Counting or discriminating the number of voices to assess binaural fusion with single-sided vocoders^{a)}

Jessica M. Wess,¹ Nathaniel J. Spencer,² and Joshua G. W. Bernstein^{1,b)}

¹National Military Audiology and Speech Pathology Center, Walter Reed National Military Medical Center, Bethesda, Maryland 20889, USA

²Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio 45433, USA

ABSTRACT:

For single-sided deafness cochlear-implant (SSD-CI) listeners, different peripheral representations for electric versus acoustic stimulation, combined with interaural frequency mismatch, might limit the ability to perceive bilaterally presented speech as a single voice. The assessment of binaural fusion often relies on subjective report, which requires listeners to have some understanding of the perceptual phenomenon of object formation. Two experiments explored whether binaural fusion could instead be assessed using judgments of the number of voices in a mixture. In an SSD-CI simulation, normal-hearing listeners were presented with one or two “diotic” voices (i.e., unprocessed in one ear and noise-vocoded in the other) in a mixture with additional monaural voices. In experiment 1, listeners reported how many voices they heard. Listeners generally counted the diotic speech as two separate voices, regardless of interaural frequency mismatch. In experiment 2, listeners identified which of two mixtures contained diotic speech. Listeners performed significantly better with interaurally frequency-matched than with frequency-mismatched stimuli. These contrasting results suggest that listeners experienced partial fusion: not enough to count the diotic speech as one voice, but enough to detect its presence. The diotic-speech detection task (experiment 2) might provide a tool to evaluate fusion and optimize frequency mapping for SSD-CI patients.

<https://doi.org/10.1121/10.0000511>

(Received 20 September 2019; revised 27 November 2019; accepted 6 December 2019; published online 30 January 2020)

[Editor: Adrian K. C. Lee]

Pages: 446–458

I. INTRODUCTION

Binaural hearing is critical to the ability to understand speech in complex, noisy environments. Individuals with single-sided deafness (SSD)—i.e., one normal-hearing (NH) ear and one deaf ear—lack these benefits, making everyday communication more challenging (Firszt *et al.*, 2017). For SSD listeners, cochlear implants (CIs) have been shown to partially restore some of the benefits of having two ears (Buss *et al.*, 2018; Galvin *et al.*, 2019), such as the ability to localize sound sources in space (Dirks *et al.*, 2019; Dorman *et al.*, 2015; Erbe *et al.*, 2015) and to better hear speech when it is spatially separated from noise (Vermeire and Van de Heyning, 2009). These benefits are generally understood not to require the computation of differences between the signals arriving at the two ears (Culling *et al.*, 2012; Zurek, 1993). Most of the reported speech-perception benefits provided by SSD-CIs in noise arise from head-shadow effects (Arndt *et al.*, 2017; Bernstein *et al.*, 2017; Vermeire and Van de Heyning, 2009), whereby listeners can take advantage of the better signal-to-noise ratio (SNR) in the CI ear when it is closest to the target talker or farthest from the interfering noise, or from binaural-summation effects (Arndt *et al.*, 2017),

whereby listeners benefit from having redundant copies of the speech information.

Recent evidence suggests that, in addition to head-shadow and binaural-summation benefits, SSD-CI listeners or NH listeners presented with SSD-CI vocoder stimulations can also benefit from the binaural integration of signals across the two ears to improve speech understanding in situations involving spatially separated competing talkers. A series of recent studies (Bernstein *et al.*, 2015, 2016; Wess *et al.*, 2017) employed a “contralateral-unmasking” paradigm, whereby a mixture of target and competing speech was presented monaurally to a NH ear. When a copy of the competing speech was then added to the CI (or a vocoder simulation of CI listening) in the opposite ear, listeners showed an improvement in speech understanding relative to the monaural condition. This performance improvement was interpreted as evidence that listeners were able to perceptually fuse the copies of the competing speech that were presented to the two ears. The reasoning was that listeners were able to binaurally fuse an unprocessed version of the competing speech in one ear with the CI- or vocoder-processed version in the other ear into a single perceptual object, thus enabling them to perceive a spatial difference between the competing speech (perceived as a diffuse image or a relatively compact image in the center of the head) and the monaural target speech. For NH listeners, a perceived difference in the spatial locations of target and competing speech can improve speech intelligibility (e.g., Freyman *et al.*, 2001).

^{a)}Preliminary data from this study were presented at the 172nd Meeting of the Acoustical Society of America, Honolulu, HI, December 2016.

^{b)}Electronic mail: joshua.g.bernstein.civ@mail.mil

Binaural fusion is defined as the subjective experience of the perception of one unified sound (instead of two) when listeners are presented with a pair of correlated signals, one to each ear (Steel *et al.*, 2015). For NH listeners, the binaural fusion of correlated signals occurs effortlessly, with the interaural binding effect so strong that it is impossible to perceive the same signal present to both ears as two separate objects (Sayers and Cherry, 1957). The perceptual process of binaural fusion is consistent with the hypothesis of a neural mechanism of binaural integration based on coincidence detection that allows for the computation of interaural cross-correlation (Colburn and Durlach, 1978; Saberi *et al.*, 1998; Sayers and Cherry, 1957). This binaural integration occurs in the mammalian superior olivary complex in the brainstem, with the higher-order auditory areas believed to receive a more complete and summed auditory object after subcortical binaural processing (Moore, 2000).

Much of our understanding of binaural fusion is derived from studies in which NH listeners are presented with noise to both ears with different levels of interaural correlation (e.g., Blauert and Lindemann, 1986). When the two noises are identical (i.e., interaural correlation = 1), these signals are perceived as a single perceptual object. As the degree of interaural correlation is parametrically decreased, the percept becomes systematically more diffuse, that is, no longer sounding like a punctate object at one spatial location, but instead a more diffuse sound with a wide perceived source width or as multiple separate sound sources (Whitmer *et al.*, 2012).

The fact that binaural fusion for NH listeners depends heavily on interaural correlation poses a potential problem for CI listeners. Bilateral CI (BI-CI) listeners, who receive similar electric signals in the two ears, sometimes report multiple sound images when the same stimulus is presented to the two ears, but also frequently report a fused auditory image (Fitzgerald *et al.*, 2015; Kan *et al.*, 2013). SSD-CI listeners receive very different signals in the two ears, which is likely to reduce the interaural correlation relative to more symmetric listeners. Interaural frequency mismatch between the stimuli presented to the acoustic and CI ears (Landsberger *et al.*, 2015) could further reduce interaural correlation and therefore binaural fusion (Aronoff *et al.*, 2015; Kan *et al.*, 2013; Suneel *et al.*, 2017; Wess *et al.*, 2017). Reduced binaural fusion could have important consequences on outcomes for CI listeners. If they are not able to properly make sense of the auditory scene by perceptually organizing it into the various sound sources present, this could impair their ability to attend to a target speech in the presence of a competing background (Bernstein *et al.*, 2016) and increase the amount of effort required for sustained auditory attention (Sladen *et al.*, 2018). On the other hand, the fact that NH and CI listeners have been shown to be sensitive to interaural envelope correlation for narrow-band or single-channel stimuli (Goupell and Litovsky, 2014, 2015) suggests that SSD-CI listeners might be able to perceive a diotic object as fused, despite the interaural differences in the temporal fine structure, which is discarded by CI processing. Aronoff *et al.* (2015) and Suneel *et al.*

(2017) found that when stimuli were spectrally matched in the two ears, NH listeners tended to report vocoded speech as a single fused perceptual object.

The most common ways to assess binaural fusion involve the elicitation of subjective responses from listeners, either as a binary fused/unfused response (Aronoff *et al.*, 2015; Suneel *et al.*, 2017), on a sliding scale from punctate to diffuse (Suneel *et al.*, 2017), or by having them point out where they heard one or more sounds on a one-dimensional sliding scale of lateral position within the head (Blauert and Lindemann, 1986; Goupell *et al.*, 2013; Kan *et al.*, 2013; Stakhovskaya and Goupell, 2017) or via multi-dimensional drawings of the spatial extent of sounds relative to the head in space (Cubick *et al.*, 2018; Whitmer *et al.*, 2012). Although these subjective approaches assess the quality of fusion, these methods are based on the underlying assumption that the listener understands what is meant by the question and is able to characterize his/her auditory perception at this level of abstraction. Unfortunately, this level of abstraction might not even be possible for some CI users. It has been argued that the subjective concept of binaural fusion may not be a concept to which CI users can relate to as well as NH listeners, due to the degradation of signals via the distortion of interaural correlation caused by independent CI processors (Fitzgerald *et al.*, 2015). An alternative strategy is to make the task more objective. One approach has been to present signals to opposite ears and ask the listener how many sounds they hear. For both SSD-CI and BI-CI listeners, this approach has shown that single-channel pulse-train stimuli delivered to opposite ears are sometimes perceived as separate sounds (e.g., Francart *et al.*, 2018; Kan *et al.*, 2013). It is not known, however, whether such an approach can be used to assess fusion for CI listeners presented with more complex stimuli involving a mixture of concurrent speech.

The goal of this study was to evaluate two potential methods for assessing binaural fusion for speech stimuli presented to NH adult listeners in a vocoder simulation of SSD-CI listening. Listeners were asked to make judgements regarding the number of voices present in a mixture as a way of determining whether two copies of a speech stimulus, unprocessed in one ear and vocoded in the other ear, were perceived as a single fused voice or two separate voices. Two experiments presented unprocessed speech signals to the left ear and a vocoder-processed copy of the same speech signals to the right ear. In some cases, additional monaural voices were added to the mixture to increase task difficulty. In experiment 1, listeners were asked to identify the total number of concurrent voices they heard in the mixture. Because some of the speech signals in the mixture were presented to both ears simultaneously, the degree of binaural fusion could be deduced depending on whether listeners counted each bilaterally presented speech signal as one voice or two. In experiment 2, listeners were asked to identify which of two mixtures contained a bilaterally presented speech signal.

The main manipulation employed in these experiments was to spectrally mismatch the vocoded signal presented to

the right ear relative to the unprocessed copy of the signal presented to the left ear. Because the CI electrode array does not reach the apex of the cochlea, but CI sound processors are generally programmed to provide the full range of speech information (approximately 100–8000 Hz), many SSD-CI listeners will experience a mismatch between the cochlear places of stimulation in the acoustic-hearing and CI ears (Bernstein *et al.*, 2018; Landsberger *et al.*, 2015). The introduction of spectral mismatch has been shown to reduce the likelihood that NH or CI listeners report the presence of a single interaurally fused sound for narrowband pulse-train stimuli (Goupell *et al.*, 2013; Goupell and Litovsky, 2015; Kan *et al.*, 2013). For broadband speech stimuli, the introduction of a spectral mismatch has been shown to reduce both the contralateral unmasking benefits for speech understanding (Wess *et al.*, 2017), and subjectively reported fusion (Aronoff *et al.*, 2015). It was hypothesized that for a mixture of voices that included at least one voice presented to both ears, spectral mismatch would lead to an increase in the number of voices reported (experiment 1) and poor performance in identifying the mixture containing an interaurally coherent voice (experiment 2). The idea was that if these measures of binaural fusion were found to be sensitive to interaural spectral mismatch, they might provide a useful tool to evaluate the efficacy of changes to SSD-CI frequency-mapping approaches aimed at reducing this mismatch.

II. EXPERIMENT 1: NUMEROSITY JUDGEMENTS

A. Methods

1. Approach

The goal of experiment 1 was to evaluate how many individual voices the listener heard in a mixture when one or two different voices were each presented diotically to the

unprocessed (left) and vocoded (right) ears. For simplicity, the presentation of the same speech signal to both ears is referred to as “diotic” input, even though the vocoded and unprocessed copies of the speech signal were not in fact identical. On a given trial, participants listened to a short segment of a speech mixture and reported how many total voices they heard. In the main test conditions, one or two of the voices in the mixture were presented diotically. It was thought that if listeners were able to perceptually fuse the unprocessed and vocoded copies of an individual speech signal, they should report hearing a single voice. Control conditions were also included to represent what listeners should have heard depending on whether they were able to perceptually fuse the diotic voices in the main test conditions. In the “fused” control conditions, a single monaural vocoded speech signal was included in the mixture in place of each diotic speech signal. In the “unfused” control conditions, two different monaural voices—one vocoded in the left ear and one unprocessed in the right ear—were included in the mixture in place of each diotic speech signal. The experimental conditions are listed in Table I and described in more detail below (Sec. II A 2).

The main manipulation examined in this experiment was to spectrally shift the vocoded stimulus to investigate the effects of interaural frequency mismatch. In the “matched” vocoder-alignment condition, the vocoder used the same synthesis and analysis filters, thereby yielding a match in the cochlear place of stimulation across the ears. In the “mismatched” vocoder-alignment condition, radiographic insertion depth data from Landsberger *et al.* (2015) were used to shift the vocoder analysis filters to approximate the average spectral mismatch between the frequency allocation of a CI electrode array and the characteristic frequency associated with the electrode location. The vocoding process is described below in Sec. II A 3.

TABLE I. Experimental conditions for experiment 1. Each set contained one test condition and two control conditions. In the last two columns, each V_i represents one voice in the mixture. In the test conditions, copies of one or two voices were presented to both ears (e.g., V_1 in the top row). In the unfused control conditions, the diotic voice from the respective test condition was replaced by two voices, one in each ear (e.g., V_1 and V_2 in the second row). In the fused control condition, the diotic voice from the respective test condition was replaced by a single vocoded voice in the right ear (e.g., V_1 in the third row).

Set	Configuration	Condition	Total # voices	Right ear (vocoded)	Left ear (unprocessed)
I	Center	Test	1	V_1	V_1
		Unfused control ^a	2	V_1	V_2
		Fused control	1	V_1	none
II	Center/Left	Test	2	V_1	V_1+V_2
		Unfused control ^b	3	V_1	V_2+V_3
		Fused control ^a	2	V_1	V_2
III	2 Center/ 1 Left	Test	3	V_1+V_2	$V_1+V_2+V_3$
		Unfused control	5	V_1+V_2	$V_3+V_4+V_5$
		Fused control	3	V_1+V_2	V_3
IV	1 Center/ 2 Left	Test	3	V_1	$V_1+V_2+V_3$
		Unfused control	4	V_1	$V_2+V_3+V_4$
		Fused control ^b	3	V_1	V_2+V_3

^aThe fused control condition for Set II was a duplicate of the unfused control condition for Set I, and was not repeated twice.

^bThe fused control condition for Set IV was a duplicate of the unfused control condition for Set II, and was not repeated twice.

2. Test conditions

The experimental conditions are summarized in Table I. There were four sets (*I*, *II*, *III*, and *IV*) of conditions, each consisting of a test condition and two control conditions. For each row in Table I, the last two columns indicate the different voices (V_i) present in the mixture in each ear. Some of the voices were presented diotically (i.e., the same speech presented to unprocessed and vocoded ears), and are listed in both columns, while some voices were presented monaurally (to only one ear), and are listed in only one column. The test conditions all involved the diotic presentation of one or two of the voices in the mixture. The control conditions, containing no diotic voices, represented how many voices should have been reported if each diotic voice in the test condition were completely fused and perceived as a single voice (fused control), or completely unfused and perceived as two separate voices (unfused control).

Set *I* had the fewest total number of voices. The test condition consisted of a single diotic voice (V_1) presented to both ears, with the goal of determining whether the subject reported the diotic voice as one (fused) voice, or two (unfused) voices. This condition is referred to as “center” because if listeners were able to fuse the unprocessed and vocoded copies of the diotic speech signal, they should have heard one voice (V_1) in the center. In the unfused control condition, two different speech signals (V_1 and V_2) were presented to the vocoded and unprocessed ears in lieu of the single diotic voice (V_1) from the test condition. The idea was that if a listener perceived the diotic voice in the test condition as unfused, their response (number of voices) should match the case where a different speech signal was presented to each ear. In the fused control condition, a single vocoded voice (V_1) was presented to the right, vocoded ear. Here, the idea was that if listeners perceived the diotic voice (V_1) as a fused combination of the unprocessed speech in one ear and the vocoded speech in the other, the reported number of voices should match the case where one vocoded speech signal was presented.

As shown in Table I, sets *II*, *III*, and *IV* introduced additional voices into the mixture. The test and control conditions for set *II* were similar to those from set *I*, except that in each case, an extra unprocessed voice was added to the left ear. This condition is referred to as “center/left” because if listeners were able to fuse the unprocessed and vocoded copies of the diotic speech signal, they should have heard one voice (V_1) from the center, plus an unprocessed monaural voice (V_2) on the left. For set *III* (“2 center / 1 left”), the test condition involved two diotic voices (V_1 and V_2) and one monaural voice (V_3). For set *IV* (“1 center / 2 left”), the test condition involved one diotic voice (V_1) and two monaural voices (V_2 and V_3). As was the case for set *I*, the two control conditions in sets *II–IV* were designed to represent either the number of voices a listener would have reported if they completely fused the diotic speech signals (fused control), or if they heard these speech signals as two

separate voices (unfused control). In two cases, the fused control conditions from one set were identical to the unfused control conditions from another set (footnotes in Table I). These conditions were presented only once.

3. Stimuli

This experiment used the City University of New York (CUNY) topic-sentence corpus (Boothroyd and Kishon-Rabin, 1988) consisting of lists of sentences pertaining to 12 different topic areas such as food, work, family, or weather. For example, “The thunder and lightning from the storm last night woke up all of us” was a sentence pertaining to the weather. The speech materials were derived from recordings of the corpus produced by two different talkers, one male and one female. To create additional male and female talkers, these recordings were modified using PRAAT software (Boersma, 2001) to change the fundamental frequencies (F0s) and spectral envelopes of the speech recordings. A total of eight voices were used: the two original talkers and three additional male and three additional female talkers that were derivatives of the original two recordings. From the original recordings, the additional talkers were synthesized by calculating the F0 and spectral envelope every 10 ms using overlapping 40-ms wide windows. For each sample identified as a voiced segment, the F0 was multiplied by a fixed factor for each synthesized talker, while for all segments, the assumed vocal tract length was adjusted to effectively shift the formant frequencies by a different fixed factor. The new speech waveforms were then reconstructed using the pitch-synchronous overlap-and-add method (Moulines and Charpentier, 1990). The three additional male talkers were generated from the recordings of the original male talker (mean F0 = 115 Hz) using F0 factors of 0.8, 1.2, or 1.3 and spectral-envelope factors of 0.87, 1.05, or 0.91, respectively. The three additional female talkers were generated from the recordings of the original female talker (mean F0 = 165 Hz) using F0 factors of 0.9, 1.18, or 1.13 and spectral-envelope factors of 0.91, 1.11, or 0.91, respectively.

Sentences for a given topic area and talker were concatenated, thereby creating 96 different paragraphs (eight talkers $\times$ 12 topics) of 118–138 s each. On each trial, a mixture of speech segments was presented to the listener. Each of the speech segments in the mixture was produced by a different talker (selected at random from the eight possible talkers, without regard to talker gender) and was derived from a different topic paragraph (selected at random from the 12 possible topics), with a two-second segment selected at random from the paragraph. The individual speech segments in the mixture were presented simultaneously, each at a level of 60 dB sound pressure level, before mixing.

The left ear was always presented with an unprocessed mixture of voices. The right ear was always presented with a noise-vocoded mixture of voices. For the right ear, the vocoding process was applied after the concurrent speech

signals were added together. To vocode, speech envelopes were extracted in seven frequency channels (Fig. 1) then used to modulate seven noise bands to excite specified regions of the cochlea. Noise carriers were used, rather than pulse carriers that have been proposed as more analogous to biphasic pulsatile electrical stimulation provided by CIs (e.g., Mesnildrey *et al.*, 2016), to ensure that the stimuli were similar to those used in our previous studies of speech perception in a multitalker situation for noise-vocoder-simulated SSD-CI listening (Bernstein *et al.*, 2015, 2016; Wess *et al.*, 2017; Wess and Bernstein, 2019). To extract the envelopes, stimuli were passed through a bank of fourth-order Butterworth “analysis” filters. The envelope of the signal in each channel was extracted via half-wave rectification, then low-passed filtered at 400 Hz (second-order Butterworth). Each envelope was then multiplied by a white-noise carrier, with the resulting signal then passed through a series band-pass “synthesis” filters. The level of the resulting signal in each vocoder channel was adjusted to be equal to the root-mean squared level of the input signal for that channel. Finally, the signals were summed across vocoder channels to create the noise-vocoded signal.

The frequency cutoffs for the analysis and synthesis bands in the matched and mismatched conditions are shown in Fig. 1. For the matched vocoder-alignment condition, the synthesis filters were identical to the analysis filters, such that a given spectral component of the speech signal was delivered to the same frequency region in both ears. For the mismatched vocoder-alignment condition, interaural spectral mismatch was introduced using synthesis filters that did not match the analysis filters used to extract the envelope, such that a given spectral component would be delivered to two different frequency regions in the unprocessed and vocoded ears. For a CI listener, each electrode stimulates a specific cochlear place that cannot be changed through reprogramming. In the vocoder simulation, this was emulated by having a fixed set of synthesis bands, so

FIG. 1. Analysis- and synthesis-filter frequency cutoffs for the matched and mismatched vocoder-alignment conditions. The synthesis filters were the same for both vocoder-alignment conditions to simulate a fixed physical location of the electrode array on the basilar membrane. For the mismatched vocoder condition, the analysis-band cutoff frequencies were shifted downwards to represent a standard frequency allocation. For the matched vocoder condition, the analysis-band cutoff frequency cutoffs were set to be equal to the synthesis-band frequency cutoffs.

that the vocoder was always stimulating the same set of locations on the cochlea. In contrast to the fixed electrode location, clinical speech-processor programming software allows for the reallocation of the frequency bands delivered to each electrode. Changes to the sound processor allocation were simulated by either setting the frequency cutoffs based on a standard clinical allocation representing a shift from the synthesis bands (mismatched map), or by setting the cutoffs to equal the synthesis bands (matched map).

The synthesis-band cutoffs (and the analysis-band cutoffs in the matched condition) shown in Fig. 1 were determined based on average insertion angles for a standard MED-EL electrode array from radiographic data reported by Landsberger *et al.* (2015) in combination with the frequency allocations recommended by MED-EL MAESTRO clinical programming software in the case where only seven electrodes are active. The data from Landsberger *et al.* (2015) suggest that on average, the most apical electrode will only reach the cochlear place associated with a characteristic frequency of about 400 Hz. This was represented by the setting low-frequency cutoff of the lowest synthesis band to 438 Hz. The upper frequency cutoff for the highest synthesis band was set to 8500 Hz, the highest frequency cutoff allowed for a MED-EL speech processor. This was done because in the matched condition, the analysis and synthesis bands were the same, and the highest analysis-band cutoff was set based on a standard CI frequency allocation. The frequency cutoffs for the seven synthesis bands were then set based on the cutoffs automatically generated by the CI programming software for a seven-electrode configuration with an overall bandwidth of 438–8500 Hz.

The analysis-band cutoffs for the mismatched condition (Fig. 1) were set to be lower than the synthesis bands, to deliver frequency information to a more basal location than in an acoustic ear, as would be expected for a typical CI insertion and standard frequency allocation. This was done by converting the synthesis-filter band frequency cutoffs via a logarithmic interpolation of the Stakhovskaya *et al.* (2007) spiral-ganglion map (representing the relationship between electrode location and the best acoustic frequency associated with that location), and a logarithmic interpolation of the mean radiography data from Landsberger *et al.* (2015) (representing the relationship between electrode location and the actual clinical frequencies delivered to that location in a standard clinical map). For example, the upper synthesis-filter cutoff frequency was 8500 Hz. According to Stakhovskaya *et al.*, a spiral-ganglion neuron tuned to 8500 Hz corresponds to an insertion angle of 62°. From the radiographic data of Landsberger *et al.*, an average electrode located at this insertion angle is tuned to a frequency of 3849 Hz, which was assigned as the upper cutoff frequency for the highest analysis band in the “mismatched” condition.

4. Procedure

On each trial, the listener was presented with a mixture of voices. After each trial, a graphical user interface (GUI)

appeared which displayed the question “How many total voices do you hear?” The listener then used the computer mouse to select the button corresponding to the number of voices they heard (1–6). No feedback was provided following each response, to allow for the measurement of the listener’s subjective impression of the number of voices in the test conditions (with no assumed correct answer).

The experiment consisted of blocks of 80 trials each. For each trial within the block, both the vocoder-alignment condition (matched or mismatched) and the test or control condition (Table I) was selected at random. A total of five blocks were presented for each listener, for a total of 20 trials for each combination of vocoder-alignment condition and test or control condition, and 400 trials in the experiment.

The main experiment was preceded by a training session, which was similar to the experimental session, except that only the six control conditions were presented (no test conditions) and listeners were provided feedback at the end of each trial via blinking of the numbered button corresponding to the correct total number of voices in the mixture. As in the main experiment, the vocoder-alignment condition and control condition were selected randomly on each trial. Listeners were presented with two blocks of 60 trials each, for total of 10 trials for each combination of vocoder-alignment and control condition, and a total of 120 training trials.

5. Participants

Ten paid listeners (age range 18–30) participated in experiment 1. These listeners were all highly trained in psychoacoustic experiments, although they had no previous exposure to vocoded speech, were members of a listener panel that is paid to participate in psychoacoustic experiments on an ongoing basis at the Air Force Research Laboratory, Wright Patterson Air Force Base, Ohio. All listeners had normal audiometric thresholds, defined as equal to or better than

20 dB hearing level at octave frequencies between 125 and 8000 Hz, and reported no cognitive or neurological disorders.

6. Apparatus

Listeners were seated in a sound booth and directed their attention to a computer screen. The speech stimuli were generated by MATLAB and played via an RME Hammerfall (Haimhausen, Germany) sound card and presented over Sennheiser HD 280 headphones.

B. Results

The results of experiment 1 are shown in Fig. 2. Each panel plots the mean reported number of voices as a function of listening condition for a given set of conditions (i.e., one test condition and two control conditions). The two vocoder-alignment conditions are represented by black (matched) and gray (mismatched) bars. The horizontal lines in each plot represent the number of voices the listener should have reported in the test condition if they had perceived the diotic voice(s) as completely unfused (upper solid line) or completely fused (lower dashed line).

Two main trends are apparent in the data. First, an examination of the results for the fused and unfused control conditions provides information regarding listeners’ accuracy in reporting the number of voices in the mixture. Although the number of reported voices generally increased with an increasing number of actual voices in the mixture, listeners were most accurate in counting the voices when there were relatively few voices in the mixture (i.e., one to three voices in sets I and II). When there were a greater number of voices in the mixture (i.e., three to five voices in sets III and IV), listeners tended to under-estimate the total number of voices, which may be related to a maximum limit of about four voices that can be counted in an auditory

FIG. 2. Results from experiment 1 showing the mean reported number of perceived number of voices as a function of the experimental conditions listed in Table I and the two vocoder-alignment conditions. Each panel represents one set of test, fused control and unfused control conditions listed in Table I. The dashed horizontal lines in each panel indicate the expected number of reported voices if the listener were able to fuse the unprocessed and vocoded copies of the speech signals presented to the two ears, while the solid horizontal lines in each panel represent the expected number of reported voices if listeners if they perceived the copies as two separate voices. Error bars represent $\pm$ one standard error of the mean reported number of voices across listeners.

scene (Zhong and Yost, 2017). Second, and most importantly for the main goal of the study, a comparison of the results for the test and control conditions, and for the matched and mismatched vocoder-alignment conditions, allowed for an assessment of whether listeners responded as if they were perceptually fusing the diotic voices in the mixture. Figure 2 clearly shows that the reported number of voices in the test condition generally corresponded to the reported number of voices in the unfused control condition, suggesting a lack of fusion. This was generally true regardless of how many voices were in the mixture and regardless of vocoder-alignment condition, with matched and mismatched vocoders producing the same result.

To assess these trends statistically, repeated-measures analyses of variance (ANOVAs) were conducted for each set of conditions. These analyses examined the effects of vocoder-alignment condition (two levels) and experimental condition (three levels: one test and two control conditions) on the reported number of voices, with Bonferroni corrections applied for four comparisons (criterion $p < 0.0125$). For all four sets, there was a significant main effect of condition (test versus unfused versus fused control, $p < 0.00001$), but no significant main effect of vocoder-alignment condition or significant interactions between the two factors ($p > 0.0125$). In all four sets of conditions, pairwise *post hoc* t-tests (Bonferroni corrected for three comparisons, i.e., criterion $p < 0.0167$) found statistically significant differences between the test and fused control conditions, indicating that the number of reported voices in the diotic case was more than would be expected if listeners were perceptually fusing the diotic stimuli. There were also significant differences between the fused and unfused control conditions, as expected. The only instance of a significant difference between the test and unfused control conditions was in set II. But in this case, the mean number of reported voices was slightly higher in the test condition (3.10 voices) than in the unfused control condition (2.93 voices), which is the opposite direction from what would be expected if listeners were partially fusing the diotic signal. For the other three sets, the number of reported voices in each test condition was not statistically different from the number reported for the corresponding unfused control condition, suggesting that listeners were counting the diotically presented speech as two separate voices.

III. EXPERIMENT 2: SPEECH-MIXTURE DISCRIMINATION

A. Rationale

In experiment 1, listeners counted each diotic voice in each test condition as if two separate voices were presented to the two ears (i.e., the number of voices reported in the unfused control condition), suggesting a lack of fusion. This result contrasts with the results of Wess *et al.* (2017) showing that NH listeners were able to integrate across the ears an unprocessed and a vocoded copy of a speech signal to perceptually separate competing speech, and that this ability

was negatively impacted by interaural frequency mismatch. One possible explanation for this discrepancy is that listeners experienced only partial fusion of the diotic speech signal—enough to facilitate the speech-perception advantage reported by Wess *et al.* (2017), but not enough to report the speech as a single unified stimulus. To investigate the possibility that listeners were able to at least partially fuse the diotic speech, experiment 2 tested whether listeners were able to detect the presence of interaural coherence between the unprocessed speech signal presented to one ear and the vocoded copy presented to the other ear.

B. Methods

1. Approach

The experiment was a two-interval, two-alternative forced-choice task where the listener was required to identify which interval contained a diotic speech signal. The signal interval contained two copies of the same speech waveform, presented unprocessed to the left and vocoded to the right ear (like the diotic test conditions in experiment 1). The reference interval replaced the diotic speech signal with two different talkers saying different things, again unprocessed in the left and vocoded in the right ear (like the unfused control conditions in experiment 1). Performance in identifying the interval with the diotic speech signal was measured as additional voices were added to the mixture in the unprocessed left ear. There was always only one vocoded voice presented to the right ear, which either matched (signal interval) or did not match (reference interval) one of the unprocessed voices in the left ear. Thus, the task involved the identification of the interval where the unprocessed mixture in the left ear contained the same speech signal as the vocoded version in the right ear.

2. Procedure

Nine paid listeners with normal audiometric thresholds (age range 18–30) participated in experiment 2, seven of whom had also participated in experiment 1. The speech stimuli and vocoder processing were the same that were used in experiment 1. Two successive two-second speech mixtures were presented, separated by 500 ms of silence. After each trial, a graphical-user interface appeared on the screen and the listener was prompted with the question: “Which interval contained one ‘stereo’ voice that was the same in both ears?” The listener used the computer mouse to select the button corresponding to the first or second interval. Each block consisted of 48 trials, with the vocoder-alignment condition (mismatched or matched) held fixed throughout the block, but the total number of unprocessed voices (1, 2, 3, 4, 5, or 7, including the diotic voice) presented to left ear was selected randomly on each trial. Following the response, listeners were provided feedback via blinking of the numbered button corresponding to the correct signal interval. The talkers, topics, and two-second segments in each interval were selected at random.

The training portion of the experiment consisted of a subset of the complete set of conditions involving 1, 2 or 3 unprocessed voices presented to the left ear and both vocoder conditions (mismatched and matched). Listeners were presented with 4 blocks (2 with each vocoder type) of 30 trials each. This yielded 20 trials for each combination of number of voices and vocoder condition, and a total of 120 trials for the training portion of the experiment.

For the main experiment, listeners were presented with a total of six blocks (three blocks for each vocoder-alignment condition). Thus, there were a total of 24 trials for each combination of number of voices and vocoder-alignment condition, and 288 trials in the experiment.

C. Results

The results of experiment 2 are shown in Fig. 3. The mean proportion correct in identifying the interval containing a diotic voice is plotted as a function of the total number of unprocessed voices presented to the left ear. With only a single voice in each ear (i.e., a diotic voice in the signal interval and two different voices in the reference interval), performance was very high regardless of the vocoder type. With an increasing number of unprocessed voices in the left ear, performance decreased, but more rapidly for the mismatched than for the matched vocoder.

The data were analyzed using binary logistic regression. This analysis, while similar in concept to a traditional ANOVA in that it provides information regarding the main effects and interactions between the study variables, has two important advantages. First, because the data were binary in nature (i.e., correct or incorrect) the regression

analysis accounted for the likelihood that percentage-correct scores were different based on the number of trials presented. Second, by incorporating a logistic link function into the analysis, the regression analysis also accounted for the fact that variability is reduced at the extremes of the psychometric function, such that a given change in percentage-correct performance is more meaningful near the end of the psychometric function (e.g., 95% versus 90%) than near the middle (e.g., 65% versus 60%).

The logistic-regression analysis revealed significant main effects of vocoder-alignment condition [$\chi^2(1) = 10.40$, $p < 0.001$] and number of voices [$\chi^2(5) = 89.89$, $p < 0.001$] and a significant two-way interaction between vocoder-alignment condition and number of voices [$\chi^2(5) = 205.87$, $p < 0.001$]. *Post hoc* tests were performed on the difference between the vocoder-alignment conditions at each number of voices condition. After Bonferroni corrections for six comparisons, the vocoder-alignment conditions were found to differ significantly in the case where there were a total of five voices presented unprocessed to the left ear ($p = 0.002$, asterisk in Fig. 3).

IV. DISCUSSION

A. Summary of main findings

This study was designed to evaluate a measure of perceptual fusion based on counting or discriminating the number of voices in a mixture. Specifically, the question was posed whether such a measure would be sensitive to interaural frequency mismatch between an unprocessed speech signal presented to one ear and a vocoded copy of the speech signal presented to the other ear. The two experiments yielded distinctly different results. In experiment 1, listeners were asked to report the number of voices they heard in the mixture. The number of voices they reported in the test condition containing one or two diotic voices was no different from the unfused control conditions, where two different voices were presented to opposite ears instead of an unprocessed and a vocoded copy of the same speech waveform. Thus, the results of experiment 1 suggest that listeners perceived the unprocessed and vocoded copies as two distinct voices. Furthermore, they reported the same number of voices whether the vocoded signals in the right ear were spectrally shifted or frequency-matched with the unprocessed signals in the right ear.

In experiment 2, listeners were asked to discriminate between a speech mixture that contained a diotic voice (unprocessed in one ear and vocoded in the other) and a mixture that instead contained two different voices in the two ears. Listeners were able to detect the diotic speech when there were relatively few voices in the mixture, but performance declined as additional unprocessed voices were added to the left ear. Critically, listeners were significantly better at detecting the interval with a diotic voice when the vocoded copy of the speech segment was spectrally matched to the unprocessed speech. This result suggests that listeners were experiencing enough fusion of the diotic voice to recognize that a “stereo” voice was present

FIG. 3. Results from experiment 2 plotting the proportion correct in identifying the interval containing the diotic voice that was presented unprocessed in a mixture of voices in the left ear and vocoded in isolation to the right ear. The horizontal line indicates chance performance. Error bars represent $\pm$ one standard error of the mean across listeners.

in the mixture, and that this fusion decreased with interaural frequency mismatch.

B. Interpretation of the results with respect to fusion

Our working understanding of binaural fusion is that two correlated signals that are presented to opposite ears are perceived by a NH listener as a single unified sound, rather than two separate sounds. The measurement of fusion is a difficult problem to address because it depends on listeners having the ability to understand what is meant by the concept of a single auditory object. The difficulty in assessing perceptual fusion is highlighted by the fact that the two approaches employed in this study yielded different sets of results. The contrasting results from experiments 1 and 2 suggest that listeners were experiencing partial or incomplete fusion of the unprocessed and vocoded signals. On one hand, the diotic stimulus might not have been sufficiently fused for a listener to identify it as one voice instead of two when asked for a free answer (experiment 1). On the other hand, the diotic signal might have been sufficiently fused for listeners to detect a single diotic input voice in the mixture (experiment 2). These results suggest that traditional measures of fusion based on a subjective report might be less sensitive to changes in interaural coherence than a discrimination task that has a correct answer.

One important difference between the current study and previous studies of fusion in real or simulated CI listeners is that these experiments involved a mixture of voices rather than a single voice. Most studies of binaural fusion for CI listeners or NH listeners presented with simulated CI listeners involve only a single non-speech sound presented to each ear such as a tone, noise, or pulse train (Goupell *et al.*, 2013; Kan *et al.*, 2013; Reiss *et al.*, 2014a), or a segment of speech presented to each ear (Aronoff *et al.*, 2015; Suneel *et al.*, 2017). The complex mixture of voices used in the current study likely required the use of stream segregation and object formation in addition to binaural fusion. By evaluating fusion in the context of a complex mixture of voices, this study revealed the negative effects of the spectrally mismatched vocoder on binaural hearing processes. This result was not apparent in the simplest condition involving a single voice, but instead emerged when a complex mixture of voices was presented to the listener. In the case of a single diotic voice, listeners likely had access to cues other than a subjective perception of binaural fusion to complete the task. For example, with only one voice that was the same in each ear, listeners might have been able to monitor the two ears for common words or speech rhythms. This would have allowed them to identify the interval containing the diotic voice, even if the matched and unmatched vocoders sounded qualitatively different. In the more complex case where the listener was presented with a mixture of unprocessed voices in one ear, monitoring each ear for similar patterns would have been much more difficult, forcing listeners to instead rely on a sensation of binaural fusion arising from interaural correlation.

C. Effects of interaural spectral mismatch

By identifying an effect of interaural mismatch on performance in a discrimination task that requires the detection of interaural coherence to obtain a correct answer, the results of the current study extend other results in the literature that have examined subjectively reported binaural fusion for CI and vocoder listeners. Aronoff *et al.* (2015) examined the effect of spectral compression on subjectively reported binaural fusion for NH listeners listening to speech through vocoder simulations where one ear was spectrally mismatched relative to the other, and concluded that spectral mismatch resulted in reduced subjective fusion. Goupell *et al.* (2013) examined the effect of spectral mismatch on binaural fusion in vocoder stimulations of BI-CI listeners using simple acoustic pulse train stimuli that simulated CI single-electrode constant-amplitude electric pulsatile stimulation. They varied the spectral mismatch and measured the perceived image location on a GUI that displayed a face, which the listener could click on. They found that as spectral mismatch increased between the ears, listeners tended to increasingly report more than one auditory image and that the perception was biased towards the ear where the stimulus had a higher carrier frequency. In a follow-up study, Kan *et al.* (2013) found a similar result for BI-CI listeners. Reiss *et al.* (2014a) examined fusion and spectral integration for bimodal CI listeners (i.e., one ear with a cochlear implant and one with a hearing aid). They found that many of the listeners reported fusion, even when listeners had very large spectral mismatch between their CI and hearing-aid ear. This suggests that at least for some listeners with prolonged experience with spectral mismatch, that some plasticity is possible, a possibility not captured by the current study.

The decline in performance in the mismatched vocoder-alignment condition may be related to a degradation of envelope correlation between the ears, with the spectral mismatch disrupting the interaural envelope correlation between the vocoded signal and the acoustic signal. Correlated envelope information has been shown to facilitate auditory object formation and binaural fusion (Carrell and Opie, 1992). Frequency mismatch shifted the envelope in one ear to higher frequencies, thus lowering interaural correlation in each frequency band. The perceptual consequence of this reduced interaural correlation would be to reduce temporal-coherence cues that have been shown to be critical for auditory grouping (Shamma *et al.*, 2011). In fact, temporal integration has been shown to be negatively impacted by spectral mismatch in CI listeners (Poon *et al.*, 2009).

The current study was focused on the assessment of the subjective quality of binaural fusion, and the results do not provide a direct connection between binaural fusion and the functional consequences for the everyday experience of SSD-CI listeners who face an interaural frequency mismatch. However, it is likely that reduced fusion has an impact on speech-reception performance (e.g., Wess *et al.*, 2017) and the amount of effort that listeners need to expend

to communicate in everyday settings (e.g., [Steel et al., 2015](#)). Whereas [Bernstein et al. \(2015, 2016\)](#) found that the addition of a vocoded copy of the masker signal to the ear contralateral to the target speech facilitated contralateral unmasking, [Wess et al. \(2017\)](#) found that this benefit was eliminated or largely diminished after CI distortions were implemented in the vocoder. The current results suggest that the loss of contralateral unmasking might be explained in terms of lost binaural fusion ability or improper object formation. Although experiment 1 suggested little fusion between the voices across the ears, the results from experiment 2 likely reflect some degree of binaural integration of the coherent speech signals, in that the listeners could identify interaural coherence enough that they could tell the difference between one voice and two voices. Thus, the conclusion from experiment 2 is that listeners were at least receiving partial fusion, but not fully integrated auditory object formation. This partial fusion was likely sufficient to generate the unmasking benefit observed previously ([Bernstein et al., 2015, 2016; Wess et al., 2017](#)).

Another area where reduced binaural fusion could have an impact is in terms of the increased cognitive load required for an SSD-CI listener to achieve sufficient speech understanding to be able to function adequately in a complex communication setting. Listening effort is the subjective experience of the additional attentional resources needed to devote to achieving sufficient speech understanding for effective communication ([Pichora-Fuller et al., 2016](#)). Listening effort has been assessed in CI and vocoder listeners using various techniques, including the measurement of pupillary response ([Winn et al., 2015](#)), performance on a secondary task in a dual-task paradigm ([Perreau et al., 2017](#)), and self-report ([Sladen et al., 2018](#)). While these approaches generally conclude that cochlear-implant listeners require more effort to perform speech-reception tasks than their acoustic-hearing listeners, evidence is limited that these methods can distinguish between different types of processing for a given hearing-impaired individual ([Ohlenforst et al., 2017](#)). Preliminary reports suggest that measures of listening effort may uncover some of the subjective benefits that a CI provides to individuals with SSD ([Litovsky et al., 2016](#)). It is an open question whether such approaches would be sensitive to differences in CI programming analogous to the interaural frequency-mismatch conditions tested here.

D. Clinical implications for SSD-CI listeners

Currently, CIs are programmed for the profoundly deaf with little or no attention given to binaural hearing ability after CI implantation. With a typical array insertion and standard (mismatched) frequency allocation, interaural spectral mismatch is highly likely for SSD-CI listeners ([Landsberger et al., 2015](#)), and may limit the opportunities for listeners to benefit from binaural hearing ([Goupell et al., 2013; Kan et al., 2013; Litovsky et al., 2012; Wess et al.,](#)

[2017](#)). Theoretically, this issue could be easily overcome from a technical standpoint, if it was known which characteristic frequencies of the auditory nerve fibers being stimulated by each electrode in the array.

Several methods for determining the effective stimulus locations for individual electrodes have been proposed, including psychometric methods such as pitch matching ([Schatzer et al., 2014](#)) or interaural time-difference discrimination ([Bernstein et al., 2018; Francart et al., 2018; Kan et al., 2013](#)), and electrophysiological methods such as the binaural interaction component of the auditory brainstem response ([Hu and Dietz, 2015](#)). As possible candidates for clinical use, these approaches each have their limitations. Pitch matching is susceptible to non-sensory biases ([Carlyon et al., 2010; Goupell et al., 2019](#)) and plasticity ([Reiss et al., 2014b](#)), while measures of binaural processing are time consuming and difficult for the listener and experimenter ([Bernstein et al., 2018; Hu and Dietz, 2015](#)). CT scans or radiographs are fast and therefore clinically feasible and provide information regarding cochlear electrode location ([Landsberger et al., 2015; Noble et al., 2014](#)), but they do not provide information about the neural response patterns. The discrimination paradigm in experiment 2 could potentially be used to determine the efficacy of new frequency-shifted maps for SSD-CI listeners. Critically, this binaural-coherence detection task employed speech stimuli but did not rely on measuring intelligibility. This is important because speech intelligibility requires an adjustment period to adapt to programming changes, which makes intelligibility an imperfect tool for assessing acute map changes. This approach also has the potential to serve as a clinical tool to identify an optimal frequency map to reduce perceptual mismatch. Using the paradigm introduced in experiment 2, an audiologist could potentially fit a listener with a few different maps and determine which map is best for binaural fusion. Relative to the above psychophysical techniques, it is fast and easy for subjects to understand, and sensitive to mismatch, and could therefore be used as the outcome measure for an optimization procedure.

E. Study limitations

The current study utilized vocoder simulation of SSD-CI listening. Although valuable for studying the various effects of CI processing on auditory perception, vocoder simulations are an imperfect approximation of what actual CI listeners hear ([Whitmal et al., 2007](#)). For example, duration of deafness, amount of time after implantation, listener age, and residual acoustic hearing ([Blamey et al., 2013](#)), as well as factors related to electrode placement ([Chakravorti et al., 2019](#)), can all impact CI outcomes. Although vocoder simulations allow for the reduction of intersubject variability in these dimensions and enable the study of specific aspects of CI processing on auditory perception, the simulations are far from perfect because they do not include these listener-specific elements. Here, a critical shortcoming of the vocoder approach is that actual CI listeners have time to

adapt to their mismatched frequency maps, whereas vocoder listeners do not, at least in these acute experiments. Siciliano *et al.* (2010) found that even after 10 plus hours of training with a unilateral frequency shifted vocoder, listeners received no binaural benefits, i.e., they could not learn the shifted map. This indicates a constraint on the limits of binaural plasticity, at least for NH vocoder listeners. There is some limited evidence that post-lingually deafened monaural adult CI listeners and bimodal users can adapt to this frequency mismatch between their implanted ear and their acoustic ear (Reiss *et al.*, 2014b; Svirsky *et al.*, 2004). However, this plasticity is likely to be incomplete, and these listeners would probably benefit from a more aligned spectral mapping, especially as it pertains to binaural hearing. For SSD-CI listeners, plasticity mechanisms could theoretically overcome some of the limitations imposed by spectral mismatch (Svirsky *et al.*, 2004) and achieve better fusion without remapping. However, adaptation might occur faster if interaural frequency alignment occurs right at implant activation (Svirsky *et al.*, 2015). It is also important to point out that the time course, magnitude, and factors underlying the adaptation process are likely to differ between adults, who were the focus of the current study, and children, who were not. For example, Steel *et al.* (2015) found that pediatric BI-CI listeners who were later implanted and had access to some acoustic hearing during early development were more likely to show evidence of binaural fusion than those who were early implanted and only had access to electrical stimulation.

Although the different outcomes of experiment 1 (no evidence of fusion) and experiment 2 (evidence of fusion) were interpreted in terms of the subjective versus forced-choice nature of the two experiments, another difference in the methods employed for the two experiments was that in experiment 1, trials for the two vocoder conditions were interleaved within each block, whereas for experiment 2, the matched and mismatched vocoder conditions were presented in separate blocks. It is an open question whether listeners might have perceived the matched-vocoder condition in experiment as more fused if they had an opportunity to listen to a single map for an extended period of time.

Although one interpretation of the results of experiment 2 was that the listeners were achieving partial fusion, an alternative possible explanation for the results is that listeners were independently monitoring the speech in the two ears to perform the task rather than detecting binaural coherence. Since the fusion interval contained a talker saying the same words in both ears, the listeners could have potentially been monitoring each ear for common words to complete the task. Given that the intelligibility likely decreased when the vocoded signals were mismatched, a reduction in speech intelligibility, rather than the mismatch *per se*, might have been responsible for the deterioration of performance in the task. However, this strategy would likely only be useful when the total number of voices was low, since individual words in the mixture would become more difficult to track with an increasing number of voices.

Therefore, it is much more likely the listeners were attending to interaural coherence, rather than independently monitoring the words that appears in both ears, especially in the case of five concurrent talkers where significant differences between vocoder-alignment conditions were observed.

While the current study was focused on the potential impact of a spectral mismatch on binaural fusion, there are two other important types of mismatch between an acoustic and a CI ear not addressed here that could play an important role in limiting binaural fusion: level and timing. Wess and Bernstein (2019) found that differences in loudness growth between the ears affected the ability to perceptually separate concurrent voices in a vocoder simulation. Zirn *et al.* (2015) found that SSD-CI processing introduced an interaural delay, which was ascribed to differences between the CI processing time and the traveling wave delay in an acoustic ear. By compensating for this interaural delay, Seebacher *et al.* (2019) showed that SSD-CI localization performance could be improved. It remains unknown whether this interaural delay would also result in reduced binaural fusion. Wess *et al.* (2017) found that the benefit to speech-reception performance in a competing-talker scenario provided by a single-sided vocoder was not reduced by an interaural timing mismatch until the mismatch exceeded 24 ms, which was ascribed to the long echo threshold for speech (Litovsky *et al.*, 1999). The discrimination paradigm employed in experiment 2, which was found to be sensitive to an interaural frequency mismatch, could also be useful in posing questions regarding the impact of interaural mismatch in loudness growth or timing on binaural fusion.

V. SUMMARY AND CONCLUSIONS

Two experiments examined how spectral mismatch in vocoder simulations of SSD-CI listening affected the ability of listeners to perceptually fuse the speech presented to the two ears. Experiment 1 indicated that neither a matched nor a spectrally mismatched vocoder provided sufficient fusion cues for the listener to count a given speech signal presented to both ears as a single voice. This suggests that whatever partial fusion was present, even in the spectrally matched case, was not sufficient for listeners to perceive a common speech signal in both ears as a single perceptual object. Experiment 2 was a discrimination task where listeners were required to identify the interval containing the same speech signal in the two ears among a mixture of voices. This task was found to be sensitive to interaural frequency mismatch, suggesting that partial fusion of the unprocessed and vocoded signals was enough to detect the presence of interaural coherence. Although this acute vocoder experiment did not capture the possible adaptation to mismatch that CI listeners might experience, the results nevertheless corroborate other results in the literature suggesting that for a SSD-CI listener to have the best chance of binaural hearing with their implant, steps should be taken to reduce spectral mismatch between their acoustic ear and their CI. The fact that the coherence-discrimination task in

experiment 2 was easy to understand and sensitive to mismatch makes this measurement technique a potentially useful tool to evaluate outcomes of binaural integration for SSD-CI users.

ACKNOWLEDGMENTS

Research reported was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award No. R01DC015798 (J.G.W.B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The identification of specific products or scientific instrumentation does not constitute endorsement or implied endorsement on the part of the author, DoD, or any component agency. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government. We thank Matthew Goupell for insightful suggestions in the planning of this study, Hector Galloza and Matthew Makashay for assistance with signal processing, and Sandra Gordon-Salant, Douglas Brungart, Jonathan Simon, and Matthew Goupell for their helpful comments on an earlier version of this manuscript. We thank Brian Simpson and Matt Ankrom for the recruitment and coordination of the subject panel at the Air Force Research Laboratory.

- Arndt, A., Laszig, R., Aschendorff, A., Hassepass, F., Beck, R., and Wesarg, T. (2017). "Cochlear implant treatment of patients with single-sided deafness or asymmetric hearing loss," *HNO* **65**, S98–S108.
- Aronoff, J. M., Shayman, C., Prasad, A., Suneel, D., and Stelmach, J. (2015). "Unilateral spectral and temporal compression reduces binaural fusion for normal hearing listeners with cochlear implant simulations," *Hear. Res.* **320**, 24–29.
- Bernstein, J. G. W., Goupell, M. J., Schuchman, G. I., Rivera, A. L., and Brungart, D. S. (2016). "Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees," *Ear Hear.* **37**, 289–302.
- Bernstein, J. G. W., Iyer, N., and Brungart, D. S. (2015). "Release from informational masking in a monaural competing-speech task with vocoded copies of the maskers presented contralaterally," *J. Acoust. Soc. Am.* **137**, 702–713.
- Bernstein, J. G. W., Schuchman, G. I., and Rivera, A. L. (2017). "Head shadow and binaural squelch for unilaterally deaf cochlear implantees," *Otol. Neurotol.* **38**, e195–e202.
- Bernstein, J. G. W., Stakhovskaya, O. A., Schuchman, G. I., Jensen, K. K., and Goupell, M. J. (2018). "Interaural time-difference discrimination as a measure of place of stimulation for cochlear-implant users with single-sided deafness," *Trends Hear.* **22**, 2331216518765514.
- Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P. H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., and Lazard, D. S. (2013). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients," *Audiol. Neurotol.* **18**, 36–47.
- Blauert, J., and Lindemann, W. (1986). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *J. Acoust. Soc. Am.* **79**, 806–813.
- Boersma, P. (2001). "Praat, a system for doing phonetics by computer," *Glott Int.* **5**, 341–345.
- Boothroyd, A., and Kishon-Rabin, L. (1988). "Voice fundamental frequency as an auditory supplement to speech reading of sentences," *Ear Hear.* **9**, 306–312.
- Buss, E., Dillon, M. T., Rooth, M. A., King, E. R., Deres, E. J., Buchman, C. A., Pillsbury, H. C., and Brown, K. D. (2018). "Effects of cochlear implantation on binaural hearing in adults with unilateral hearing loss," *Trends Hear.* **22**, 2331216518771173.
- Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., Baguley, D. M., Briggs, J., Deeks, J. M., Braire, J. J., Barreau, X., and Dauman, R. (2010). "Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear," *J. Assoc. Res. Otolaryngol.* **11**, 625–640.
- Carrell, T. D., and Opie, J. M. (1992). "The effect of amplitude comodulation on auditory object formation in sentence perception," *Percept. Psychophys.* **52**, 437–445.
- Chakravorti, S., Noble, J. H., Gifford, R. H., and Dawant, B. M. (2019). "Further evidence of the relationship between cochlear implant electrode positioning and hearing outcomes," *Otol. Neurotol.* **40**, 617–624.
- Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in *Hearing*, Vol. IV of *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic Press, Los Angeles), pp. 467–518.
- Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., and Dau, T. (2018). "Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners," *J. Acoust. Soc. Am.* **144**, 2896–2905.
- Culling, J. F., Jelfs, S., Talbert, A., Grange, J., and Backhouse, S. (2012). "The benefit of bilateral versus unilateral cochlear implantation to speech intelligibility in noise," *Ear Hear.* **33**, 673–682.
- Dirks, C., Nelson, P. B., Sladen, D. P., and Oxenham, A. J. (2019). "Mechanisms of localization and speech perception with colocated and spatially separated noise and speech maskers under single-sided deafness with a cochlear implant," *Ear Hear.* **40**(6), 1293–1306.
- Dorman, M. F., Zeitler, D., Cook, S. J., Loiselle, L., Yost, W. A., Wanna, G. B., and Gifford, R. H. (2015). "Interaural level difference cues determine sound source localization by single-sided deaf patients fit with a cochlear implant," *Audiol. Neurotol.* **20**, 183–188.
- Erbele, I. D., Bernstein, J. G. W., Schuchman, G. I., Brungart, D. S., and Rivera, A. (2015). "An initial experience of cochlear implantation for patients with single-sided deafness after prior osseointegrated hearing device," *Otol. Neurotol.* **36**, e24–e29.
- Firszt, J. B., Reeder, R. M., and Holden, L. K. (2017). "Unilateral hearing loss: Understanding speech recognition and localization variability—Implications for cochlear implant candidacy," *Ear Hear.* **38**, 159–173.
- Fitzgerald, M. B., Kan, A., and Goupell, M. J. (2015). "Bilateral loudness balancing and distorted spatial maps in recipients of bilateral cochlear implants," *Ear Hear.* **36**, e225–e236.
- Francart, T., Wiebe, K., and Wesarg, T. (2018). "Interaural time difference perception with a cochlear implant and a normal ear," *J. Assoc. Res. Otolaryngol.* **715**, 703–715.
- Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (2001). "Spatial and spectral factors in release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**, 2112–2122.
- Galvin, J. J., Fu, Q., Wilkinson, E. P., Mills, D., Hagan, S. C., Lupo, J. E., Padilla, M., and Shannon, R. V. (2019). "Benefits of cochlear implantation for single-sided deafness: Data from the House Clinic-University of Southern California-University of California, Los Angeles clinical trial," *Ear Hear.* **40**, 766–781.
- Goupell, M. J., Cosentino, S., Stakhovskaya, O. A., and Bernstein, J. G. W. (2019). "Interaural pitch-discrimination range effects for bilateral and single-sided-deafness cochlear-implant users," *J. Assoc. Res. Otolaryngol.* **20**, 187–203.
- Goupell, M. J., and Litovsky, R. Y. (2014). "The effect of interaural fluctuation rate on correlation change discrimination," *J. Assoc. Res. Otolaryngol.* **15**, 115–129.
- 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., Stoelb, C., Kan, A., and Litovsky, R. Y. (2013). "Effect of mismatched place-of-stimulation on the salience of binaural cues in

- conditions that simulate bilateral cochlear-implant listening,” *J. Acoust. Soc. Am.* **133**, 2272–2287.
- Hu, H., and Dietz, M. (2015). “Comparison of interaural electrode pairing methods for bilateral cochlear implants,” *Trends Hear.* **19**, 2331216515617143.
- 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.
- Landsberger, D. M., Svrakic, M., Roland, J. T., and Svirsky, M. (2015). “The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants,” *Ear Hear.* **36**, e207–e213.
- Litovsky, R., Misurelli, S., Godar, S., Thakkar, T., Kan, A., and Winn, M. (2016). “The impact of cochlear implantation on spatial hearing and listening effort,” *J. Acoust. Soc. Am.* **140**, 3100 (A).
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). “The precedence effect,” *J. Acoust. Soc. Am.* **106**, 1633–1654.
- Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., Agrawal, S., Kan, A., Todd, A., Hess, C., and Misurelli, S. (2012). “Studies on bilateral cochlear implants at the University of Wisconsin’s Binaural Hearing and Speech Laboratory,” *J. Am. Acad. Audiol.* **23**, 476–494.
- Mesnildrey, Q., Hilkuysen, G., and Macherey, O. (2016). “Pulse-spreading harmonic complex as an alternative carrier for vocoder simulations of cochlear implants,” *J. Acoust. Soc. Am.* **139**, 986–991.
- Moore, J. K. (2000). “Organization of the human superior olivary complex,” *Microsc. Res. Tech.* **51**, 403–412.
- Moulines, E., and Charpentier, F. (1990). “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,” *Speech Commun.* **9**, 453–467.
- Noble, J. H., Gifford, R. H., Hedley-Williams, A. J., Dawant, B. M., and Labadie, R. F. (2014). “Clinical evaluation of an image-guided cochlear implant programming strategy,” *Audiol. Neurotol.* **19**, 400–411.
- Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., and Kramer, S. E. (2017). “Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review,” *Ear Hear.* **38**, 267–281.
- Perreau, A. E., Wu, Y.-H., Tatge, B., Irwin, D., and Corts, D. (2017). “Listening effort measured in adults with normal hearing and cochlear implants,” *J. Am. Acad. Audiol.* **28**, 685–697.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., and Wingfield, A. (2016). “Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL),” *Ear Hear.* **37**, 5S–27S.
- Poon, B. B., Eddington, D. K., Noel, V., and Colburn, H. S. (2009). “Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing,” *J. Acoust. Soc. Am.* **126**, 806–815.
- Reiss, L. A. J., Ito, R. A., Eggleston, J. L., and Wozny, D. R. (2014a). “Abnormal binaural spectral integration in cochlear implant users,” *J. Assoc. Res. Otolaryngol.* **15**, 235–248.
- Reiss, L. A. J., Turner, C. W., Karsten, S. A., and Gantz, B. J. (2014b). “Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation,” *Neuroscience* **256**, 43–52.
- Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., and Farahbod, H. (1998). “Effects of interaural decorrelation on neural and behavioral detection of spatial cues,” *Neuron* **21**, 789–798.
- Sayers, B. M., and Cherry, E. C. (1957). “Mechanism of binaural fusion in the hearing of speech,” *J. Acoust. Soc. Am.* **29**, 973–987.
- Schatzer, R., Vermeire, K., Visser, D., Krenmayr, A., Kals, M., Voormolen, M., Van de Heyning, P., and Zierhofer, C. (2014). “Electric-acoustic pitch comparisons in single-sided-deaf cochlear implant users: Frequency-place functions and rate pitch,” *Hear. Res.* **309**, 26–35.
- Seebacher, J., Franke-Triege, A., Weichbold, V., Zorowka, P., and Stephan, K. (2019). “Improved interaural timing of acoustic nerve stimulation affects sound localization in single-sided deaf cochlear implant listeners,” *Hear. Res.* **371**, 19–27.
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). “Temporal coherence and attention in auditory scene analysis,” *Trends Neurosci.* **34**, 114–123.
- Siciliano, C. M., Faulkner, A., Rosen, S., and Mair, K. (2010). “Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants,” *J. Acoust. Soc. Am.* **127**, 1645–1660.
- Sladen, D. P., Nie, Y., and Berg, K. (2018). “Investigating speech recognition and listening effort with different device configurations in adult cochlear implant users,” *Cochlear Implants Int.* **19**, 119–130.
- Stakhovskaya, O., Sridhar, D., Bonham, B. H., and Leake, P. A. (2007). “Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants,” *J. Assoc. Res. Otolaryngol.* **8**, 220–233.
- Stakhovskaya, O. A., and Goupell, M. J. (2017). “Lateralization of interaural level differences with multiple electrode stimulation in bilateral cochlear-implant listeners,” *Ear Hear.* **38**, e22–e38.
- Steel, M. M., Papsin, B. C., and Gordon, K. A. (2015). “Binaural fusion and listening effort in children who use bilateral cochlear implants: A psychoacoustic and pupillometric study,” *PLoS One* **10**, e0141945.
- Suneel, D., Staisloff, H., Shayman, C. S., Stelmach, J., and Aronoff, J. M. (2017). “Localization performance correlates with binaural fusion for interaurally mismatched vocoded speech,” *J. Acoust. Soc. Am.* **142**, EL276–EL280.
- Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., and Suárez, H. (2004). “Long-term auditory adaptation to a modified peripheral frequency map,” *Acta Otolaryngol.* **124**, 381–386.
- Svirsky, M. A., Talavage, T. M., Sinha, S., Neuburger, H., and Azadpour, M. (2015). “Gradual adaptation to auditory frequency mismatch,” *Hear. Res.* **322**, 163–170.
- Vermeire, K., and Van de Heyning, P. (2009). “Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus,” *Audiol. Neurotol.* **14**, 163–171.
- Wess, J. M., and Bernstein, J. G. W. (2019). “The effect of nonlinear amplitude growth on the speech perception benefits provided by a single-sided vocoder,” *J. Speech Lang. Hear. Res.* **62**, 745–757.
- Wess, J. M., Brungart, D. S., and Bernstein, J. G. W. (2017). “The effect of interaural mismatches on contralateral unmasking with single-sided vocoders,” *Ear Hear.* **38**, 374–386.
- Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (2007). “Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience,” *J. Acoust. Soc. Am.* **122**, 2376–2388.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2012). “Apparent auditory source width insensitivity in older hearing-impaired individuals,” *J. Acoust. Soc. Am.* **132**, 369–379.
- Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). “The impact of auditory spectral resolution on listening effort revealed by pupil dilation,” *Ear Hear.* **36**, e153–e165.
- Zhong, X., and Yost, W. A. (2017). “How many images are in an auditory scene?,” *J. Acoust. Soc. Am.* **141**, 2882–2892.
- Zirn, S., Arndt, S., Aschendorff, A., and Wesarg, T. (2015). “Interaural stimulation timing in single sided deaf cochlear implant users,” *Hear. Res.* **328**, 148–156.
- Zurek, P. (1993). “Binaural advantages and directional effects in speech intelligibility,” in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston, MA), pp. 266–276.