



# Dichotic listening performance with cochlear-implant simulations of ear asymmetry is consistent with difficulty ignoring clearer speech

Matthew J. Goupell<sup>1</sup> · Daniel Eisenberg<sup>1</sup> · Kristina DeRoy Milvae<sup>1</sup>

Accepted: 7 January 2021 / Published online: 29 March 2021  
© The Psychonomic Society, Inc. 2021

## Abstract

There are an increasing number of bilateral and single-sided-deafness cochlear-implant (CI) users who hope to achieve improved spatial-hearing abilities through access to sound in both ears. It is, however, unclear how speech is processed when inputs are functionally asymmetrical, which may have an impact on spatial-hearing abilities. Therefore, functionally asymmetrical hearing was controlled and parametrically manipulated using a channel vocoder as a CI simulation. In Experiment 1, normal-hearing (NH) listeners performed a dichotic listening task (i.e., selective attention to one ear, ignoring the other) using asymmetrical signal degradation. Spectral resolution varied independently in each ear (4, 8, 16 channels, and unprocessed control). Performance decreased with decreasing resolution in the target ear and increasing resolution in the interferer ear. In Experiment 2, these results were replicated using a divided attention task (attend to both ears, report one after sentence completion) in both NH and bilateral CI listeners, although overall performance was lower than in Experiment 1. In Experiment 3, frequency-to-place mismatch simulated shallow CI insertion depths (0, 3, 6-mm shifts, and unprocessed control). Performance mostly decreased with increasing shift in the target ear and decreasing shift in the interferer ear; however, performance nonmonotonicities occurred. The worst performance occurred when the shift matched across ears, suggesting that pitch similarity increases difficulty. The results show that it is more difficult to attend an ear that is relatively degraded or distorted, which may set spatial-hearing limitations for CI users when trying to attend to a target in complex auditory scenes.

**Keywords** Selective attention · Divided attention · Vocoding · Cochlear implants

Having two ears and two eyes allows humans to precisely encode the spatial location of sound sources and visual objects, respectively. The inputs in a typical sensory system are ideally symmetrical. When asymmetry occurs, such as in cases of asymmetrical hearing loss, there is a vast array of possible interventions that can be provided (e.g., hearing aids and bionic auditory prostheses called cochlear implants [CIs]). Bilateral hearing aids are now commonly dispensed across the life span (Kochkin, 2009). It is recommended that children who are born deaf receive bilateral CIs around 1 year of age and with less than 18 months between implantations so that there is some hope of developing useable binaural hearing pathways and spatial-hearing abilities (Litovsky & Gordon, 2016). The effectiveness of these interventions varies greatly

depending on numerous factors (Knudsen et al., 2010; Litovsky et al., 2012). It is unclear, however, how asymmetrical auditory inputs can diminish hearing abilities and how development contributes to these processing problems (Gordon et al., 2015; Kral et al., 2013; Tillein et al., 2016). The purpose of this study is to better understand the effect of functional asymmetry in absence of developmental problems that can occur with clinical populations.

Functionally asymmetrical sound inputs appear to diminish binaural hearing abilities. Binaural hearing is critical for sound localization in the horizontal plane (Wightman & Kistler, 1997) and speech understanding in background noise (Zurek, 1992). Adults that were born with acoustic hearing, lost their hearing as adults, and received bilateral CIs demonstrated improved speech understanding when targets and interferers originated from different spatial locations (they experienced spatial release from masking; Bernstein et al., 2016). Those listeners were argued to be high performers with relatively symmetrical inputs, and bilateral CI listeners with more asymmetrical hearing appeared not to experience the spatial release from masking benefit. Asymmetry partially

---

✉ Matthew J. Goupell  
goupell@umd.edu

<sup>1</sup> Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, USA

explains the discrepancy between the Bernstein et al. (2016) study and another study by Goupell, Stakhovskaya, et al. (2018) that used the same speech-on-speech masking tests; the bilateral CI listeners in the latter study experienced interference instead of an unmasking benefit. This latter study also included a subset of listeners with earlier onsets of hearing loss, so developmental effects may have also contributed to the observed interference.

To simplify the speech-on-speech spatial release from masking task to a more controlled scenario, dichotic listening can be used, in which different speech samples are presented to opposite ears. This paradigm eliminates energetic masking (energy from target and interferer are not analyzed in the same auditory filters in the cochlea or peripheral auditory neurons; e.g., Best et al., 2013); the target is presented to one ear and the interferer is presented to the other. Such a situation sets up separate streams of information for each ear. Dichotic listening, where a person selectively attends to one ear and ignores the other, is often thought to be a relatively simple task (Brungart & Simpson, 2002; Cherry, 1953; Gallun et al., 2007b; Goupell et al., 2016; Wood & Cowan, 1995a). A human's ability to selectively attend to a single ear is so good that it is unlikely that one would hear his or her own name in a stream of speech presented to the unattended ear (Wood & Cowan, 1995b). Stimulus information and task complexity, however, play a role in the typically excellent performance in dichotic listening. Interference from the nontarget ear becomes increasingly common as task demands increase by adding additional interfering sound streams or altering the similarity of target and interferer streams, which is possibly a result of how the listener expends a finite amount of resources to undertake the task (Brungart & Simpson, 2002, 2004, 2007; Gallun et al., 2007a, 2007b). Such an explanation is consistent with the relative difficulty of dichotic listening seen in children compared with adults (Wightman et al., 2003; Wightman & Kistler, 2005; Wightman et al., 2006; Wightman et al., 2010) because children may have fewer resources and/or they allocate them inefficiently (Lutfi et al., 2003; Wightman & Kistler, 2005).

The ease of dichotic listening and selective attention to a single ear also occurs in most bilateral CI listeners. Goupell et al. (2016) tested 11 adult bilateral CI listeners and found nine listeners had relatively symmetric abilities in attending to the right or left ear. The other two listeners could easily attend to their right ear, but demonstrated great difficulty attending to their left ear despite explicit instruction. The inability to easily attend to the left ear was also found in two of 10 children with bilateral CIs—those who had the longest interimplant durations (Misurelli et al., 2020). Other bilateral CI listeners in a separate study (Goupell, Stakhovskaya, et al., 2018) had a similar difficulty perceiving sound presented to one of their ears with dichotic presentation, but this did not occur with monaural presentation. This subset of results in bilateral CI

listeners appears to have some similarities to the loss of binocular vision caused by the amblyopia phenomenon (Barrett et al., 2004; Kaplan et al., 2016; Whitton & Polley, 2011), as one explanation for the data was that these few listeners were unable to attend to an ear because its input was not perceived under dichotic presentation.

While there is mounting evidence for difficulty in understanding speech with asymmetrical inputs, particularly for accessing spatial-hearing benefits, the heterogeneity of the CI population (differences in performance due to biological, surgical, and device-related factors; Litovsky et al., 2012) makes it difficult to separate effects due to functionally asymmetrical performance from developmental problems like an amblyopia-like effect (i.e., amblyaudia; see Kaplan et al., 2016, for a review). Therefore, the purpose of this study was to determine whether there are consequences to auditory processing of speech through asymmetrical inputs in the absence of developmental issues, which are often difficult to assess (Whitton & Polley, 2011). To avoid the variability that occurs in clinical populations, we produced asymmetrical inputs through a signal processing technique called “channel vocoding” (Dudley, 1939; Shannon et al., 1995), degrading the signals parametrically and independently in each ear. We hypothesized that it would be more difficult to attend to a relatively poorer ear and ignore a relatively better ear. We simulated two forms of hearing asymmetry with a CI in this study. In Experiments 1 and 2, we simulated differential spectral resolution, which is related to physical placement of the electrodes, how electrical fields interact and overlap in the cochlea, and neural degeneration (e.g., Croghan et al., 2017; Friesen et al., 2001). In Experiment 3, we simulated differential electrode array insertion depths across the ears to produce a frequency-to-place mismatch or “shift” of frequency information. These conditions simulate some of the asymmetrical ear differences that could occur in the CI population.

## Experiment 1: Selective attention and asymmetrical spectral resolution

### Listeners and equipment

Ten normal-hearing (NH) listeners between the ages of 20 and 35 years were tested in this experiment (mean age = 24.7 years), a sample size based on previous similar work using vocoded speech with this type of experiment (Goupell et al., 2016). The listeners had typical hearing thresholds ( $\leq 20$  dB hearing level [HL] air conduction thresholds) at 0.25, 0.5, 1, 2, 4, and 8 kHz, measured with an audiometer (Maico, MA41; Berlin, Germany). In addition, listeners were screened for asymmetrical hearing such that no listener had an interaural difference in threshold at any tested frequency of  $>10$  dB.

Seven listeners had previous exposure to vocoded speech, and all were native English speakers.

Stimuli were created on a personal computer using MATLAB (The MathWorks; Natick, MA). Listeners were presented with the stimuli via open-backed circumaural headphones (Sennheiser, HD650; Hanover, Germany). All testing was performed in a double-walled sound-attenuating booth (IAC; Bronx, NY) at the University of Maryland, College Park.

**Stimuli**

Stimuli were nonsense sentences with five keywords, each consisting of a name, verb, number, adjective, and object, and were the original recordings from Kidd et al. (2008). Each category had a closed set with eight possibilities per keyword, which are shown in Table 1. Words within each category had a range of similarity and confusability (e.g., “old” vs. “cold”).

The keywords were randomly chosen with replacement for each ear; the same word could occur in both ears, which limits using a strategy of eliminating target word possibilities by identifying the interferer word (e.g., Bernstein et al., 2016). The individual words had different lengths and onset times. This occurred for even the same word because two different talkers produced the target and interferer words. Therefore, the words were not time aligned across the ears.

The talkers in the left and right ears were female (Talker 10, F0 = 209.5 Hz) and male (Talker 2, F0 = 109.8 Hz), respectively, similar to the stimuli used in Goupell et al. (2016). The stimuli were presented at an A-weighted sound pressure level of 65 dB.

Stimuli were either unprocessed or processed using a noise vocoder. The vocoding process included an analysis stage in which the unprocessed stimuli were passed through a filter bank with 4, 8, or 16 contiguous channels. The bandpass filters in the filter bank were fourth-order Butterworth filters. The corner frequencies on the bandpass filters were logarithmically spaced and covered a frequency range from 300 to 8500 Hz (see Table 2). The envelope from each channel was extracted using a second-order low-pass filter with a 400-Hz

**Table 2** Corner frequencies for the bandpass filters, arithmetic center frequencies (CFs), and bandwidths (BW) for the channels used in the vocoder for Experiment 1

Channels	Corner frequency (Hz)		CF (Hz)	BW (Hz)
	Lower	Upper		
4	300.0	692.1	496.1	392.1
	692.1	1596.9	1144.5	904.7
	1596.9	3684.2	2640.5	2087.3
	3684.2	8500.0	6092.1	4815.8
8	300.0	455.7	377.8	155.7
	455.7	692.1	573.9	236.5
	692.1	1051.3	871.7	359.2
	1051.3	1596.9	1324.1	545.6
	1596.9	2425.5	2011.2	828.7
	2425.5	3684.2	3054.9	1258.7
	3684.2	5596.1	4640.1	1911.8
	5596.1	8500.0	7048.0	2903.9
16	300.0	369.7	334.9	69.7
	369.7	455.7	412.7	85.9
	455.7	561.6	508.6	105.9
	561.6	692.1	626.9	130.5
	692.1	853.0	772.6	160.9
	853.0	1051.3	952.2	198.3
	1051.3	1295.7	1173.5	244.4
	1295.7	1596.9	1446.3	301.2
	1596.9	1968.1	1782.5	371.2
	1968.1	2425.5	2196.8	457.5
	2425.5	2989.3	2707.4	563.8
	2989.3	3684.2	3336.8	694.9
3684.2	4540.6	4112.4	856.4	
4540.6	5596.1	5068.3	1055.5	
5596.1	6896.8	6246.4	1300.8	
6896.8	8500.0	7698.4	1603.2	

**Table 1** Matrix word corpus

Name	Verb	Number	Adjective	Object
Jane	took	two	new	toys
Gene	gave	three	old	hats
Pat	lost	four	big	shoes
Bob	found	five	small	cards
Sue	bought	six	red	pens
Mike	sold	eight	blue	socks
Lynn	held	nine	cold	bags
Jill	saw	ten	hot	gloves

cutoff frequency. The envelopes were used to modulate narrowband noise carriers after the carriers were filtered with the same bandpass analysis filters. The vocoded stimuli were synthesized by summing the channels into the acoustic waveform and were normalized to have the same root-mean-square energy as the unprocessed stimuli.

Note that changing the number of channels from 4 to 8 to 16 channels changes more than just spectral resolution. With 4 channels, there is likely to be much greater mismatch between the input and output frequencies than with 16 channels (e.g., formant information is smeared relatively more with fewer channels). In addition, the interaurally uncorrelated narrowband noise carriers add random fluctuations to the temporal envelopes of each channel, and the rate of those fluctuations are dependent on the narrowband noise bandwidth (e.g., Goupell & Litovsky, 2014).

## Procedure

Listeners were seated at a computer, and the experiment was performed with a graphical user interface that was programmed in MATLAB (The MathWorks; Natick, MA). The listener initiated each trial by pressing a button on the computer interface. Stimuli were presented dichotically over headphones, meaning different sentences were presented to each ear simultaneously and spoken by a different talker. The instructions to the listeners were to report the words in the left ear and ignore the words in the right ear. They responded by selecting one of eight possible keyword choices for each category on a grid. Another button on the screen was then pressed to confirm the selections and end the trial. Listeners were forced to guess if they did not know one of the words in the sentence. The words presented in each ear were chosen at random and, in some cases, the same word was presented to both ears.

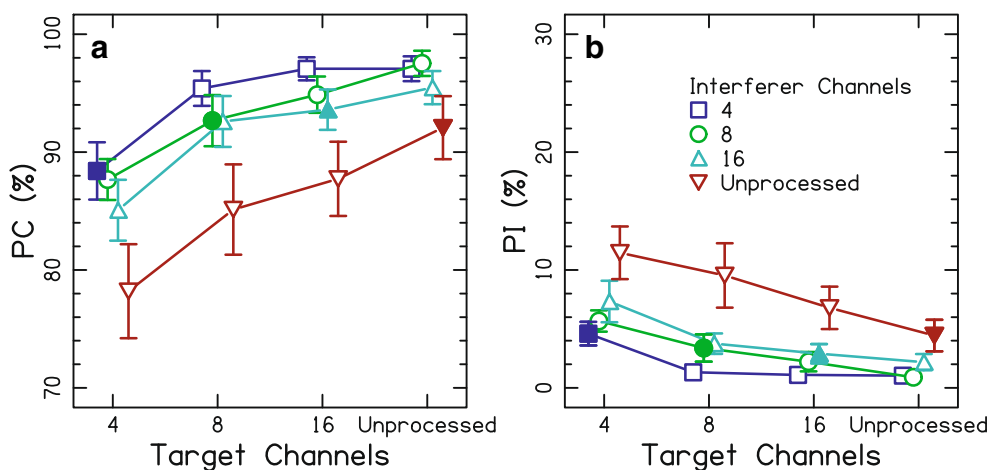
The number of channels was independently varied in each ear. Listeners were presented with 4 target ear  $\times$  4 interferer ear = 16 combinations of resolution levels (4, 8, 16 channels, and unprocessed control). A monaural control measurement was not performed because performance would have likely been close to 100% correct for most listeners and most conditions, at least for 8 channels or more (Waked et al., 2017); the strong ceiling effect would have limited its usefulness to interpret the amount of interference caused by the contralaterally presented interferer. The experiment was performed using a method of constant stimuli where all 16 combinations of stimuli were presented 10 times in a randomized order per block. Listeners performed three blocks of 160 trials for a total of 480 trials. Therefore, there were 150 total keywords per condition (5 keywords  $\times$  10 trials  $\times$  3 blocks). Testing for this experiment took approximately 1.5 hours.

## Data analysis

The percentage of correct responses (PC) and percentage of across-ear confusions or intrusions (PI; i.e., reporting the word presented to the nontarget ear) were calculated (Brungart et al., 2001; Bryden et al., 1983). When the target and interferer were the same word, the response was counted as correct and not an intrusion. PC and PI scores were transformed to rationalized arcsine units (RAUs; Studebaker, 1985) to better follow the assumption of homogeneity of variance required for an analysis of variance (ANOVA). The data were analyzed using a two-way repeated-measures ANOVA, with factors target-ear channels (four levels: 4, 8, 16 channels, and unprocessed control) and interferer-ear channels (four levels: 4, 8, 16 channels, and unprocessed control). In cases where the assumption of sphericity was violated, a Greenhouse-Geisser correction was used. Bonferroni-corrected two-sample two-tailed paired  $t$  tests were used for post hoc comparisons.

## Results

Figure 1a shows the PC data. Performance was near ceiling for 8 and 16 channels, as well as unprocessed speech, in the target ear. As the number of channels in the target ear increased, performance increased,  $F(1.5, 13.8) = 29.0, p < .0001, \eta_p^2 = 0.76$ . Post hoc tests showed that all conditions were different than the others ( $p < .05$  for all six comparisons). As the number of channels in the interferer ear increased, performance decreased,  $F(3, 27) = 97.6, p < .0001, \eta_p^2 = 0.92$ . Post hoc tests showed that all conditions were different than the others ( $p < .001$  for all six comparisons). There was no significant interaction between target-ear channels and interferer-ear channels,  $F(4.2, 37.9) = 1.9, p = .13, \eta_p^2 = 0.18$ . Finally, the matched conditions (i.e., same number of channels in both



**Fig. 1** Average percentage of correct responses (PC; **a**) and intrusions (PI; **b**) as a function of the channels in the target ear for the selective attention task of Experiment 1. Filled symbols highlight the conditions

where the channels are equal in the target and interferer ears. Error bars represent  $\pm 1$  standard error. Note that the scale on the y-axis changes from **a** to **b**, but the relative size of the scale does not change



ears; filled symbols of Fig. 1a) were compared in post hoc tests. The matched 4-channel condition was lower than the other three conditions ( $p < .05$  for all three comparisons), but no other comparisons were significantly different ( $p > .05$ ).

Figure 1b shows the PI data. As the number of channels in the target ear increased, intrusions decreased,  $F(3, 27) = 75.4$ ,  $p < .0001$ ,  $\eta_p^2 = 0.89$ . Post hoc tests showed that all comparisons were significantly different ( $p < .05$  for all). As the number of channels in the interferer ear increased, intrusions increased,  $F(3, 27) = 31.9$ ,  $p < .0001$ ,  $\eta_p^2 = 0.78$ . Post hoc tests showed that all comparisons were significantly different ( $p < .05$  for all). There was no significant interaction between target-ear channels and interferer-ear channels,  $F(9, 81) = 1.47$ ,  $p = .17$ ,  $\eta_p^2 = 0.14$ . Finally, the matched conditions (filled symbols of Fig. 1b) were compared in post hoc tests. The matched 4-channel condition was higher than the 8-channel condition ( $p = .016$ ) and the 16-channel condition ( $p = .042$ ). There were no differences for the other comparisons ( $p > .05$  for all four comparisons).

## Discussion

Selective attention to a single ear for dichotically presented sentences was affected by the spectral resolution in both target and interferer ears. Figure 1a shows that increased spectral resolution produced increased speech understanding for the target ear, consistent with other studies investigating spectral resolution with vocoders (e.g., Friesen et al., 2001). The key finding of this experiment was that the spectral resolution of the *interfering* ear also affected speech understanding in the target ear, where increasing spectral resolution in the interferer ear produced decreased speech understanding for the target ear. The highest performance was not for the unprocessed conditions for each ear; rather, highest performance was achieved for the 4-channel interferer conditions. This suggests that the signals in the interfering ear are being processed, and it takes some resources to ignore them.

The results are in line with those found for NH listeners attending to an unprocessed target with a same-sex interferer in the same ear, and trying to ignore a noise-vocoded interferer in the other ear with the number of channels parametrically varied (Brungart et al., 2005). These data showed that the intelligibility of the interfering ear is critical; the more intelligible the interferer speech, the more difficult it was to selectively attend to the target ear. Similar results concerning the effect of interferer intelligibility occurred for vocoded stimuli with matched spectral resolution, but nonoverlapping frequency bands across the ears when there was a less intense but more intelligible interferer (Gallun et al., 2007a) or in the presence of masking noise (Kidd et al., 2005). Together, these data are evidence that listeners had difficulty attending to the ear with poorer resolution and speech

understanding compared with the ear with better resolution and speech understanding. It may be that it is the speech understanding performance, rather than the resolution of the stimuli, that drives the size of the effect (Dai et al., 2017).

Figure 1b provides further insight on why performance decreased as the target ear becomes more degraded than the interfering ear. PI decreased as the target-ear resolution increased. The largest PI occurred when the interfering ear was unprocessed. This further supports the interpretation of the PC data that the listeners were having difficulty ignoring the ear with the signal that is easier to understand because they more often reported the words from the interferer ear. The average intrusion rate was 43.7% (normalized over the number of incorrect responses), which is much lower than the intrusion rate of about 90% seen in several other studies (Arbogast et al., 2002; Brungart & Simpson, 2002; Brungart et al., 2001; Kidd et al., 2005) and closer to the intrusion rate of less than 65% seen in Gallun et al. (2007a).

## Experiment 2: Divided attention in NH and CI listeners

Experiment 1 showed many PC scores near ceiling performance. To mitigate possible ceiling effects, we performed a divided attention task, which is much more demanding for listeners because they attend to the words presented to both ears. Both NH and CI listeners were tested because there is no data on this task yet in the CI population.

## Methods

The same 10 NH listeners were tested as in Experiment 1, and the same stimuli were used. The procedure remained the same, except that a divided attention task was used rather than a selective attention task. In the divided attention task, listeners were to attend to both ears. After stimulus presentation, they were asked to report the words presented to just the left or right ear, the ear randomly chosen in each trial with 50% a priori probability (Gallun et al., 2007b).

In addition, seven postlingually deafened bilateral CI listeners performed this task. They were ages 44–73 years (mean age = 59.0 years). They all used Cochlear Ltd. devices (Cochlear Ltd.; Sydney, Australia). CI listener information is located in Table 3. They were presented stimuli to their clinically fit everyday sound processors set to their most commonly used program. Stimuli were presented over Freedom TV/HiFi cables (Cochlear Ltd.; Sydney, Australia) to the direct audio input, similar to the methods used in Goupell et al. (2016) and Misurelli et al. (2020). The stimuli were nominally presented at 65 dB-A. A loudness adjustment procedure was performed before the testing to ensure that stimuli were

**Table 3** CI listener demographics in Experiment 2

Code	Age (yr)	Left ear		Right ear	
		Onset of deafness (yr)	Duration of CI use (yr)	Onset of deafness (yr)	Duration of CI use (yr)
S1	73	60	3	60	9
S2	61	51	4	51	3
S3	63	55	5	47	7
S4	44	39	1	39	0.5
S5	50	43	1	43	5
S6	55	48	5	48	0.5
S7	67	n/a	n/a	n/a	n/a

Note. Listener S7 was not able to provide this information

presented at a comfortable loudness in each ear separately, and also that the loudness was balanced across ears.

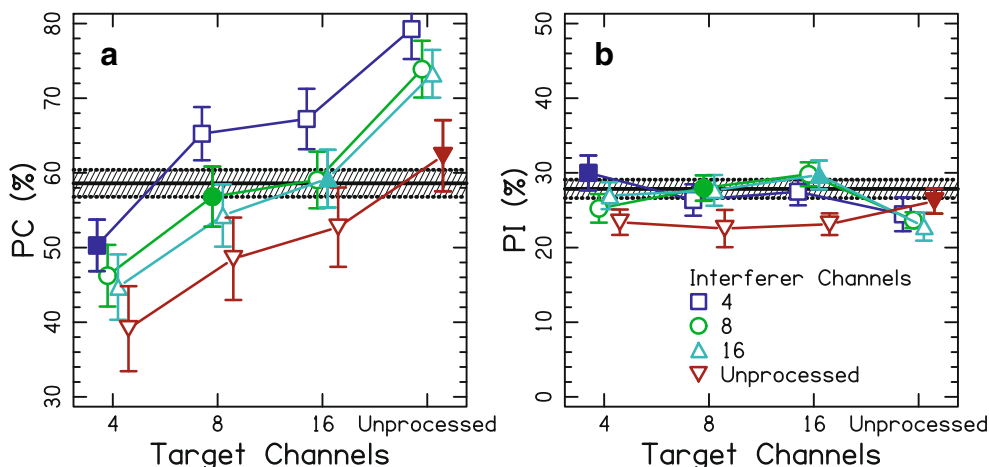
## Results

For the NH listeners, the average PC scores decreased from approximately 91.3% in Experiment 1 to 58.3% in Experiment 2, but the pattern of results across experiments was similar. Performance was symmetrical across the ears, so the data were averaged over the different right (average PC = 58.5%) and left (average PC = 58.0%) target ears. Similar to Experiment 1, separate two-way repeated-measures ANOVAs were performed on RAU-transformed PC and PI values, with factors target ear and interferer ear. Post hoc tests and corrections to statistical calculations were approached in the same way.

Figure 2a shows that as the number of channels in the target ear increased, performance increased,  $F(1.3, 11.7) = 51.6, p < .0001, \eta_p^2 = 0.85$ . Post hoc tests showed that all conditions were different than the others ( $p < .05$  for all six comparisons). As the number of channels in the interferer ear increased,

performance decreased,  $F(1.18, 10.6) = 17.1, p < .0001, \eta_p^2 = 0.66$ . Post hoc tests showed that all conditions were different than the others ( $p < .0005$  for five comparisons), except that the 8-channel interferer condition was not different than the 16-channel interferer condition ( $p > .05$ ). There was no significant interaction between target-ear resolution and interferer-ear resolution,  $F(9, 81) = 1.37, p = .22, \eta_p^2 = 0.13$ .

Figure 2b shows that as the number of channels in the target ear increased, intrusions decreased,  $F(3, 27) = 4.45, p = .012, \eta_p^2 = 0.33$ . Post hoc tests showed that the unprocessed condition was lower than 16 channels ( $p = .031$ ); no other differences were significant ( $p > .05$  for all). As the number of channels in the interferer ear increased, intrusions did not change,  $F(3, 27) = 2.62, p = .071, \eta_p^2 = 0.23$ . There was a significant interaction between target-ear resolution and interferer-ear resolution,  $F(9, 81) = 3.11, p = .003, \eta_p^2 = 0.26$ . The interaction primarily occurred because the unprocessed interferer conditions had a trend of the same or relatively fewer number of intrusions when there was 4, 8, or 16 channels in the target ear, but had relatively more intrusions



**Fig. 2** Average PC (a) and PI (b) as a function of the channels in the target ear for the divided attention task of Experiment 2. The solid line shows the average CI performance and the shaded box shows  $\pm 1$  standard

error. Note that the scale on the y axis changes from a to b, and differ from those used in Fig. 1. Otherwise, conventions are the same as in Fig. 1

when the target ear was unprocessed. Post hoc testing (Bonferroni corrected for 120 comparisons) revealed no statistically significant pairs ( $p > .05$  for all).

Data for the NH listeners can be compared with the CI listeners in this study. Generally, the CI data (horizontal lines and shaded area) best correspond to the symmetric vocoding conditions (8 or 16 channels; filled symbols in Fig. 2).

## Discussion

Experiment 2 showed a substantial drop in overall performance as would be expected from the sharing of resources to process both speech streams (Gallun et al., 2007b), and yet still confirmed the effects in Experiment 1. Both Experiments 1 and 2 (Figs. 1a and 2a, respectively) showed that performance increased with increasing number of channels in the target ear and decreased with increasing number of channels in the interferer ear. The intrusions for this experiment revealed a different pattern than in Experiment 1, where small differences occurred between conditions, and the fewest overall intrusions occurred for the unprocessed target (see Fig. 2b). The reason for the different pattern was likely a result of listeners attempting to process and remember the stimuli in both ears simultaneously. Also, a new pattern emerged in this experiment in that the cases with matched number of channels generally scored near the highest intrusions, and there was a significant interaction (see Fig. 2b). One interpretation of this is that it was more common to report the word in the incorrect ear if the stimuli sounded similar. Finally, there was no evidence of a right-ear bias in selective attention or advantage for targets in that ear. This is contrary to Gallun et al. (2007b), who found a greater number of intrusions from the right ear during divided listening. The difference in findings may have been a result of the differences in the stimuli, where Gallun et al. (2007b) used five-channel tone-vocoded stimuli in background noise, the carrier frequencies were randomly chosen and different across the ears, and the speech materials were the call-response-measure sentences (Bolia et al., 2000). Alternatively, it could have been a result of practice attending to the left ear from Experiment 1, since the same listeners participated in both experiments.

Data from adult bilateral CI listeners were also collected for comparison. They showed no relative difficulty in performing the divided attention task compared with the NH listeners. This is consistent with the similar performance between these groups on the selective attention task performed in Goupell et al. (2016). The CI listeners had average performance similar to the NH listeners at the matched channel conditions (8 or 16 channels; see filled symbols in Fig. 2a). Comparison of the demographic information in Table 3 shows that for six listeners, durations of CI use were less than 6 years different across the ears, and the duration of deafness was the same except for S3. Therefore, these CI listeners could be argued

to be a relatively symmetric group (e.g., compare against the CI listeners in Goupell, Stakhovskaya, et al., 2018). No formal statistical comparison was performed because of the age confound (young NH listeners compared with middle-aged and older CI listeners). Dichotic listening becomes more difficult with age (Humes et al., 2006), and aging effects in binaural tasks have been recently reported (Bernstein et al., 2020). Note that asking our listeners to remember two sentences each with five keywords each put a massive demand on their short-term memory (see Baddeley et al., 2015, for a review). Demands on short-term memory likely exacerbate the age confound between our NH and CI listeners. Future work could consider using fewer keywords and providing age-matched controls to this task (Cleary et al., 2018).

Finally, it should be noted that symmetry in the CI listeners cannot fully be assumed. Monaural control measurements should be added using a relatively difficult open-set speech corpus. Doing so would allow us to evaluate the effect of asymmetry in the CI listeners (Bernstein et al., 2020; Goupell, Stakhovskaya, et al., 2018).

## Experiment 3: Asymmetrical spectral shift

Speech degradation through a vocoder can occur with a number of different signal manipulations (Shannon et al., 1998). One of the most detrimental signal manipulations is to introduce frequency-to-place mismatch (Dorman et al., 1997; Rosen et al., 1999), which simulates shallow CI insertion depths (Landsberger et al., 2015). Therefore, we hypothesized that this type of degradation would produce similar asymmetries in the ability to selectively attend to an ear.

### Listeners and equipment

Five NH listeners between the ages of 19 and 38 years were tested (mean age = 24.4 years). Similar to Experiment 1, they had hearing thresholds  $\leq 20$  dB HL between 250 and 8000 Hz and  $\leq 10$  dB interaural asymmetry in thresholds. Three listeners had previous exposure to vocoded speech.

### Stimuli

As in previous experiments, the stimuli were either unprocessed or vocoded. However, several aspects of the vocoder differed from the previous two experiments to optimize testing with simulated frequency-to-place mismatch, which we will simply call “shift.” Instead of changing the number of channels used across conditions, this variable was held constant at 8 channels. The corner frequencies on the bandpass analysis filter bank were still logarithmically spaced, but were changed to cover a frequency range from 200 to 5000 Hz, which was lower in frequency than in Experiments 1 and 2 and more similar to

the range used in Rosen et al. (1999). The rationale behind this change was to avoid shifting carriers to frequencies where hearing status was unknown (e.g., well above 8 kHz).

The envelope was extracted using a second-order low-pass filter with a 50-Hz cutoff frequency, lower than that used in Experiments 1 and 2. The rationale behind the lower cutoff frequency was to avoid the sidebands of the modulated carriers (i.e., the additional spectral components added by introducing modulations) from falling into separate auditory filters than the carrier (i.e., they are resolved), which is known to greatly improve performance (Souza & Rosen, 2009). For the stimuli used in this experiment, depending on the envelope cutoff frequency, the unshifted conditions could have had resolved sidebands and the shifted conditions could have had mostly unresolved sidebands, thus introducing a confound that we wanted to avoid.

The envelopes were used to modulate tonal carriers instead of noise bands. The rationale behind this change was to avoid the interaural decorrelation of the carriers (Goupell et al., 2013; Goupell, Stoelb, et al., 2018), which would be perceived as diffuse in the head when the carriers overlapped in frequency across the ears (Whitmer et al., 2012). The frequencies of the tonal carriers were either at the center frequency of the channel or spectrally shifted by 3 or 6 mm to simulate a shallow insertion depth (see Table 4). The shift was implemented by starting with the carrier frequencies in Hz, converting them to a distance in cochlear location in mm using the frequency-to-place conversion equation (Greenwood, 1990), shifting them by a certain number of mm, and then converting them back to Hz.

A loudness correction was performed to diminish any differences in performance across the conditions based on loudness or audibility, particularly those shifted to much higher frequencies (Faulkner et al., 2003; Waked et al., 2017). The frequency-specific loudness compensation adjusted the level in dB by 50% between the threshold for the unshifted and shifted carrier frequencies, and the threshold was based on the minimum audible field (MAF) curve. For example, Channel 7 with a CF = 2734.4 Hz is shifted by 3 mm to a CF = 4213.4 Hz, and this band is amplified by 0.6 dB (see Table 4).

As before, the vocoded stimuli were synthesized by summing the channels into the acoustic waveform and were normalized to have the same root-mean-square energy as the unprocessed stimuli. Stimuli were calibrated to have an A-weighted sound pressure level of 65 dB for the unshifted conditions. The shifted conditions had the same root-mean-square energy before shifting and loudness correction.

Note that changing the mismatch between the input and output stimuli is dictated by the specific vocoder parameters. Since this experiment used tonal carriers, the envelope of each vocoder channel is relatively well represented and lacks the random envelope fluctuations of the narrowband noise carriers from Experiments 1 and 2.

## Procedure

Listeners completed two types of blocks of trials, alternating between testing and training blocks. Training is necessary to

**Table 4** Conversion of frequency (in Hz) to place (in mm) for the shifted conditions of Experiment 3

Channel	Initial CF (Hz)	Initial CF (mm)	MAF (dB SPL)	Shift (mm)	Final CF (mm)	Final CF (Hz)	MAF (dB SPL)	Level Change/2 (dB)
1	244.6	6.2	11	3	9.2	444.9	7	-2.0
2	365.7	8.2	4.5	3	11.2	628.3	2.1	-1.2
3	546.9	10.4	2.5	3	13.4	902.5	1.75	-0.4
4	817.8	12.8	1.75	3	15.8	1312.5	2.1	0.2
5	1222.8	15.3	2.1	3	18.3	1925.6	-0.5	-1.3
6	1828.6	17.9	0.5	3	20.9	2842.4	-6	-3.3
7	2734.4	20.7	-6.2	3	23.7	4213.4	-5	0.6
8	4088.8	23.5	-6	3	26.5	6263.4	4	5.0
1	244.6	6.2	11	6	12.2	748.2	2	-4.5
2	365.7	8.2	4.5	6	14.2	1025.7	2	-1.3
3	546.9	10.4	2.5	6	16.4	1440.7	2.2	-0.2
4	817.8	12.8	1.75	6	18.8	2061.3	-2.1	-1.9
5	1222.8	15.3	2.1	6	21.3	2989.3	-6.1	-4.1
6	1828.6	17.9	0.5	6	23.9	4376.9	-5	-2.8
7	2734.4	20.7	-6.2	6	26.7	6452.0	3.75	5.0
8	4088.8	23.5	-6	6	29.5	9554.9	11.75	8.9



evaluate the effect of shift in vocoded speech understanding because listeners improve substantially over time (Rosen et al., 1999). As a measure of acute performance without training, the first block was a testing block with no feedback. For testing blocks, listeners were presented with combinations of conditions with unprocessed speech and sine-vocoded speech with 0, 3, or 6 mm of shift. For the testing blocks, stimuli were again presented dichotically over headphones, and listeners were asked to report the sentences in the target ear. A selective attention task was used as in Experiment 1. Listeners were told to attend only to the left ear (target ear) and ignore the right ear (interfering ear), and as in previous experiments the words in both ears could be the same because of the completely random selection.

For training blocks, listeners were presented with only 6 mm of shift, only in the left ear (thus, only the target was presented, no interferer), and were provided with feedback on their performance after each trial. The correct words on the computer interface were highlighted, then the unprocessed sentence was played, and finally the vocoded sentence was repeated (Davis et al., 2005).

The experiment was performed in a method of constant stimuli where all 16 combinations of stimuli were presented 10 times in a randomized order per block. Testing occurred over 2 days. On each day, listeners completed three blocks of 160 trials in each testing block, interspersed with training blocks containing 60 trials each. Therefore, each session consisted of testing and training blocks in the following order. Day 1 consisted of testing (Fig. 3: block number 1, all conditions, no feedback), training (6-mm shift only, with feedback), testing (block number 2), training, and testing (block number 3). Day 2 consisted of testing (Fig. 3: block number 4), training, testing (block number 5), training, and testing (block number 6). Each session lasted 2 hours. Therefore, the experiment took 4 hours, which is comparable to the amount of training needed to see saturation in performance for 6 mm of shift (Rosen et al., 1999; Waked et al., 2017).

## Data analysis

The data were analyzed using a three-way repeated-measures ANOVA with factors block number, target-ear shift (four levels: unprocessed, 0, 3, and 6 mm), and interferer-ear shift (four levels: unprocessed, 0, 3, and 6 mm). As in Experiment 1, transformation of percentages to RAUs, use of Greenhouse-Geisser correction for sphericity violations, and Bonferroni-corrected two-tailed two-sample *t* tests for post hoc comparisons were used. Because improvement due to training was assumed to be monotonic, Helmert contrasts were used to evaluate performance between a given block number and subsequent scores of the remaining blocks. This evaluated the number of blocks needed to show asymptotic performance.

## Results

### PC data

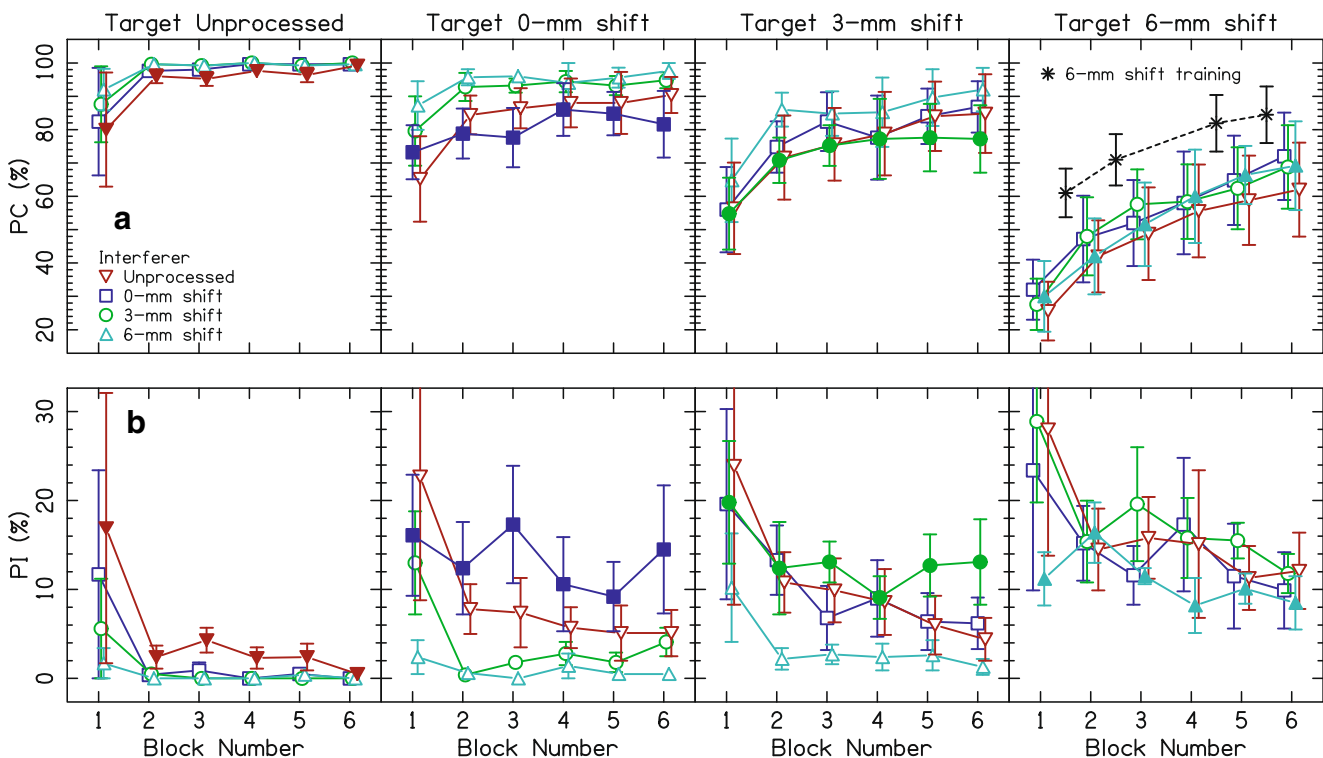
Figure 3A shows the PC values for the testing blocks as a function of block number. The rightmost panel also includes the PC values for the monaural training blocks. The repeated-measures ANOVA results are shown in Table 5. Performance increased with increasing block number ( $p = .045$ ). Helmert contrasts showed that blocks 2, 3, and 5 were different than the subsequent blocks ( $p < .05$  for all three), and blocks 1 and 4 were not different than the subsequent blocks ( $p > .05$  for both). For simpler visualization of the data, the PC values from the last block (i.e., block 6) in Fig. 3a are shown in Fig. 4a.

Performance was near ceiling when the target was unprocessed. In the target ear, performance significantly *decreased* with increasing shift ( $p = .004$ ). Post hoc tests showed that all target-shift conditions were significantly different than the others ( $p < .0001$  for all). In the interferer ear, performance significantly increased with increasing shift ( $p = .010$ ). Post hoc tests showed that performance for the unprocessed interferer condition was not different than the 0-mm interferer-shift condition ( $p > .05$ ), but was significantly lower than the 3-mm and 6-mm interferer-shift conditions ( $p < .05$  for both). Performance for the 0-mm interferer-shift condition was not different than the 3-mm interferer-shift condition ( $p > .05$ ), but was significantly lower than the 6-mm interferer-shift condition ( $p < .0001$ ). Performance for the 3-mm interferer-shift condition was significantly lower than the 6-mm interferer-shift condition ( $p < .0001$ ).

There was a significant target-shift  $\times$  interferer-shift interaction ( $p = .001$ ). This interaction occurred partially because of the relatively smaller differences in interferer conditions for the unprocessed and 6-mm target-shift conditions compared with the 0-mm and 3-mm shift conditions. This interaction also occurred partially because conditions that were matched in shift (e.g., 3-mm target shift and 3-mm interferer shift) showed performance that was equal to or less than all the other conditions at each target shift. This is highlighted in Figs. 3a and 4a with filled symbols for the matched target-ear and interferer-ear shift conditions being the lowest curve in each panel in Fig. 3a or point in Fig. 4a, except for the 6-mm target condition where there were no differences between interferer conditions. All pairwise comparisons for PC are shown in Table 6. The target-shift  $\times$  block number, interferer-shift  $\times$  block number, target-shift  $\times$  interferer-shift  $\times$  block number interactions were not significant ( $p > .05$  for all).

### PI data

Figure 3b shows the PI data for the testing blocks. The repeated-measures ANOVA results are shown in Table 5. Intrusions did not significantly change with block number ( $p$



**Fig. 3** Average PC (row **a**) and PI (row **b**) as a function of testing block number for different amounts of shift in the target ear (different panels) and interferer ear (different symbols) for the selective attention task of Experiment 3. Average PC for the monotic 6-mm shift training blocks is

shown in the rightmost panel of row **a**. Note that the scale on the y-axis changes from rows **a** to **b**, both in absolute numbers and relative size. Otherwise, conventions are the same as in Fig. 1

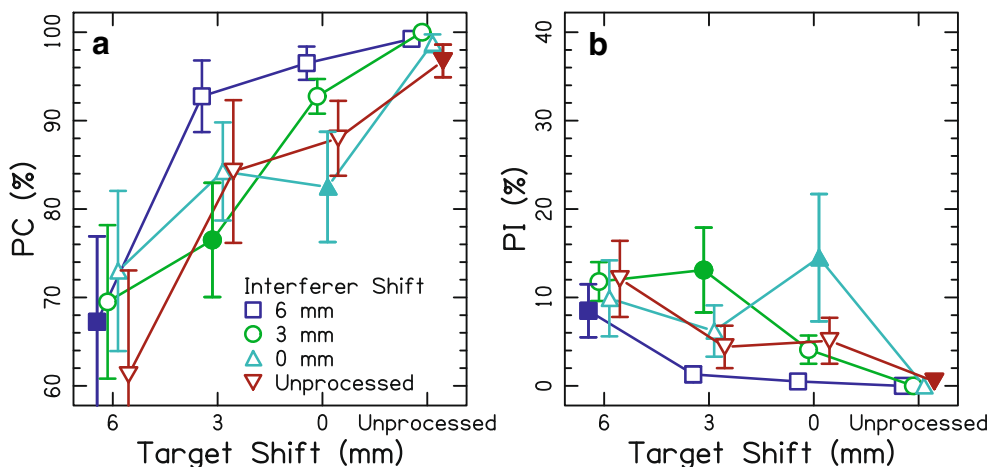
= .21). For simpler visualization of the data, the PI values from the last block (i.e., block 6) in Fig. 3b are shown in Fig. 4b. In the target ear, intrusions significantly increased with increasing shift ( $p < .0001$ ). Post hoc tests showed that all target-shift conditions were significantly different than the others ( $p < .0001$  for all comparisons). In the interferer ear, intrusions significantly decreased with increasing shift ( $p = .041$ ). Post

hoc tests showed that intrusions for the 6-mm interferer-shift condition were lower than for the unprocessed, 0-mm, and 3-mm interferer-shift conditions ( $p < .0001$  for all), but were otherwise not different ( $p > .05$  for all).

There was a significant target-shift  $\times$  interferer-shift interaction ( $p = .005$ ), which occurred because conditions that were matched in shift (e.g., 3-mm target shift and 3-mm

**Table 5** Repeated-measures ANOVA results for Experiment 3 for the two dependent variables (DVs) PC and PI

DV	Factor	<i>F</i>	<i>df</i>	<i>p</i>	$\eta_p^2$
PC	<b>Block</b>	<b>6.78</b>	<b>1.2, 4.9</b>	<b>.045</b>	<b>0.63</b>
	<b>Target-Shift</b>	<b>33.7</b>	<b>1.0, 4.2</b>	<b>.004</b>	<b>0.89</b>
	<b>Interferer-Shift</b>	<b>6.03</b>	<b>3, 12</b>	<b>.01</b>	<b>0.60</b>
	<b>Target-Shift <math>\times</math> Interferer-Shift</b>	<b>12.8</b>	<b>2.8, 11.3</b>	<b>.001</b>	<b>0.76</b>
	Target-Shift $\times$ Block	1.60	1.8, 7.4	.27	0.28
	Interferer-Shift $\times$ Block	1.07	2.9, 11.6	.40	0.21
	Target-Shift $\times$ Interferer-Shift $\times$ Block	0.71	3.2, 12.9	.57	0.15
PI	Block	2.22	1.1, 4.4	.21	0.36
	<b>Target-Shift</b>	<b>22.9</b>	<b>3, 12</b>	<b>&lt;.0001</b>	<b>0.85</b>
	<b>Interferer-Shift</b>	<b>7.77</b>	<b>1.1, 4.5</b>	<b>.041</b>	<b>0.66</b>
	<b>Target-Shift <math>\times</math> Interferer-Shift</b>	<b>8.32</b>	<b>2.5, 10.2</b>	<b>.005</b>	<b>0.68</b>
	Target-Shift $\times$ Block	0.59	2.8, 11.2	.63	0.13
	Interferer-Shift $\times$ Block	1.60	1.8, 7.3	.26	0.29
	Target-Shift $\times$ Interferer-Shift $\times$ Block	1.00	3.3, 13.0	.43	0.20



**Fig. 4** Average PC (a) and PI (b) for the last testing block as a function of target-ear shift for the selective attention task of Experiment 3. Filled symbols highlight the conditions where the shift is equal in the target

and interferer ears. Error bars represent  $\pm 1$  standard error. Note that the scale on the y-axis changes from a to b, but the relative size of the scale does not change

interferer shift) showed intrusions that were equal to or greater than all the other conditions at a fixed level of target shift except for the 6-mm target-shift and interferer-shift condition. This is highlighted in Figs. 3b and 4b with filled symbols for the matched target-ear and interferer-ear shift conditions being the highest curve in each panel, which mirrors the pattern observed in Figs. 3a and 4a for PC. All pairwise comparisons for PI are shown in Table 6. The target-shift  $\times$  block number, the interferer-shift  $\times$  block number, and the target-shift  $\times$  interferer-shift  $\times$  block number interactions were not significant ( $p > .05$  for all).

**Discussion**

The dichotic listening selective attention task used in Experiment 3 simulated a relatively good and poor ear with a vocoder, as in the previous experiments, but this time an 8-channel vocoder with different amounts of frequency-to-place mismatch was used. An important aspect of using this type of vocoder is that while initial performance can be very poor (Dorman et al., 1997; Shannon et al., 1998), improvement with practice and training can recover much of the decrement for relatively large (e.g., 6 mm) shifts (Rosen et al., 1999; Waked et al., 2017). Therefore, training was also implemented in this experiment to determine how performance changed over the course of 4 hours when alternating between conditions with feedback and those that included testing.

The PC data are shown in Figs. 3a and 4a. Mostly similar to Experiments 1 and 2, the results showed that dichotic listening performance was worse as the target ear had more degradation and more shift, but almost always improved as the interferer ear had more degradation and more shift.

One interesting feature in the data is that the matched conditions (shown by closed symbols in Figs. 3a and 4a) had the same or worse performance for each target-ear condition. In

other words, the effect of interferer-ear shift was nonmonotonic depending on the target-ear shift. Such a pattern of data did not appear to occur in the PC data of Experiments 1 and 2 (see Figs. 1a and 2a), but bears some similarity to the PI data in Experiment 2 (see Fig. 2b). Given that shift distorts the mapping of acoustic cues to phonemes, listeners likely must learn novel or adapt current mappings for the 3-mm and 6-mm shift conditions. There is also an additional pitch cue introduced for conditions that have different shifts; more shift increases the carrier frequencies and would produce higher pitches. The pattern of data suggests that confusion occurs most when there is the most similarity of the sounds across the ears. Such similarity has also been shown when attempting to perform dichotic listening with unprocessed talkers and interferers (e.g., Treisman, 1964). Specifically, less interference is seen when using steady-state noise interferers (Brungart & Simpson, 2002) and speech-like modulations are important for producing interference (Brungart et al., 2005). Such an interpretation of the data is also supported by the intrusions shown in Fig. 3b. Generally, the PI data of Experiment 3 mirror the PC data, meaning that for the matched conditions, PC was the worst because PI was the highest. In other words, the reason PC was relatively low was that listeners were more inclined to respond to the words in the interferer ear.

There was an improvement in performance with block number, and improvement was largest for the conditions with 3 and 6 mm of shift. Such a result was expected given previous literature showing the effects of training and adaptation with shifted vocoders (Rosen et al., 1999). For this specific corpus of matrix sentences, 4 hours of training and interleaved testing previously showed that listeners can improve to approximately 60% correct with a 6-mm shift (Waked et al., 2017). During the four training blocks of the current study, which used monotic presentation, PC for the 6-mm shift

Table 6 Bonferroni-corrected paired comparisons (two-sample two-tailed *t* tests) for different target-interferer shift combinations for PC and PI in Experiment 3

PC Target	Target Interferer	U U	U U	U 3	U 6	U U	0 0	0 0	0 3	0 6	3 U	3 U	3 0	3 3	3 3	3 6	6 U	6 0	6 3	6 6	6 6	6 6	6 6	6 6			
																									U	U	
U	U	--	--	↓*	--	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
U	0	--	--	--	--	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
U	3	--	--	--	--	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
U	6	--	--	--	--	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
0	U	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
0	0	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
0	3	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
0	6	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
3	U	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
3	0	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
3	3	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
3	6	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
6	U	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
6	0	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
6	3	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
6	6	--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
		--	--	↓	↓***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***	↑***
PI Target	Target Interferer	U U	U U	U 3	U 6	U U	0 0	0 0	0 3	0 6	3 U	3 U	3 0	3 3	3 3	3 6	6 U	6 0	6 3	6 6	6 6	6 6	6 6	6 6	6 6	6 6	
U	U	--	--	↑*	--	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
U	0	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
U	3	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
U	6	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
0	U	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
0	0	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
0	3	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
0	6	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
3	U	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
3	0	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
3	3	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
3	6	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
6	U	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
6	0	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
6	3	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
6	6	--	--	↑	↑	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	

Note. “U” denotes the unprocessed condition. Significant differences: \**p* < .05, \*\**p* < .01, \*\*\**p* < .001. The upward arrows show when the condition for the row is higher than the condition for the column.

condition improved from 61.0% for training block 1 to 84.5% for training block 4. The difference between the studies may have been the result of different listeners or different testing conditions. The current study had longer training blocks, 60 trials per block compared with 30 per block in Waked et al. (2017).

PC for the monotic training blocks was approximately 10%–20% higher than PC for the dichotic testing blocks. Improvement occurred for all three vocoded conditions. For unshifted vocoded speech, there is usually rapid improvement on the order of a few sentences (Davis et al., 2005). For this speech corpus specifically, performance saturated after one session of training with feedback for a 0-mm shift. Performance saturated by four feedback training blocks at the group level for a 6-mm shift, although variability showed some individuals improving even after 11 blocks of training and testing (Waked et al., 2017). In the current study, Helmert contrasts did not show a convincing plateau in performance and, therefore, listeners appeared to need more training and testing time to demonstrate a saturation in performance. It could be that not only did the listeners improve in understanding the vocoded speech, but they also improved at attending to the target ear in the dichotic listening task (Tallus et al., 2015). On one hand, PI significantly decreased with increasing block number, supporting the idea that listeners may have improved at focusing their attention, not just identifying the correct words. On the other hand, the monotic training and dichotic testing conditions were parallel for the 6-mm shift conditions (see Fig. 3a, rightmost panel), suggesting that the increased PC in both conditions were primarily driven by improvements in shifted vocoded speech understanding.

Other studies have performed dichotic or binaural hearing studies using different shifts between the ears. Siciliano et al. (2010) tested 6-channel vocoded and shifted maps using channels that were interleaved across the ears. They found that listeners were unable to integrate the information across the ears and performance never exceeded the performance of the unshifted condition with three channels in one ear, despite the fact that the shifted ear had three channels of potentially useful but admittedly distorted speech information. It could be that the listeners were unable to support interaural integration with only three channels in each ear, and interaural integration may have been possible if more information (i.e., more channels) were available. Such a result would be more consistent with the findings of the current study that showed some interference and interaural integration. Goupell, Stoelb, et al. (2018) measured spatial release from masking (i.e., the improvement in performance when comparing colocated and spatially separated talkers) using vocoders and shift. They found that spatial release from masking

was reduced as interaural mismatch was increased to 6 mm. Interestingly, the change in spatial release from masking was asymmetrical across ears based on the spatial location of the target. Spatial release from masking decreased more as the ear with the better signal-to-noise ratio (SNR) was shifted to higher frequencies compared with when the worse SNR ear was shifted. In other words, it appeared that listeners used the speech information in the ear with the worse SNR because it was more intelligible; the trade-off of SNR and intelligibility meant that attending to a 6-mm shifted ear with a relatively better SNR did not emerge. Such a result is broadly consistent with the findings of this experiment; it appears that it is easier to ignore a poor signal and to attend to a good one, even when the instructions are to direct attention to the poorer ear.

## General discussion

### Overview

In typical human sensory systems like hearing and vision, there is bilaterally symmetrical processing of inputs. With an increase in the number of people receiving hearing rehabilitation with a CI, there are more people experiencing large functional asymmetries in their hearing, which may have consequences for their ability to perform spatial-hearing tasks. This paper investigated the effects of functional hearing asymmetry on dichotic listening by parametrically and independently varying the spectral resolution or shift in both the target (left) and interferer (right) ears. We hypothesized that it would be more difficult to attend to a relatively functionally poorer ear (i.e., less resolution or more shift) and ignore a relatively better ear, therefore simulating some of the asymmetrical ear differences that could occur in the CI population. All three experiments that varied the type of vocoder (noise-excited/number of channels [Experiments 1 and 2] vs. tone-excited/shift [Experiment 3]), type of dichotic listening (selective [Experiments 1 and 3] vs. divided attention [Experiment 2]), and training (Experiment 3) consistently supported this hypothesis (see Figs. 1a, 2a, 3a, and 4a). The interpretation was further supported by observing the intrusions, or times when the listener reported the words presented in the non-target ear (see Figs. 1b, 2b, 3b, and 4b).

### Results viewed through the shared resource and integrated strategy models

Critical features in the data of these experiments were the nonmonotonic PC and PI functions; the worst performance and most intrusions occurred for the conditions where the vocoder was matched across the ears in Experiment 3 (see



Figs. 3 and 4). The PI values of Experiment 2 (Fig. 2B) showed a similar but less pronounced pattern. There was little evidence of such a pattern in Experiment 1 (Fig. 1).

Two models have been proposed to explain within-ear and across-ear interference effects over a number of studies, which are relevant to our data set. Brungart and Simpson (2002), in an effort to explain why listeners experienced more across-ear interference when the target ear was at relatively lower (negative) target-to-masker ratio, proposed a model based on a “shared resource” limitation. Under this model, the greater the difficulty of the listening task in the target ear, the fewer the resources that would be left to perform subsequent speech segregation (e.g., suppressing a contralateral interferer in a dichotic listening experiment). Such a model is consistent with Gallun et al. (2007a) and Gallun et al. (2007b), who suggested that an important factor in these types of tasks is the task-based demands on processing resources. Listeners appeared to combine information across ears more often when spectrotemporal similarity and task demands were high. Therefore, stimulus information and task complexity should play a role in dichotic listening, and interference across ears should occur when more resources are allocated to more demanding tasks.

Data from Kidd et al. (2003), however, were not consistent with the idea that the most interference is observed when the task is most difficult in the target ear. They found that 1-kHz tone detection in a set of random-frequency tone maskers (which was relatively easy to segregate) showed more across-ear interference than when there was a fixed-frequency tone masker (which was relatively difficult to segregate). Data from that experiment and those in Brungart and Simpson (2007), prompted the development of an alternative “integrated strategy model” of dichotic listening. Under this model, it is assumed that listeners can focus attention using a single listening strategy (e.g., spatial, pitch, loudness differences), and this strategy is applied to the combined information arriving at the two ears. The strictest form of this model forces the listener to adopt a single listening strategy, presumably the one that provides the best overall segregation. Then this strategy is applied to each stream. For example, listeners had varying levels of difficulty ignoring a contralateral speech interferer when there was an ipsilateral interferer also in the target ear, depending on similarity of talkers (i.e., the difficulty of the segregation) in the target ear (Brungart & Simpson, 2002, 2004, 2007).

In summary, the “shared resource” model of dichotic attention predicts that across-ear interference depends primarily on the difficulty of the within-ear listening task. The “integrated strategy” model predicts performance depends on the within-ear and across-ear target interferer similarity. The most across-ear interference would occur when the across-ear interferer is more similar to the target talker than the within-ear interferer

(for the cases of having interferers in both ears as in Brungart & Simpson, 2007, not in cases of pure dichotic listening with no interferer in the target ear, as were performed in our current study). This is because the optimal strategy to segregate target and interferer in the attended ear would be suboptimal for eliminating interference from the unattended ear interferer. An important prediction for the integrated strategy model is that for a dichotic listening task with the target in one ear and the interferer in the other ear, the listener would presumably choose the optimal strategy of attending to the target ear, and therefore minimal or no across-ear interference would occur (Brungart & Simpson, 2007).

Our data appear to be broadly consistent with the data in Brungart and Simpson (2007), where the most interference occurred when there was the most similarity between across-ear target and interferer; this finding partially supports the integrated strategy model. Our data are also consistent with Brungart et al. (2005), where performance decreased systematically as the number of channels in a vocoder interferer ear was increased (i.e., there was more interference as the interferer became more intelligible); this finding partially supports the shared resource model if it were extended to include both the difficulty of processing both the target and interferer ears. Both models could be combined into a “relative salience” model. In such a model, resources would be allocated to perform multiple tasks, like understanding vocoded speech or allocating space to short-term memory. Resources for segregation would be limited by these other tasks, and signals that had relatively more similarity would require more resources for segregation. If those resources are limited by other tasks, this would produce an increasing number of errors in dichotic listening.

Specifically, for the data collected in this study, a relative salience model might explain an optimal listening strategy in the following way. First, a fixed amount of resources would need to be allocated to short-term memory. Experiment 2 would have a particularly high memory demand because listeners had to remember two speech streams, each five items long. Listeners would know how much memory to preallocate from performing the numerous trials of each experiment. Second, the level of degradation of the vocoding would make the listener use different amounts of resources in understanding the speech, more as the signal became poorer in resolution or had more shift. This allocation would be unknown to the listener from trial-to-trial given the randomization of vocoding conditions within blocks. Then, the strict form of the integrated strategy model suggests that all speech would then be combined and then segregation occurs based on a single best rule. For all three experiments in the current study, attention to the target by ear selection should then occur and be constant across studies. However, the striking nonmonotonicities of Experiment 3 (Figs. 3 and 4) suggest that the stream segregation by ear is imperfect. In some cases and studies, there is the

assumption that ear selection is nearly perfect (e.g., Brungart & Simpson, 2007, p. 1725), while others acknowledge that some of the stream in the nonselected ear remains and is only attenuated (e.g., Brungart & Simpson, 2007, p. 1733; Treisman, 1964).

It should be noted here that Experiment 3 likely had the largest task demands for understanding the 6-mm shifted vocoded speech given the poor performance on this condition, thus leaving the fewest resources for stream segregation. The nonmonotonicities in Experiment 3 could be explained if a secondary segregation were added to the model; the secondary segregation would be based on pitch differences because the pitch of the vocoded speech will depend on the amount of shift. This secondary segregation could occur before the ear selection and would appear consistent with the idea of having multiple selection modes, such as early modes based on acoustic stimulus properties (Wightman et al., 2010) and late modes based on semantic analysis (Broadbent, 1958; Johnston & Heinz, 1978). Alternatively, the secondary segregation could occur after that selection if it is imperfect and leaves some residual trace of an attenuated interferer stream. A reasonable assumption for having primary and secondary segregation stages would be that the primary would be the more effective of the two, perhaps having more resources allocated to it.

A relative salience model could also explain the lack of nonmonotonicities in Experiment 1. In that experiment, the noise vocoder would have produced stimuli that had no appreciable pitch differences (or other salient segregation cue, assuming intelligibility is not a strong cue) for different resolutions, meaning the stimuli would have sounded very similar across conditions. This would have provided less of a secondary segregation cue, and therefore produced data that lack non-monotonicities for the matched-resolution conditions. The fact that there was a hint of nonmonotonicity for PI in Experiment 2 might be explained by the relatively large short-term memory demands on a limited amount of available resources before stream segregation, or by the fact that a divided attention task might naturally have a higher propensity of responding to the nontarget ear.

There are two alternative explanations that could predict the nonmonotonicities in Figs. 3 and 4. It could be that listeners were segregating via pitch and not ear; such an explanation would better align with the strict integrated strategy model where listeners only had the ability to use the single best rule. Such an explanation is compelling because of its simplicity and would suggest pitch cues are stronger segregation cues than spatial cues. Given the assumed ease of dichotic listening in most cases, however, it is not clear segregation via pitch would indeed be stronger for these stimuli. It could also be that the words for the matched shifts had a greater propensity to fuse, yielding the resultant data patterns. Such an explanation would only apply for the tone vocoder of

Experiment 3, and not for the noise vocoder used in Experiments 1 and 2. As mentioned previously, because the corpus has different words and different talkers, the envelopes were quite different across the ears and likely limited the perceived fusion.

In summary, the current data set seems consistent with many aspects of the shared resource (Brungart & Simpson, 2002) and integrated strategy (Brungart & Simpson, 2007) models. A combined relative salience model was proposed that has a finite amount of resources, attends to the intelligibility of both the target *and* interferer ear stimuli (including the relative differences across the two ears), and perhaps includes more than one stream segregation. Such a conceptual framework would appear to explain the lack of nonmonotonicities in Experiment 1 and the clear existence of them in Experiment 3. Admittedly, this relative salience model is more complicated than either of the other original dichotic listening models, but still should capture the effects seen in the studies that motivated the other models (Brungart & Simpson, 2002, 2004, 2007; Brungart et al., 2005; Brungart et al., 2001; Gallun et al., 2007a; Kidd et al., 2003; Kidd et al., 2005).

### Comparison with CI performance

While the vocoding in this study only crudely simulates the degradations of what is actually experienced by real CI listeners, it provided a simple signal processing tool to simulate functionally asymmetrical hearing. The results from these simulations could partially explain performance differences seen in actual CI listeners. For example, most bilateral CI listeners demonstrate a relatively symmetrical ability to perform selective attention; however, there were two of 11 adult (Goupell et al., 2016) and two of 10 child (Misurelli et al., 2020) bilateral CI listeners with highly asymmetric performance. The data from those studies can be compared with the simulated asymmetrical selective attention performance in Experiments 1 and 3 (Fig. 1a and 3/4, respectively). The larger differences from asymmetry for the real CI listeners compared with the smaller differences in simulated CI listeners suggest that differences in resolution and/or shift might only partially explain the CI data. Comparisons between NH and CI listeners were directly shown in Experiment 2 (Fig. 2) for divided attention, for which there was good correspondence in the data sets. Similarly, Goupell, Stakhovskaya, et al. (2018) intentionally recruited bilateral CI listeners that could have asymmetrical hearing abilities, and showed in a speech-on-speech masking task that binaural unmasking and head shadow were reduced. In the broader literature, individual bilateral CI and hearing-aid users show variable changes in performance for adding a second ear; those that show little benefit or even a decrement may be a result of asymmetrical hearing (Mosnier et al., 2009; Polonenko et al., 2018; Reeder et al., 2014; Walden & Walden, 2005).

Single-sided deafness is arguably the most asymmetrical listening condition that can be achieved because there is normal acoustic hearing in one ear and a CI in the other. Investigations involving real single-sided-deafness CI listeners are relevant from two studies with contrasting results (Bernstein et al., 2016; Bernstein et al., 2020). They found that binaural unmasking of speech could be produced if the target was in the NH ear, but showed that interference could occur if the target was in the CI ear. Such a result is consistent with the finding of the current study broadly showing that attending to a target in a relatively poorer ear is difficult when a clearer interfering talker is present.

### Limitations and future directions

While the current study is an important first step in understanding asymmetrical hearing, dichotic listening is only related to the cocktail party phenomenon, and is not a realistic configuration that people typically experience. Rather, listeners would experience the signals from both target and interferer in both ears. When performing spatial release from masking from more realistic sound sources, interaurally asymmetrical shifts decrease binaural benefits (Goupell, Stoelb, et al., 2018; Xu et al., 2020). In addition, listeners attend to a functionally better (i.e., less shifted) ear, even when they experience a worse target-to-masker ratio (Goupell, Stoelb, et al., 2018), similar to the main finding of the current study that it is relatively difficult to ignore clearer speech. Therefore, another avenue to explore is adapting this type of study to a more realistic listening situation, such as for different talkers that are realistically spatially separated. In such cases, a wide range of listeners report a spatial release from masking. The physics of the head can introduce interaural time and level differences to convey spatially separated sound locations. In contrast, the dichotic listening experiment in this study was effectively an infinite interaural level difference. The changing of resolution across ears would affect spatial cues and perceived separation of sources, but the results of the present study suggest that spatial release from masking might also be affected by attention mechanisms.

Another additional important set of conditions that was omitted in this study was monotic control measurements. Figure 3 shows that there is a decrease in performance of 10%–20% from monotic to dichotic listening, similar to the decrease found in Goupell et al. (2016) and Misurelli et al. (2020). Monotic control measurements would have explicitly reflected asymmetries across conditions, which would be helpful in interpreting effects based on the relative intelligibility of the target and interferer speech. Such control conditions would be critical as a baseline for actual CI listeners.

Future work should also consider further characterizing the NH listeners. There were no questionnaires inquiring about asymmetrical and chronic ear infection history. In addition,

use of the left ear only as the target ear in the selective attention tasks (Experiments 1 and 3) may have affected performance because of the right-ear advantage (Cooper et al., 1967; Kimura, 1967). It could be that performance was modulated more greatly because of this, since intrusions may be more likely when attending the left ear (Hugdahl et al., 2001). Investigating ear effects is a possible future direction of this type of work.

One limitation of this study was the use of this particular corpus, which had sentences that were syntactically correct, but sometimes semantically irregular. There was also a range of confusability within word categories (see Table 1). Confusions within word categories may have decreased PC and PI slightly. For example, if the target word was “Blue” and the interferer word was “Old,” and the subject responded with “Cold,” this would not count as an intrusion, but would be a legitimate confusion, given the small differences between “Old” and “Cold,” especially for spectrally degraded speech. While this is likely a small effect that may particularly alter the most degraded conditions, future studies could attempt using a different speech corpus.

Another procedural manipulation that could be explored is how the different conditions were randomly presented across trials. This is in contrast to how actual CI users experience degraded sounds, where they listen to a fixed resolution in each ear. The results may differ if the vocoder condition were fixed in blocks.

Broadly, this study shows how the brain prefers to attend to a more salient signal or one that is easier to understand over a poorer one. Such work on asymmetries has previously been studied in the auditory system, particularly when considering auditory development (Gordon et al., 2015; Kral et al., 2013; Tillein et al., 2016; Whitton & Polley, 2011). There may be some similarities to the much more well studied and understood phenomenon of amblyopia, but it remains unclear whether the effects of asymmetry are more similar or different when comparing hearing and vision (Gordon & Kral, 2019; Tillein et al., 2016). What is most interesting from this study is that effects of asymmetrical processing occur in the absence of developmental changes.

**Acknowledgments** We would like to thank Anna Tinnemore, Miranda Cleary, Bobby Gibbs, Josh Bernstein, and Doug Brungart for helpful discussions about this work. We would like to thank Cochlear Ltd. for providing the testing equipment and technical support. Research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R01DC015798 (M.J.G. and Joshua G. W. Bernstein). The word corpus was funded by NIH-NIDCD Grant P30DC04663 (BU Hearing Research Center core grant). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Portions of this work were presented at the Association for Research in Otolaryngology 35th Midwinter Meeting and the 16th Conference on Implantable Auditory Prostheses.

## References

- Arbogast, T. L., Mason, C. R., & Kidd, G., Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America*, *112*(5 Pt. 1), 2086–2098. <https://doi.org/10.1121/1.1510141>
- Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2015). *Memory* (2nd ed.). London, United Kingdom: Psychology Press.
- Barrett, B. T., Bradley, A., & McGraw, P. V. (2004). Understanding the neural basis of amblyopia. *Neuroscientist*, *10*(2), 106–117. <https://doi.org/10.1177/1073858403262153>
- Bernstein, J. G. W., Goupell, M. J., Schuchman, G., Rivera, A., & Brungart, D. S. (2016). Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees. *Ear and Hearing*, *37*(3), 282–288. <https://doi.org/10.1097/AUD.0000000000000284>
- Bernstein, J. G. W., Stakhovskaya, O. A., Jensen, K. K., & Goupell, M. J. (2020). Acoustic hearing can interfere with single-sided deafness cochlear-implant speech perception. *Ear and Hearing*, *41*(4), 747–761. <https://doi.org/10.1097/AUD.0000000000000805>
- Best, V., Thompson, E. R., Mason, C. R., & Kidd, G., Jr. (2013). An energetic limit on spatial release from masking. *Journal of the Association for Research in Otolaryngology*, *14*(4), 603–610. <https://doi.org/10.1007/s10162-013-0392-1>
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065–1066. <https://doi.org/10.1121/1.428288>
- Broadbent, D. E. (1958). *Perception and communication*. USA: Oxford University Press.
- Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *The Journal of the Acoustical Society of America*, *112*(6), 2985–2995. <https://doi.org/10.1121/1.1512703>
- Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *The Journal of the Acoustical Society of America*, *115*(1), 301–310. <https://doi.org/10.1121/1.1628683>
- Brungart, D. S., & Simpson, B. D. (2007). Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task. *The Journal of the Acoustical Society of America*, *122*(3), 1724–1734. <https://doi.org/10.1121/1.2756797>
- Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd, G., Jr. (2005). Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *The Journal of the Acoustical Society of America*, *117*(1), 292–304. <https://doi.org/10.1121/1.1835509>
- Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5 Pt. 1), 2527–2538. <https://doi.org/10.1121/1.1628683>
- Bryden, M. P., Munhall, K., & Allard, F. (1983). Attentional biases and the right-ear effect in dichotic listening. *Brain and Language*, *18*(2), 236–248. [https://doi.org/10.1016/0093-934x\(83\)90018-4](https://doi.org/10.1016/0093-934x(83)90018-4)
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. <https://doi.org/10.1121/1.1907229>
- Cleary, M., Wilkinson, T., Wilson, L., & Goupell, M. J. (2018). Memory span for spoken digits in adults with cochlear implants or typical hearing: Effects of age and identification ability. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2099–2114. [https://doi.org/10.1044/2018\\_JSLHR-H-17-0245](https://doi.org/10.1044/2018_JSLHR-H-17-0245)
- Cooper, A., Achenbach, K., Satz, P., & Levy, C. M. (1967). Order of report and ear asymmetry in dichotic listening. *Psychonomic Science*, *9*(2), 97–98. <https://doi.org/10.3758/BF03330777>
- Croghan, N. B. H., Duran, S. I., & Smith, Z. M. (2017). Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility. *The Journal of the Acoustical Society of America*, *142*(6), EL537–EL543. <https://doi.org/10.1121/1.5016044>
- Dai, B., McQueen, J. M., Hagoort, P., & Kosem, A. (2017). Pure linguistic interference during comprehension of competing speech signals. *The Journal of the Acoustical Society of America*, *141*(3), EL249–EL254. <https://doi.org/10.1121/1.4977590>
- Davis, M. H., Johnsruide, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. <https://doi.org/10.1037/0096-3445.134.2.222>
- Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *The Journal of the Acoustical Society of America*, *102*(5 Pt. 1), 2993–2996. <https://doi.org/10.1121/1.420354>
- Dudley, H. W. (1939). Remaking speech. *The Journal of the Acoustical Society of America*, *11*(2), 1969–1977. <https://doi.org/10.1121/1.1916020>
- Faulkner, A., Rosen, S., & Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *The Journal of the Acoustical Society of America*, *113*(2), 1073–1080. <https://doi.org/10.1121/1.1536928>
- Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. <https://doi.org/10.1121/1.1381538>
- Gallun, F. J., Mason, C. R., & Kidd, G., Jr. (2007a). The ability to listen with independent ears. *The Journal of the Acoustical Society of America*, *122*(5), 2814–2825. <https://doi.org/10.1121/1.2780143>
- Gallun, F. J., Mason, C. R., & Kidd, G., Jr. (2007b). Task-dependent costs in processing two simultaneous auditory stimuli. *Perception & Psychophysics*, *69*(5), 757–771. <https://doi.org/10.3758/bf03193777>
- Gordon, K. A., Henkin, Y., & Kral, A. (2015). Asymmetric hearing during development: The aural preference syndrome and treatment options. *Pediatrics*, *136*(1), 141–153. <https://doi.org/10.1542/peds.2014-3520>
- Gordon, K. A., & Kral, A. (2019). Animal and human studies on developmental monaural hearing loss. *Hearing Research*, *380*, 60–74. <https://doi.org/10.1016/j.heares.2019.05.011>
- Goupell, M. J., Kan, A., & Litovsky, R. Y. (2016). Spatial attention in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, *140*(3), 1652–1662. <https://doi.org/10.1121/1.4962378>
- Goupell, M. J., & Litovsky, R. Y. (2014). The effect of interaural fluctuation rate on correlation change discrimination. *Journal of the Association for Research in Otolaryngology*, *15*(1), 115–129. <https://doi.org/10.1007/s10162-013-0426-8>
- Goupell, M. J., Stakhovskaya, O. A., & Bernstein, J. G. W. (2018). Contralateral interference caused by binaurally presented competing speech in adult bilateral cochlear-implant users. *Ear and Hearing*, *39*(1), 110–123. <https://doi.org/10.1097/AUD.0000000000000470>
- Goupell, M. J., Stoelb, C., Kan, A., & Litovsky, R. Y. (2013). Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening. *The Journal of the Acoustical Society of America*, *133*(4), 2272–2287. <https://doi.org/10.1121/1.4792936>



- Goupell, M. J., Stoelb, C. A., Kan, A., & Litovsky, R. Y. (2018). The effect of simulated interaural frequency mismatch on speech understanding and spatial release from masking. *Ear and Hearing, 39*(5), 895–905. <https://doi.org/10.1097/AUD.0000000000000541>
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592–2605. <https://doi.org/10.1121/1.399052>
- Hugdahl, K., Carlsson, G., & Eichele, T. (2001). Age effects in dichotic listening to consonant-vowel syllables: Interactions with attention. *Developmental Neuropsychology, 20*(1), 445–457. [https://doi.org/10.1207/S15326942DN2001\\_8](https://doi.org/10.1207/S15326942DN2001_8)
- Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America, 120*(5 Pt. 1), 2926–2937. <https://doi.org/10.1121/1.2354070>
- Johnston, W., & Heinz, S. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107*(4), 420–435. <https://doi.org/10.1037/0096-3445.107.4.420>
- Kaplan, A. B., Kozin, E. D., Remenschneider, A., Eftekhari, K., Jung, D. H., Polley, D. B., & Lee, D. J. (2016). Amblyaudia: Review of pathophysiology, clinical presentation, and treatment of a new diagnosis. *Otolaryngology—Head and Neck Surgery, 154*(2), 247–255. <https://doi.org/10.1177/0194599815615871>
- Kidd, G., Jr., Best, V., & Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *The Journal of the Acoustical Society of America, 124*(6), 3793–3802. <https://doi.org/10.1121/1.2998980>
- Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., & Simpson, B. D. (2003). Informational masking caused by contralateral stimulation. *The Journal of the Acoustical Society of America, 113*(3), 1594–1603. <https://doi.org/10.1121/1.1547440>
- Kidd, G., Jr., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *The Journal of the Acoustical Society of America, 118*(2), 982–992. <https://doi.org/10.1121/1.1953167>
- Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3*, 163–168. [https://doi.org/10.1016/S0010-9452\(67\)80010-8](https://doi.org/10.1016/S0010-9452(67)80010-8)
- Knudsen, L. V., Oberg, M., Nielsen, C., Naylor, G., & Kramer, S. E. (2010). Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature. *Trends in Hearing, 14*(3), 127–154. <https://doi.org/10.1177/1084713810385712>
- Kochkin, S. (2009, October 1). MarkeTrak VIII: 25-year trends in the hearing health market. *Hearing Review, 16*(11), 12–31. <https://www.hearingreview.com/practice-building/marketing/marketrak-viii-25-year-trends-in-the-hearing-health-market>
- Kral, A., Heid, S., Hubka, P., & Tillein, J. (2013). Unilateral hearing during development: Hemispheric specificity in plastic reorganizations. *Frontiers in Systems Neuroscience, 27*, 1–13. <https://doi.org/10.3389/fnsys.2013.00093>
- Landsberger, D. M., Svrakic, M., Roland, J. T., Jr., & Svirsky, M. (2015). The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants. *Ear and Hearing, 36*(5), e207–e213. <https://doi.org/10.1097/AUD.0000000000000163>
- Litovsky, R. Y., & Gordon, K. (2016). Bilateral cochlear implants in children: Effects of auditory experience and deprivation on auditory perception. *Hearing Research, 338*, 76–87. <https://doi.org/10.1016/j.heares.2016.01.003>
- 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., & Misurelli, S. (2012). Studies on bilateral cochlear implants at the University of Wisconsin’s Binaural Hearing and Speech Laboratory. *Journal of the American Academy of Audiology, 23*(6), 476–494. <https://doi.org/10.3766/jaaa.23.6.9>
- Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., & Callahan, M. R. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Perception & Psychophysics, 65*(3), 396–406. <https://doi.org/10.3758/bf03194571>
- Misurelli, S. M., Goupell, M. J., E, A. B., Jocewicz, R., Kan, A., & Litovsky, R. Y. (2020). Auditory attention and spatial unmasking in children with cochlear implants. *Trends in Hearing, 24*, 1–11. <https://doi.org/10.1177/2331216520946983>
- Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., Ambert-Dahan, E., & Ferrary, E. (2009). Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients. *Audiology and Neurotology, 14*(2), 106–114. <https://doi.org/10.1159/000159121>
- Polonenko, M. J., Papsin, B. C., & Gordon, K. A. (2018). Limiting asymmetric hearing improves benefits of bilateral hearing in children using cochlear implants. *Scientific Reports, 8*(1), 1–17. <https://doi.org/10.1038/s41598-018-31546-8>
- Reeder, R. M., Firszt, J. B., Holden, L. K., & Strube, M. J. (2014). A longitudinal study in adults with sequential bilateral cochlear implants: Time course for individual ear and bilateral performance. *Journal of Speech, Language, and Hearing Research, 57*(3), 1108–1126. [https://doi.org/10.1044/2014\\_jslhr-h-13-0087](https://doi.org/10.1044/2014_jslhr-h-13-0087)
- Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America, 106*(6), 3629–3636. <https://doi.org/10.1121/1.428215>
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303–304. <https://doi.org/10.1126/science.270.5234.303>
- Shannon, R. V., Zeng, F. G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *The Journal of the Acoustical Society of America, 104*(4), 2467–2476. <https://doi.org/10.1121/1.423774>
- Siciliano, C. M., Faulkner, A., Rosen, S., & Mair, K. (2010). Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *The Journal of the Acoustical Society of America, 127*(3), 1645–1660. <https://doi.org/10.1121/1.3293002>
- Souza, P., & Rosen, S. (2009). Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *The Journal of the Acoustical Society of America, 126*(2), 792–805. <https://doi.org/10.1121/1.3158835>
- Studebaker, G. A. (1985). A ‘rationalized’ arcsine transform. *Journal of Speech, Language, and Hearing Research, 28*(3), 455–462. <https://doi.org/10.1044/jslhr.2803.455>
- Tallus, J., Soveri, A., Hamalainen, H., Tuomainen, J., & Laine, M. (2015). Effects of auditory attention training with the dichotic listening task: Behavioural and neurophysiological evidence. *PLOS ONE, 10*(10), 1–14. <https://doi.org/10.1371/journal.pone.0139318>
- Tillein, J., Hubka, P., & Kral, A. (2016). Monaural congenital deafness affects aural dominance and degrades binaural processing. *Cerebral Cortex, 26*(4), 1762–1777. <https://doi.org/10.1093/cercor/bhv351>
- Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. *The American Journal of Psychology, 77*, 533–546. <https://doi.org/10.2307/1420765>
- Waked, A., Dougherty, S., & Goupell, M. J. (2017). Vocoded speech understanding with simulated shallow insertion depths in adults and children. *The Journal of the Acoustical Society of America, 141*(1), EL45–EL50. <https://doi.org/10.1121/1.4973649>
- Walden, T. C., & Walden, B. E. (2005). Unilateral versus bilateral amplification for adults with impaired hearing. *Journal of the American*



- Academy of Audiology*, 16(8), 574–584. <https://doi.org/10.3766/jaaa.16.8.6>
- Whitmer, W. M., Seeber, B. U., & Akeroyd, M. A. (2012). Apparent auditory source width insensitivity in older hearing-impaired individuals. *The Journal of the Acoustical Society of America*, 132(1), 369–379. <https://doi.org/10.1121/1.4728200>
- Whitton, J. P., & Polley, D. B. (2011). Evaluating the perceptual and pathophysiological consequences of auditory deprivation in early postnatal life: A comparison of basic and clinical studies. *Journal of the Association for Research in Otolaryngology*, 12(5), 535–547. <https://doi.org/10.1007/s10162-011-0271-6>
- Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., & Oh, E. (2003). Children's detection of pure-tone signals: Informational masking with contralateral maskers. *The Journal of the Acoustical Society of America*, 113(6), 3297–3305. <https://doi.org/10.1121/1.1570443>
- Wightman, F. L., & Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, 101(2), 1050–1063. <https://doi.org/10.1121/1.418029>
- Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, 118(5), 3164–3176. <https://doi.org/10.1121/1.2082567>
- Wightman, F. L., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America*, 119(6), 3940–3949. <https://doi.org/10.1121/1.2195121>
- Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *The Journal of the Acoustical Society of America*, 128(1), 270–279. <https://doi.org/10.1121/1.3436536>
- Wood, N. L., & Cowan, N. (1995a). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, 124(3), 243–262. <https://doi.org/10.1037//0096-3445.124.3.243>
- Wood, N. L., & Cowan, N. (1995b). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 255–260. <https://doi.org/10.1037//0278-7393.21.1.255>
- Xu, K., Willis, S., Gopen, Q., & Fu, Q. J. (2020). Effects of spectral resolution and frequency mismatch on speech understanding and spatial release from masking in simulated bilateral cochlear implants. *Ear and Hearing*, 41(5), 1362–1371. <https://doi.org/10.1097/AUD.0000000000000865>
- Zurek, P. M. (1992). *Binaural advantages and directional effects of speech intelligibility*. Boston: Allyn & Bacon.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.