standard error of the mean. Unfilled black circles show model predictions of NoSπ thresholds (see the Appendix).

Pulse rates (average pulse rates for the PA + PT and PT conditions) above 500 pps were not used for the NH participants because higher pulse rates would likely have resulted in resolvable spectral peaks, which may have provided unwanted spectral cues for detecting the NoSo target (Carlyon and Deeks, 2002). For the PT NoSπ condition, the amplitudes of the pulses of both the left and right channels were determined solely by the Hilbert envelope of the left channel of the unprocessed waveforms. Thus, the NoSπ PT stimuli differed from the NoSπ PA + PT stimuli in that the stimuli lacked interaural differences in the temporal envelope of the PAs. The PT condition was not tested in the NoSo condition because it was the same as the PA + PT condition for the NoSo stimuli. The experimental procedure was like that of experiment 1. The PT NoSπ condition was presented in the same NoSπ blocks as the PA and PA + PT conditions. Statistical testing was similar to that of experiment 1. Statistical analysis for the PT condition was done separately because the NoSo PT data were the same as the NoSo PA + PT data. As in experiment 1, only significant *post hoc* tests are reported.

C. Results and discussion

Figure 4 shows NH thresholds in the NoSo and NoSπ conditions for the PA, PA + PT, and PT conditions with the 50-Hz bandwidth stimuli. Panels from left to right show thresholds at 125 pps, 250 pps, and 500 pps. Figure 4 also shows predicted NoSπ thresholds based on numerical modeling of the NH stimuli, described in the Appendix.

The results from the 50-Hz bandwidth PA and PA + PT condition are described first. Because the phase × rate × type interaction was significant ($X_2^2 = 7.75$, $p = 0.018$), *post hoc* pairwise contrasts were examined for factors at each level of the other factors. *Post hoc* tests showed that NoSπ thresholds were lower than NoSo thresholds at each combination of type and rate ($t_{374} > 6.0$, $p < 0.0001$), indicating that participants showed binaural unmasking in all six conditions. NoSπ PA + PT thresholds were lower than NoSπ PA thresholds at 125 pps ($t_{374} = 6.37$, $p < 0.0001$) and 250 pps ($t_{374} = 3.004$, $p = 0.014$). This suggests that participants were sensitive to the interaural differences caused by the PT at 125 and 250 pps in the PA + PT condition. Furthermore, thresholds in the NoSπ PA condition were lower at 250 pps compared to 125 pps ($t_{374} = 3.63$, $p = 0.0035$) and at 500 pps compared to 125 pps ($t_{374} = 5.35$, $p < 0.0001$), suggesting that participants had more difficulty using interaural information in the PAs at 125 pps than at 250 or 500 pps.

Despite sensitivity to interaural differences in PT, there was no evidence of better performance with the PA + PT stimuli at lower rates than with the PA stimuli at the highest rate. To examine this, PA + PT thresholds at 125 and 250 pps were combined with the PA thresholds at 500 pps. The effect of rate (125, 250, and 500 pps) was not significant ($X_2^2 = 0.48$, $p = 0.776$) and neither was the phase × rate interaction ($X_2^2 = 3.58$, $p = 0.173$).

The results from the 50-Hz bandwidth PT condition are described next. As can be seen in Fig. 4, the pattern of the NoSπ PT thresholds across rates was the opposite of that of the NoSπ PA thresholds. That is, NoSπ PT thresholds were lowest at the lowest rate. The effect of phase ($X_1^2 = 45.07$, $p < 0.001$), the effect of rate ($X_2^2 = 29.18$, $p = 0.001$), and the phase × rate interaction were significant ($X_2^2 = 58.04$, $p < 0.001$). *Post hoc* tests showed that NoSπ PT thresholds were lower than NoSo PT thresholds at 125 pps ($t_{166} = 11.37$, $p < 0.0001$) and 250 pps ($t_{166} = 2.79$, $p = 0.012$), indicating that participants could use PT for binaural unmasking at 125 and 250 pps. Furthermore, NoSπ PT thresholds were lower at 125 pps compared to 250 pps ($t_{167} = -8.089$, $p < 0.0001$) and 125 pps compared to 500 pps ($t_{167} = -9.71$, $p < 0.0001$). These results suggest that participants were most sensitive to interaural differences in PT at 125 pps compared to 250 pps or 500 pps.

Figure 5 shows thresholds for the two noise-bandwidth conditions at 500 pps. There was no evidence of a difference between the PA and PA + PT thresholds with either bandwidth. The bandwidth × type interaction ($X_1^2 = 0.3$, $p = 0.599$) and bandwidth × phase × type interaction ($X_1^2 = 0.40$, $p = 0.521$) were not significant. This fails to provide evidence that noise bandwidth affects the usefulness of interaural information in PT. Furthermore, in the PT condition at 500 pps,
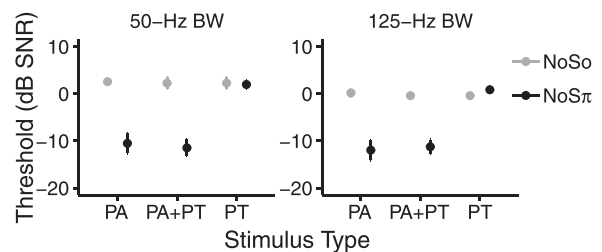

FIG. 5. NoSo (gray) and NoSπ (black) signal detection thresholds (dB SNR) in the PA, PA + PT, and PT conditions at 500 pps averaged across NH participants. Left and right panels show thresholds for the 50-Hz and 125-Hz noise-bandwidth (BW) conditions, respectively. NoSo PA + PT and NoSo PT thresholds are the same data. The 50-Hz BW data are the same as the 500-pps data in Fig. 4. Error bars show ±1 standard error of the mean.

there was no evidence of binaural unmasking at either bandwidth. The effect of phase ($X_1^2 = 0.57$, $p = 0.473$) and the bandwidth $\times$ phase interaction ($X_1^2 = 1.69$, $p = 0.228$) were not significant. The absence of an effect of phase in the PT condition suggests that listeners were not able to use the interaural information in PT with either bandwidth at 500 pps. There was a significant and very small effect of bandwidth ($X_1^2 = 6.44$, $p = 0.016$) with PA and PA + PT thresholds being lower for the wider bandwidth. The bandwidth $\times$ phase interaction ($X_1^2 = 1.89$, $p = 0.17$) was not significant.

We predicted the NH listeners' performance with the 50-Hz NoSπ stimuli using a model intermediate to the normalized correlation and normalized covariance models (Bernstein and Trahiotis, 1996b; see the Appendix). NH participants' NoSπ thresholds could be predicted relatively well from this model (Fig. 4), indicating that even though the stimuli of the present study are different from what is typically used for binaural unmasking experiments, the performance of the listeners can be understood in a way that is generally consistent with current understanding of NoSπ detection (Bernstein and Trahiotis, 2017).

Overall, the results show that PA thresholds were higher than PA + PT thresholds in the NoSπ condition at 125 and 250 pps (Fig. 4). The results suggest that at lower rates, the NH participants could use the interaural information in PT but had difficulty using interaural information in the PAs. These results are similar to those of the CI listeners in experiment 1, except that a significant difference was found between the NoSπ PA + PT and NoSπ PA conditions at 250 pps for the NH participants but not the CI participants. However, when the groups (CI and NH) were compared directly, there was no evidence of a larger difference between the PA and PA + PT conditions in the NH listeners than the CI listeners at 250 pps. Specifically, the group $\times$ type interaction ($X_1^2 = 3.28$, $p = 0.073$) and the group $\times$ phase $\times$ type interaction ($X_1^2 = 0.27$, $p = 0.62$) were not significant at 250 pps. Like with the CI listeners, the encoding of temporal fine structure in PT failed to improve NH listeners' performance beyond that which was achieved with constant-rate pulses at the highest pulse rate tested. In addition, at the highest rate tested, information in PT was not found to be more useful at a wider noise bandwidth.

IV. GENERAL DISCUSSION

For NH listeners, previous studies have shown that binaural unmasking is greater at frequencies at which listeners are sensitive to interaural differences in temporal fine structure (Eddins and Barber, 1998). For this reason, it was of interest to investigate whether presenting temporal fine structure information in electrical PT can improve binaural unmasking for listeners with CIs. Binaural unmasking was examined by comparing diotic and dichotic signal detection thresholds in conditions in which only temporal envelope information was encoded and conditions in which both temporal envelope and temporal fine structure information were encoded. Stimuli were presented using bilaterally synchronized single-electrode stimulation. Binaural unmasking was also examined in NH listeners using Gaussian-enveloped tones to simulate single-channel CI stimulation. It was

expected that lower pulse rates would result in better sensitivity to interaural information in PT but poorer sensitivity to interaural information in PAs.

For diotic signal detection, no effect of encoding temporal fine structure information in PT was found for the CI or NH listeners. There was the possibility that encoding temporal fine structure information would result in better diotic signal detection thresholds because the presence of the target tone in the noise increased the periodicity of the stimulus, especially at higher SNRs, which could have created a cue for detecting the target. We expected this cue to be most useful to listeners at the rate of 125 pps at which the noise bandwidth was the largest proportion of the average rate (Chen and Zeng, 2004). However, this cue may have been too subtle compared to the cue from the temporal envelope information in the PAs.

CI listeners showed better dichotic signal detection thresholds than diotic signal detection thresholds regardless of whether temporal fine structure information was encoded or not. This is consistent with previous findings that CI listeners show binaural unmasking when information is encoded as temporal envelope modulations of constant-rate pulses (e.g., Long et al., 2006; Goupell and Litovsky, 2015). However, CI listeners showed better dichotic signal detection thresholds with temporal fine structure information encoded in the PT than with constant pulse rates at 125 pps but not at 250 pps (Fig. 2). Dichotic signal detection thresholds for CI users were best at the rate of 125 pps when temporal fine structure was encoded; however, encoding temporal fine structure at 125 pps did not significantly improve dichotic thresholds beyond dichotic thresholds with constant pulse rates at higher rates. Rather, the benefit of encoding temporal fine structure at 125 pps made up for the higher dichotic thresholds with constant-rate pulses at 125 pps. In a similar pattern, NH listeners showed better dichotic signal detection thresholds with temporal fine structure information encoded in PT than with constant-rate pulses at 125 pps and 250 pps but not at 500 pps (Fig. 4). Although CI listeners did not show a significant effect of temporal fine structure at 250 pps, the effect at 250 pps did not differ significantly between NH and CI listeners. Notably, the trend of a higher upper limit of sensitivity to interaural information in PT found with the NH listeners compared to the CI listeners in the present study is consistent with the results of previous studies, which have examined CI and NH listeners' sensitivity to time-varying ITDs in pulse trains, generated by using slightly different pulse rates between ears. Using these stimuli, CI listeners have shown sensitivity to time-varying ITDs at 100 pps and 200 pps but not 300 pps (van Hoesel, 2007; Carlyon et al., 2008), while NH listeners have shown sensitivity to time-varying ITDs in acoustic pulse trains at 300 pps (Carlyon et al., 2008). Similar to the CI users in the present study, temporal fine structure encoded in PT did not improve dichotic thresholds of NH listeners beyond dichotic thresholds with constant pulse rates at higher rates but rather made up for higher thresholds at lower constant pulse rates. The fact that CI and NH listeners only benefited from the encoding of temporal fine structure information at low pulse rates is consistent with previous findings that the ITD

discrimination of CI users (and NH listeners with amplitude-modulated stimuli) is limited to lower rates (van Hoesel *et al.*, 2009; Bernstein and Trahiotis, 2014; Ihlefeld *et al.*, 2015).

van de Par and Kohlrausch (1997) found that NH listeners can show greater binaural unmasking with transposed stimuli at a 4-kHz center frequency than with tone-in-noise stimuli at 4 kHz. Transposed stimuli are similar to the acoustic PA + PT stimuli in the present study. From the results of van de Par and Kohlrausch (1997), one might expect listeners in the present study to have shown greater binaural unmasking in the PA + PT condition at 125 pps than in the PA condition at 500 pps, since the PA condition is similar to tone-in-noise stimuli at 4 kHz in the sense that only temporal envelope information is available for both conditions. However, the data in Fig. 4 show similar unmasking in the PA + PT condition at 125 pps and in the PA condition at 500 pps. Possibly, the difference between the two studies is a result of better performance with the acoustic pulse train PA stimuli at 500 pps than with tone-in-noise stimuli at the same center frequency.

The results of the present study are also in contrast to those of previous studies in CI users that have examined the effect of presenting temporal fine structure on binaural speech unmasking (van Hoesel *et al.*, 2008; Zirn *et al.*, 2016). These studies did not find an effect of temporal fine structure; however, stimuli were presented using multi-electrode stimulation, which may have reduced the accessibility of the temporal fine structure due to channel interactions (Shannon, 1983). In addition, these studies did not control for the rate of stimulation per electrode across conditions with and without temporal fine structure. Thus, lower rates used to present temporal fine structure were compared to higher constant rates. These issues may have contributed to these studies not finding an effect of temporal fine structure.

While CI and NH participants showed sensitivity to interaural information in PT at lower pulse rates, interaural temporal envelope information encoded in PA was utilized more poorly at lower rates. This finding is consistent with the results of Goupell and Litovsky (2015), which showed CI users have poorer envelope-based interaural correlation discrimination at 100 pps than at 1000 pps. This result may have been because lower pulse rates provided the listeners with a sparser representation of the temporal envelope. If this was the case, we might expect to see the effect of pulse rate decrease with slower temporal envelope modulation rates as occurs with narrower noise bandwidths. Indeed, Goupell (2012) failed to find an effect of pulse rate for dichotic PA thresholds with pulse trains sampling 10-Hz bandwidth noise envelopes in NH listeners. If lower rates resulted in higher dichotic signal detection thresholds due to poorer temporal envelope representation, it is unclear why pulse rate was not found to affect thresholds for diotic signal detection. An alternative explanation for higher dichotic detection thresholds with lower constant pulse rates is that at lower rates, there was an increase in the saliency of the irrelevant interaural information (0-μs ITD) in the PT, which interfered binaurally with listeners' abilities to use interaural temporal envelope information in the PAs (see Bernstein and Trahiotis, 1995). One way to test this would be to degrade the

salience of the interaural information in the PT, perhaps by modifying the pulse shape, to see if it improves dichotic signal detection thresholds with low-rate constant-rate pulses.

In summary, this study showed that for dichotic signal detection, there is a trade-off between interaural information obtained from PT and interaural information obtained from PA across low pulse rates for a binaural unmasking task. No overall benefit was demonstrated from encoding temporal fine structure in PT using low rates compared to encoding just envelope information at higher rates.

APPENDIX: NUMERICAL MODELING

A number of metrics have been used to predict NH listeners' NoSπ thresholds and interaural correlation discrimination thresholds (e.g., Osman, 1971; Gabriel and Colburn, 1981). Two of these metrics are the normalized interaural correlation and the normalized interaural covariance (e.g., Bernstein and Trahiotis, 1996b). These metrics are given by the formula

$$\rho = \frac{\sum x(t)y(t)}{\sqrt{\sum x(t)^2}\sqrt{\sum y(t)^2}}.$$

For the normalized correlation, $x(t)$ and $y(t)$ represent the stimuli in the left and right ears, respectively, and are irrespective of the mean values of the waveforms. For the normalized covariance, $x(t)$ and $y(t)$ represent the deviations of the stimuli in the left and right ears, respectively, from their respective means. Both metrics are predictive of listener performance across a wide range of conditions, especially when physiologically inspired preprocessing is added (Bernstein and Trahiotis, 1996a,b; Bernstein *et al.*, 1999). However, the normalized covariance has been shown to be a poorer predictor of listener performance with temporal envelope when modulation depth is manipulated (Bernstein and Trahiotis, 1996b). That is, for stimuli with low modulation depth, the normalized envelope correlation was better than the normalized envelope covariance in predicting listeners' detection of changes in interaural correlation. Bernstein and Trahiotis (1996b) proposed metrics intermediate to the normalized correlation and covariance, where $x(t)$ and $y(t)$ represent the deviations of the stimuli in the left and right ears, respectively, from a percentage of their respective means. However, support for this type of metric was not found. We refer to these intermediate metrics by the percentage of the mean remaining. For example, when $x(t)$ and $y(t)$ represent deviations from 70% of the mean, this metric is termed the normalized interaural covariance 30% because 30% of the mean remains in the calculation.

We calculated the normalized interaural correlation (100% of the mean remaining), covariance (0% of the mean remaining), and metrics intermediate to the correlation and covariance (from 10% to 90% of the mean remaining in 10% steps) of the NoSπ stimuli presented to the NH listeners to determine which of these metrics best predicted listeners' NoSπ thresholds across the different pulse rates and stimulus types. Calculations were based on the 25 stimulus tokens per SNR (−32- to 0-dB SNR in 2-dB steps) in the 50-Hz bandwidth NoSπ condition used in experiment 2. Metrics for SNRs higher than 0 dB were not calculated because listeners' responses at higher SNRs were assumed to be largely affected by monaural cues. Indeed, the NH average NoSπ thresholds (Fig. 4) were less than zero for each condition except the PT 500-pps condition in which NH listeners did not show binaural unmasking. Prior to calculating the metrics, the stimuli were processed to simulate auditory peripheral processing (Bernstein *et al.*, 1999; Goupell, 2012). This consisted of fourth-order gammatone filtering centered at 9.2 kHz, temporal envelope compression with a power of 0.46, half-wave rectification, and low-pass filtering. The stimuli were first low-pass filtered with a fourth-order Butterworth filter with a 425-Hz cutoff frequency. The stimuli were then low-pass filtered a second time, this time with a second-order Butterworth filter with a cutoff frequency that varied from 50 to 150 Hz in 10-Hz steps across different iterations of each metric (Bernstein and Trahiotis, 2014). Because the low-pass filtering resulted in no remaining acoustic temporal fine structure (i.e., no carrier at 9.2 kHz), the metrics were that of the temporal envelope (i.e., the pulses) and will be referred to in this regard (e.g., the normalized envelope correlation). Each metric was calculated from the middle 200 ms of the 400-ms stimuli to avoid onset and offset ramps. Figure 6 shows examples of the processed PA + PT stimuli at each pulse rate. The predicted NoSπ threshold for any given condition was taken as the SNR corresponding to a particular value of the metric, which was held constant across stimulus conditions within each metric. To find this particular value of the metric, the value of the metric was systematically varied to find the predicted thresholds which resulted in the highest percent variance explained in the observed thresholds. The percent variance explained was calculated using the formula in Bernstein and Trahiotis (1996a),

$$100 \times \left(1 - \left(\sum (O_i - P_i)^2\right) \middle/ \left(\sum (O_i - \bar{O})^2\right)\right),$$

where O_i and P_i refer to the mean observed and predicted thresholds, respectively, for individual conditions, and $\bar{O}$ refers to the grand mean of the observed thresholds across conditions. The predicted threshold in the PT 500-pps condition was set to the observed NoSo threshold for the percent variance calculation because the metric values were consistently high in this condition, which is consistent with absence of binaural unmasking.

The rows of Fig. 7 show the normalized envelope correlation, normalized envelope covariance 30%, and the normalized envelope covariance, respectively, as a function of SNR. The panels from left to right show the PA, PA + PT, and PT conditions. Pulse rate (125, 250, and 500 pps) is represented by the shading of the points. The metric values that resulted in the highest percent variance explained are shown by the dotted horizontal lines. The cutoff frequency of the second-order low-pass filter was 100 Hz in Fig. 7 as this resulted in the overall highest percent variance explained (97.9%; Table II). With each of the three metrics in Fig. 7, NoSπ threshold predictions for the PT stimuli matched the pattern of thresholds of the NH listeners. That is, at higher pulse rates, there were higher predicted thresholds (Fig. 4; observed NoSπ thresholds were −10.1-, −0.2-, and 1.9-dB SNR at 125, 250, and 500 pps, respectively; $p < 0.0001$ for −10.1 vs −0.2 and −10.1 vs 1.9). For the PA stimuli, threshold predictions with the normalized envelope covariance and normalized envelope covariance 30% (but not the normalized envelope correlation) match the pattern of thresholds of the NH listeners. That is, these two metrics predicted lower thresholds with higher pulse rates as was observed (Fig. 4; observed NoSπ thresholds were −3.6-, −8.6-, and −10.5-dB SNR at 125, 250, and 500 pps, respectively; $p = 0.0035$ for −3.6 vs −8.6 and $p < 0.0001$ for −3.6 vs −10.5), but the normalized envelope correlation predicted equivalent thresholds across rates. The failure of the normalized envelope correlation to predict performance across rates in the PA condition seems to be due to the insensitivity of the normalized envelope correlation to the interaurally synchronized inter-pulse intervals, which are more prominent at lower
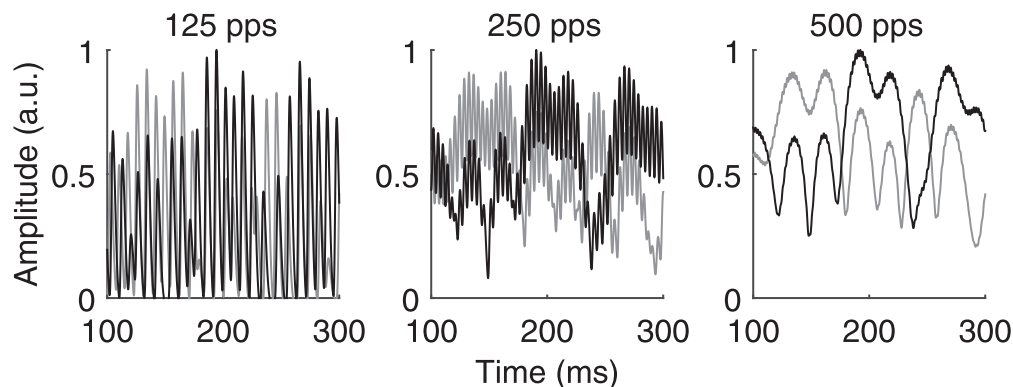

FIG. 6. Representations of an acoustic PA + PT NoSπ stimulus at 125 pps (left), 250 pps (middle), and 500 pps (right) after physiologically inspired preprocessing. Stimuli were at −4-dB SNR. Left and right channels are shown in gray and black, respectively. Preprocessing consisted of fourth-order gammatone filtering centered at 9.2 kHz, temporal envelope compression with a power of 0.46, half-wave rectification, and low-pass filtering with a fourth-order Butterworth filter with a 425-Hz cutoff frequency, and a second low-pass filtering with a second-order Butterworth filter with a 100-Hz cutoff frequency.
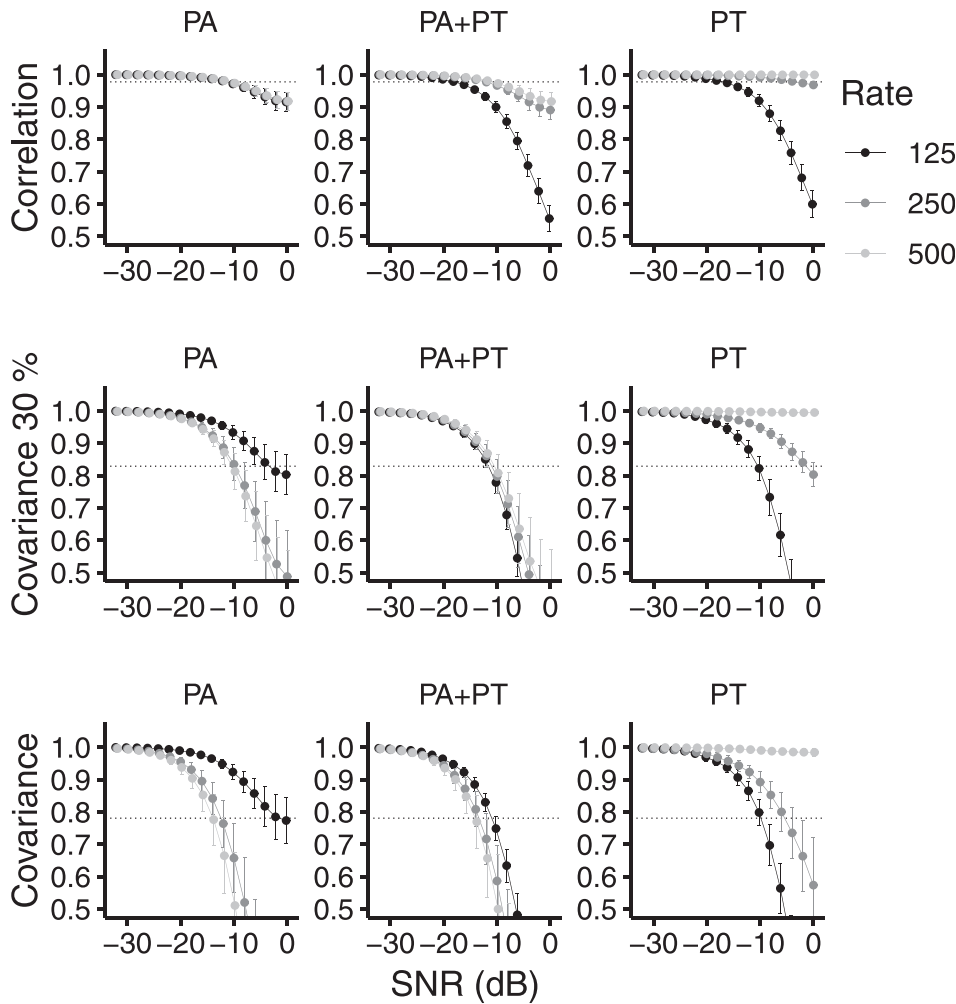
FIG. 7. Metrics of interaural "correlation" of the left and right channels of the NoSπ stimuli as a function of SNR (dB) in the PA (left column), PA + PT (middle column), and PT (right column) conditions. Stimuli were that of the NH listeners in the 50-Hz bandwith condition (experiment 2). Rows from top to bottom show the normalized envelope correlation, the normalized envelope covariance 30%, and the normalized envelope covariance, respectively. The low-pass cutoff frequency of the second-order filter was 100 Hz for all metrics in the figure. The pulse rate of the stimuli is shown with the darkness of shading. Error bars show ±1 standard deviation from the mean. Dotted lines show the metric value at which predicted thresholds (corresponding SNRs) explained the highest amount of variance in the average observed NH thresholds.

rates after peripheral processing. This is due to the fact that stimulus values that are close to zero in both the left and right channels have little effect on the normalized correlation. In contrast, the synchronized inter-pulse intervals increase the normalized covariance metrics. For the PA + PT stimuli, only the predicted thresholds of the normalized envelope covariance 30% match the pattern of thresholds of

TABLE II. The highest percent variance in the average NH NoSπ thresholds (50-Hz bandwidth stimuli) explained by the predicted thresholds. The low-pass cutoff of the second-order filter and the value of the metric at which the predicted thresholds explained the highest percent variance in the data are provided.

Metric	Variance explained (%)	Low-pass cutoff (Hz)	Threshold value of metric
Covariance	86.3	80	0.73
Covariance 10%	89.9	90	0.75
Covariance 20%	96.5	90	0.79
Covariance 30%	97.9	100	0.83
Covariance 40%	96.3	100	0.87
Covariance 50%	91.6	100	0.90
Covariance 60%	85.3	110	0.91
Covariance 70%	79.4	110	0.93
Covariance 80%	73.6	110	0.94
Covariance 90%	68.5	110	0.95
Correlation	63.9	110	0.96

the NH listeners. That is, the normalized envelope covariance 30% correctly predicted the approximately equivalent thresholds across pulse rates (Fig. 4; observed NoSπ thresholds were −11.5-, −12.0-, −11.4-dB SNR at 125, 250, and 500 pps, respectively), but the normalized envelope correlation predicted higher thresholds at higher pulse rates, and the normalized envelope covariance predicted somewhat lower thresholds at higher pulse rates. The failure of the normalized envelope correlation and covariance to predict performance across rates in the PA + PT condition seems to be due to the relative influence of the unsynchronized pulses (which are more apparent at lower rates; see Fig. 6) compared to the influence of the slow temporal envelope modulations (which are more apparent at higher rates). In the case of the normalized envelope correlation, the influence of the unsynchronized pulses is too strong relative to that of the slow temporal envelope such that the lower rates show lower correlation. In the case of the normalized covariance, the influence of the slow temporal envelope modulations is somewhat too strong relative to influence of the unsynchronized pulses such that the higher rates show lower covariance.

Table II shows the highest percent variance in the NH average thresholds that each metric was able to explain. The highest percent variance in the average NH thresholds was explained by the normalized envelope covariance 30%. This result lends support for a model of binaural hearing

J. Acoust. Soc. Am. **145** (5), May 2019

Todd *et al.*   2991

intermediate to the normalized envelope correlation and normalized envelope covariance in which correlations are calculated for deviations from a percentage of the mean as proposed by Bernstein and Trahiotis (1996b). This is in contrast to Bernstein and Trahiotis (1996b), which found little support for an intermediate metric over the normalized envelope correlation in predicting listener accuracy in discriminating NoSπ from NoSo with non-pulsatile stimuli. This was evaluated by examining symmetry about 0-dB SNR in the function relating listeners' discrimination accuracy (d') to SNR (from −30- to 30-dB SNR). Differences in the findings of that study and the present study are likely due to differences in the stimuli. While both studies examined stimuli in which only temporal envelope cues were useful, the stimuli in the present study were pulsatile, and the pulses contained varying degrees of information about the temporal fine structure and temporal envelope of the original low-frequency stimuli. In addition, a higher center frequency was used for the present study. Thus, the normalized envelope covariance 30% may only be preferable over the normalized envelope correlation for stimuli such as these. A model that could predict the data set of the present study as well as other data sets would be beneficial.

It should be noted that the highest percent variance in the average NH thresholds was explained with a low-pass second-order filter cutoff frequency of 100 Hz (Table II). Figure 8 shows the percent variance explained as a function of the low-pass filter cutoff frequency. This value of 100 Hz for the low-pass cutoff frequency is somewhat lower than that used in previous implementations of interaural correlation metrics (Goupell, 2012). The relatively low value for the low-pass cutoff frequency may be due to the high center frequency (9.2 kHz) of the stimuli in the present study; NH listeners can show poorer binaural processing performance with increasing center frequency (Bernstein and Trahiotis, 2014), which a lower low-pass cutoff frequency models. However, because we varied the value of the second-order
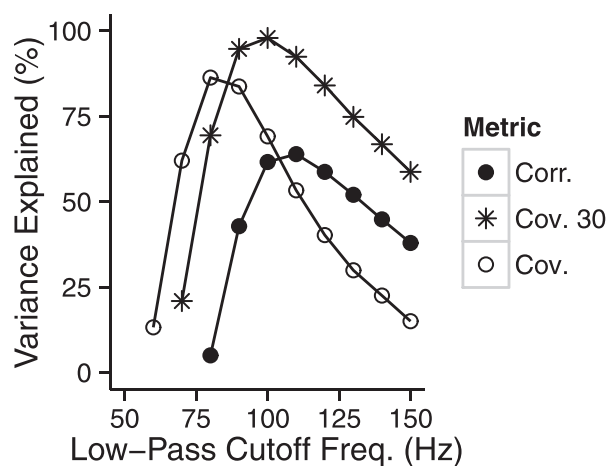

FIG. 8. The variance of the average NH NoSπ thresholds explained (%) by three metrics as a function of the low-pass second-order cutoff frequency (Hz) used in calculating the metrics. The three metrics shown are the normalized envelope correlation (Corr.), the normalized envelope covariance 30% (Cov. 30), and the normalized envelope covariance (Cov.). Each metric is shown with a different symbol and a connecting line.

TABLE III. The percent variance explained by the normalized envelope covariance 30% metric with various preprocessing parameters. The preprocessing parameters are half-wave rectification (HWR), fourth-order low-pass filtering at 425 Hz (fourth-order LP), second-order low-pass at 100 Hz (second-order LP), gammatone filtering (GT), and temporal envelope compression (Comp).

Preprocessing parameters	Cov30 variance explained (%)
No preprocessing	0
HWR, fourth-order LP	31.5
HWR, second-order LP	81.6
GT, HWR, second-order LP	70.4
Comp, HWR, second-order LP	88.5
GT, Comp, HWR, second-order LP	97.7

low-pass filter cutoff frequency to fit the data set and used a relatively small data set, this may have inflated the percent variance explained relative to that which might be found with a larger data set.

Table III shows the effect of various combinations of preprocessing parameters on the percent variance explained by the normalized envelope covariance 30%. The covariance 30% could explain 81.6% of the variance in the observed thresholds with only half-wave rectification and second-order low-pass filtering at 100 Hz. Gammatone filtering and envelope compression interacted in their effect on the percent variance explained in the presence of the 100-Hz second-order low-pass filtering. Including both gammatone filtering and envelope compression along with half-wave rectification and second-order low-pass filtering was necessary to attain the highest percent variance explained (97.7%).

With the present implementation of the model, we used the middle 200 ms of the stimulus, because this was the portion of the stimulus in which the tone was steady state. However, this relies on an assumption about the size of the temporal window of binaural processing (Akeroyd and Summerfield, 1999). Thus, we also evaluated the metrics with various stimulus durations. The normalized envelope covariance 30% was similarly predictive when we analyzed the middle 300 ms and the full duration of the stimuli. However, with the full stimulus duration (which included the ramps of the diotic noise), the normalized envelope covariance 20% was slightly better at predicting listener performance than the normalized envelope covariance 30% (97.9% vs 96.9% variance explained, respectively). Furthermore, the normalized envelope covariance was similar to the normalized envelope covariance 30% in its predictive ability (96.4% variance explained). The normalized envelope correlation remained low in its predictive ability across the different analysis durations (63.9%–65.6% variance explained). Including long durations of preceding and following silence reduced the predictive ability of all of the covariance metrics to that of the normalized envelope correlation (65.9%–65.4% variance explained across metrics). Thus, the fact that the window duration influences the predictive ability of the covariance metrics is a limitation of the covariance 30% model.

Akeroyd, M. A., and Summerfield, A. Q. (1999). "A binaural analog of gap detection," J. Acoust. Soc. Am. 105, 2807–2820.

2992   J. Acoust. Soc. Am. 145 (5), May 2019

Todd et al.

Bernstein, L. R. (**1991**). "Measurement and specification of the envelope correlation between two narrow bands of noise," Hear. Res. **52**, 189–194.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation," J. Acoust. Soc. Am. **78**, 514–523.

Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1995**). "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," J. Acoust. Soc. Am. **98**, 155–163.

Bernstein, L. R., and Trahiotis, C. (**1996a**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., and Trahiotis, C. (**1996b**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**2014**). "Sensitivity to envelope-based interaural delays at high frequencies: Center frequency affects the envelope rate-limitation," J. Acoust. Soc. Am. **135**, 808–816.

Bernstein, L. R., and Trahiotis, C. (**2017**). "An interaural-correlation-based approach that accounts for a wide variety of binaural detection data," J. Acoust. Soc. Am. **141**, 1150–1160.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: Accounting for NoSπ thresholds obtained with Gaussian and 'low-noise' masking noise," J. Acoust. Soc. Am. **106**, 870–876.

Brughera, A., Dunai, L., and Hartmann, W. M. (**2013**). "Human interaural time difference thresholds for sine tones: The high-frequency limit," J. Acoust. Soc. Am. **133**, 2839–2855.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1967**). "Release of masking for speech through interaural time delay," J. Acoust. Soc. Am. **42**, 124–138.

Carlyon, R. P., and Deeks, J. M. (**2002**). "Limitations on rate discrimination," J. Acoust. Soc. Am. **112**, 1009–1025.

Carlyon, R. P., Long, C. J., and Deeks, J. M. (**2008**). "Pulse-rate discrimination by cochlear-implant and normal-hearing listeners with and without binaural cues," J. Acoust. Soc. Am. **123**, 2276–2286.

Chen, H., and Zeng, F. G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Durlach, N. I. (**1964**). "Note on binaural masking-level differences at high frequencies," J. Acoust. Soc. Am. **36**, 576–581.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Eddins, D. A., and Barber, L. E. (**1998**). "The influence of stimulus envelope and fine structure on the binaural masking level difference," J. Acoust. Soc. Am. **103**, 2578–2589.

Egan, J. P. (**1965**). "Masking-level differences as a function of interaural disparities in intensity of signal and of noise," J. Acoust. Soc. Am. **38**, 1043–1049.

Fu, Q. J. (**2005**). "Loudness growth in cochlear implants: Effect of stimulation rate and electrode configuration," Hear. Res. **202**, 55–62.

Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.

Galvin, J. J., III, and Fu, Q. J. (**2009**). "Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users," Hear. Res. **250**, 46–54.

Goupell, M. J. (**2012**). "The role of envelope statistics in detecting changes in interaural correlation," J. Acoust. Soc. Am. **132**, 1561–1572.

Goupell, M. J., and Litovsky, R. Y. (**2015**). "Sensitivity to interaural envelope correlation changes in bilateral cochlear-implant users," J. Acoust. Soc. Am. **137**, 335–349.

Goupell, M. J., Majdak, P., and Laback, B. (**2010**). "Median-plane sound localization as a function of the number of spectral channels using a channel vocoder," J. Acoust. Soc. Am. **127**, 990–1001.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Ihlefeld, A., Carlyon, R. P., Kan, A., Churchill, T. H., and Litovsky, R. Y. (**2015**). "Limitations on monaural and binaural temporal processing in bilateral cochlear implant listeners," J. Assoc. Res. Otolaryngol. **16**, 641–652.

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (**2013**). "Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users," J. Acoust. Soc. Am. **134**, 2923–2936.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Long, C. J., Carlyon, R. P., Litovsky, R. Y., and Downs, D. H. (**2006**). "Binaural unmasking with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **7**, 352–360.

Lu, T., Carrol, J., and Fan-Gang, Z. (**2007**). "On acoustic simulations of cochlear implants," in Conference on Implantable Auditory Prostheses, Lake Tahoe, CA.

Lu, T., Litovsky, R., and Zeng, F. G. (**2011**). "Binaural unmasking with multiple adjacent masking electrodes in bilateral cochlear implant users," J. Acoust. Soc. Am. **129**, 3934–3945.

Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1511.

Schubert, E. D. (**1956**). "Some preliminary experiments on binaural time delay and intelligibility," J. Acoust. Soc. Am. **28**, 895–901.

Schubert, E. D., and Schultz, M. C. (**1962**). "Some aspects of binaural signal selection," J. Acoust. Soc. Am. **34**, 844–849.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

van de Par, S., and Kohlrausch, A. (**1995**). "Analytical expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157–3169.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

van der Heijden, M., and Joris, P. X. (**2010**). "Interaural correlation fails to account for detection in a classic binaural task: Dynamic ITDs dominate NoSpi detection," J. Assoc. Res. Otolaryngol. **11**, 113–131.

van Hoesel, R. J. M. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.

van Hoesel, R. J. M., Bohm, M., Pesch, J., Vandali, A., Battmer, R. D., and Lenarz, T. (**2008**). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies," J. Acoust. Soc. Am. **123**, 2249–2263.

van Hoesel, R. J. M., Jones, G. L., and Litovsky, R. Y. (**2009**). "Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse rate, modulation rate, and place of stimulation," J. Assoc. Res. Otolaryngol. **10**, 557–567.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

Zirn, S., Arndt, S., Aschendorff, A., Laszig, R., and Wesarg, T. (**2016**). "Perception of interaural phase differences with envelope and fine structure coding strategies in bilateral cochlear implant users," Trends Hear. **20**, 1–12.