


# A Comparison of Place-Pitch-Based Interaural Electrode Matching Methods for Bilateral Cochlear-Implant Users

Trends in Hearing  
Volume 25: 1–21  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2331216521997324  
journals.sagepub.com/home/tia  


Kenneth K. Jensen<sup>1</sup>, Stefano Cosentino<sup>2</sup>, Joshua G. W. Bernstein<sup>1</sup>,  
Olga A. Stakhovskaya<sup>1,2</sup>, and Matthew J. Goupell<sup>2</sup> 

## Abstract

Interaural place-of-stimulation mismatch for bilateral cochlear-implant (BI-CI) listeners is often evaluated using pitch-comparison tasks that can be susceptible to procedural biases. Bias effects were compared for three sequential interaural pitch-comparison tasks in six BI-CI listeners using single-electrode direct stimulation. The reference (right ear) was a single basal, middle, or apical electrode. The comparison electrode (left ear) was chosen from one of three ranges: basal half, full array, or apical half. In Experiment 1 (discrimination), interaural pairs were chosen randomly (method of constant stimuli). In Experiment 2 (ranking), an efficient adaptive procedure rank ordered 3 reference and 6 or 11 comparison electrodes. In Experiment 3 (matching), listeners adjusted the comparison electrode to pitch match the reference. Each experiment was evaluated for testing-range bias (point of subjective equality [PSE] vs. comparison-range midpoint) and reference-electrode slope bias (PSE vs. reference electrode). Discrimination showed large biases for both metrics; matching showed a smaller but significant reference-electrode bias; ranking showed no significant biases in either dimension. Ranking and matching were also evaluated for starting-point bias (PSE vs. adaptive-track starting point), but neither showed significant effects. A response-distribution truncation model explained a nonsignificant bias for ranking but it could not fully explain the observed biases for discrimination or matching. It is concluded that (a) BI-CI interaural pitch comparisons are inconsistent across test methods; (b) biases must be evaluated in more than one dimension before accepting the results as valid; and (c) of the three methods tested, ranking was least susceptible to biases and therefore emerged as the optimal approach.

## Keywords

cochlear implants, place pitch, interaural mismatch, binaural hearing

Received 12 January 2021; Revised 12 January 2021; accepted 29 January 2021

Cochlear-implant (CI) users have the potential to gain binaural advantages from the use and fitting of a second device. For many bilateral CI (BI-CI) users, however, such advantages are far smaller than those observed for normal-hearing (NH) listeners. The gap in performance may be attributable to sources of monaural and binaural impairments affecting the auditory system (e.g., asymmetric neural survival or loss of spiral ganglia; Bierer & Faulkner, 2010), as well as to the signal processing of current devices (e.g., unmatched compression functions or lack of synchronous stimulation; Kan & Litovsky, 2015; Litovsky et al., 2012). The limiting factor addressed in this study is interaural mismatch in

the cochlear sites of electrical stimulation. The binaural system is designed to compute spatial location cues or interaural differences from narrowly tuned

<sup>1</sup>National Military Audiology and Speech Pathology Center, Walter Reed National Military Medical Center, Bethesda, Maryland, United States

<sup>2</sup>Department of Hearing and Speech Sciences, University of Maryland, College Park, United States

### Corresponding Author:

Matthew J. Goupell, Department of Hearing and Speech Sciences, University of Maryland, 0241 Lefrak Hall, College Park, MD 20742, United States.

Email: goupell@umd.edu



frequency-matched inputs (Blanks et al., 2007; Fischer & Peña, 2009; Kan et al., 2013, 2015). If the two corresponding electrodes with the same frequency allocations in the two ears stimulate neural populations with different characteristic frequencies, they may not optimally excite the binaural-processing circuits located in the superior olivary complex. In these cases, adjusting frequency-to-electrode allocation in one or both ears to minimize interaural place mismatch might increase the binaural advantage for BI-CI users.

While the minimization of interaural place mismatch has clear theoretical underpinnings to suggest the possible benefits of this approach, it is unclear what is the best method to measure it. One could use a psychoacoustical measurement of binaural processing (Hu & Dietz, 2015; Noel & Eddington, 2013) or computed tomography (CT) scans (Reda et al., 2014; Zhao et al., 2014) to determine matched intracochlear electrode locations. The most widely used method to evaluate the relative place of stimulation in the two ears for BI-CI listeners, and the focus of this study, involves the identification of interaural electrode pairs that give rise to the same psychophysical percept of place pitch (e.g., Aronoff et al., 2016; Goupell, 2015; Kan et al., 2013, 2015; Litovsky et al., 2010; Long et al., 2003; van Hoesel, 2007; van Hoesel et al., 2009). Interestingly, even though the eventual goal of this approach is to improve binaural function, interaural pitch-comparison procedures are far more prevalent than binaural-processing-based psychophysical techniques (Hu & Dietz, 2015; Noel & Eddington, 2013), presumably because they are arguably less time-consuming, less technically complex, and more easily understood for many BI-CI users. In contrast, many CI users experience difficulty in perceiving the interaural-difference cues required to perform binaural tasks, and CI users with early onset of deafness appear to lack access to binaural cues (Ehlers et al., 2017; Litovsky et al., 2010).

Despite the prevalence of pitch-comparison techniques to measure interaural place mismatch, the methodology has varied greatly between studies and has not been thoroughly evaluated. The rationale for conducting interaural place-pitch measurements extends from evidence that unilateral CI listeners are sensitive to differences in the site of electrical stimulation within one ear and perceive these differences as a change in pitch (e.g., Carlyon et al., 2013; McKay et al., 1999). However, it is also well known that the perceived pitch is affected by other factors beyond the site of electrical stimulation, including the rate and level of the stimulation (Carlyon, Lynch, et al., 2010; Landsberger et al., 2016; McKay et al., 1999). When sounds vary on multiple perceptual dimensions, it makes it more difficult for a listener to attend to a single perceptual dimension, such as place pitch (Carlyon, Macherey, et al., 2010; Goupell

et al., 2019). Interaural pitch comparisons involve an additional perceptual dimension related to the fact that sounds are presented to different ears. This could further complicate the interpretation of pitch-comparison data and could be of even greater concern than rate and level effects in cases where the quality of sound differs across the ears for sequentially implanted BI-CI users, a common occurrence for postlingually deafened adults (Firszt et al., 2018).

It has been established that interaural pitch comparisons are susceptible to procedural biases, whereby changes in the stimuli presented can strongly influence the results. In particular, Carlyon, Macherey, et al. (2010) and Goupell et al. (2019) showed that for CI listeners, interaural pitch comparisons often reflected a range bias, whereby the point of subjective equality (PSE; the stimulus electrode or frequency in one ear that yields a perceptually comparable pitch to the reference electrode or frequency in the other) tends to gravitate toward the middle of the testing range. This phenomenon is similar to the centering bias defined by Poulton (1979) and also referred to as regression effect by Stevens and Greenbaum (1966). In a perceptual task where no feedback is provided, once the observer has learned the range of available responses, they select a response that is closer to the middle of the range; these responses appear to cluster in the middle of the range when the task is difficult or the cue is somewhat ambiguous. Harris (1948) evaluated the subjective pitches of individual comparison tones (up to 20 Hz different) relative to a preceding reference tone for NH listeners. In one experiment, the reference was always a 1000-Hz tone. In another experiment, the reference was shifted by 3 Hz after 40 practice trials, without communicating this to the listeners. In the critical third experiment, the reference was discontinued entirely after the first 40 trials. Psychometric functions did not differ significantly for these experiments, suggesting that listeners relied predominantly on the context of the range of comparison tones presented, and did not attend to the reference tone, at least after the first 40 trials. Subsequent experiments randomizing the reference had little effect on the resulting PSEs, again suggesting that the comparison-tone range was the most important factor.

For CI listeners, whose perception of place pitch is weaker than for NH listeners, the existence of a range bias can have a dramatic effect on interaural pitch-comparison results. Goupell et al. (2019) examined a pitch-discrimination task employing a method of constant stimuli and fitted psychometric functions to estimate an interaural PSE for multiple reference electrodes. They found that this method was particularly susceptible to range bias effects for both BI-CI and single-sided-deafness (SSD)-CI listeners, with about 2/3 of the change in PSE attributable to changes in testing range

and only 1/3 attributable to changes in the intracochlear electrode position. Goupell et al. (2019) did not, however, conduct similar evaluations of procedures whereby the place-pitch estimates are progressively refined following each stimulus presentation, either through adaptive psychophysical procedures (e.g., Long et al., 2005) or by direct listener control of the pitch of the comparison stimulus (e.g., Aronoff et al., 2016). Carlyon, Macherey, et al. (2010) identified individual cases where range effects influenced the PSE for SSD-CI listeners and NH listeners using both constant-stimuli and adaptive methods, but these approaches were not compared in a systematic way or evaluated for BI-CI listeners.

This study further expands the work from Carlyon, Macherey, et al. (2010) and Goupell et al. (2019) to validate and understand interaural place-pitch comparisons. First, the study aimed to evaluate the efficiency and susceptibility to range-bias effects for three different pitch-comparison procedures presented to the same BI-CI listeners. Whereas the results of Goupell et al. (2019) showed large range-bias effects for a discrimination procedure that employed a method of constant stimuli, we hypothesized that an adaptive procedure where the listener had some control over the stimuli presented would show less bias because the range of possible comparison stimuli would quickly converge near the PSE. Second, the study aimed to evaluate the extent to which these methods of evaluating relative interaural place pitch yielded results that were consistent with one another. Three pitch-comparison methods—*discrimination*, *ranking*, and *matching*—employed sequential pitch judgments to estimate the comparison electrode in one CI ear that yielded an equivalent percept of place pitch to the reference electrode in the other ear. In each case, pitch judgments were made for combinations of each of three reference electrode locations in one ear (apical, middle, and basal) and three ranges of comparison electrodes in the other ear (full range, apical half, and basal half).

Experiment 1 examined *pitch discrimination*. Listeners were presented with pairs of sequential single-electrode stimuli, one in each ear, and asked to judge which was higher in pitch (two-interval two-alternative forced choice). A range of electrode pairs was presented in random order (method of constant stimuli), and psychometric functions were calculated to estimate the PSE based on the 50% crossover point (e.g., Carlyon, Macherey, et al., 2010; Goupell et al., 2019; Hu & Dietz, 2015; Reiss et al., 2011, 2014; Svirsky et al., 2015).

Experiment 2 examined *pitch ranking*. This method also required listeners to judge the relative pitches of two sequential sounds in a two-interval two-alternative forced choice paradigm, but instead of constructing a

psychometric function, the procedure employed an adaptive algorithm that allowed for the rank ordering of all electrodes of interest in both ears. The electrode pairs selected for presentation on each trial depended on the previous responses (i.e., were chosen adaptively) in such a way as to require substantially fewer trials than the method of constant stimuli employed in Experiment 1. Pitch-ranking tasks have been employed in studies with unilateral CI users (e.g., Adel et al., 2019; Collins et al., 1997; Kenway et al., 2015; Vickers et al., 2016) and NH listeners (e.g., Carlyon, Macherey, et al., 2010) but have not been used with BI-CI listeners.

Experiment 3 examined *pitch matching*. While still a two-interval sequential pitch-comparison task, the major difference here was that in each testing block, the stimulus presented to one ear was fixed (i.e., the reference) and the listener controlled the stimulus in the other ear (i.e., the comparison), changing the stimulated electrode until the pitches were subjectively best matched. This approach, which is both more subjective and a more direct way of obtaining a PSE estimate, has also been employed numerous times in the literature (e.g., Aronoff et al., 2016; Carlyon, Macherey, et al., 2010; Kan et al., 2013; Tan et al., 2017).

## Methods

Data from six BI-CI users of Cochlear-brand devices (Cochlear Ltd.; Sydney, Australia) were collected and analyzed in the tasks of this study: discrimination (Experiment 1), ranking (Experiment 2), and matching (Experiment 3). Experiment 1 was performed in Goupell et al. (2019), and the data are reported here for comparison with results from Experiments 2 and 3.

### Listeners and Stimuli

The same seven BI-CI users who participated in the discrimination study of Goupell et al. (2019) were recruited for this study (Table 1). Data for listener BCI7 were collected for all three tasks but were excluded from this report due to a notably impaired ability to perform within-ear pitch comparisons in the left ear and interaural pitch comparisons (see also Goupell et al., 2019, Figures 1 and 2 for the issues with the discrimination task for listener BCI7). Stimuli consisted of 300-ms pulse trains delivered at a rate of 1,000 pulses per second (pps) using monopolar stimulation. Every pulse was biphasic with 25- or 35- $\mu$ s phase duration and 8- $\mu$ s interphase gap. All stimuli were loudness balanced. The stimuli were delivered using two bilaterally synchronized L34 research processors (Cochlear Ltd., Sydney, Australia) and controlled using MATLAB software (MATLAB 6.1, The MathWorks Inc., Natick, MA,

**Table 1.** Listener Hearing History and Demographic Information.

Listener	Age (years)	Array type		Duration of deafness (years)		CI experience (years)		Etiology		
		Left	Right	Left	Right	Left	Right	Left	Right	
BCI1	CAB	71	CI24M	CI24R	36	44	20	13	Unknown	Unknown
BCI2	CAE	65	CI24RE	CI24RE	4	3	10	11	Hereditary	Hereditary
BCI3	CAK	70	CI422	CI24R(CS)	0	0	13	1	Sinus surgery	Unknown
BCI4	CAQ	59	CI24RE	CI24RE	5	6	9	5	Meniere's	Meniere's
BCI5	CBF	59	CI24RE	CI422	5	10	7	2	Hereditary	Hereditary
BCI6	CCA	77	CI24RE	CI512	61	1	3	6	Measles, antibiotics	Measles, antibiotics
BCI7	CCI	66	CI24R(CS)	CI24RE	2	8	15	4	Possibly otosclerosis	Possibly otosclerosis

Note. CI = cochlear-implant.

2000) running the Nucleus Implant Communicator (version 2).

### Procedure

**Loudness Balancing.** Loudness balancing for all stimuli was performed in two stages. First, most comfortable levels were measured using a standard clinical loudness chart. The listeners were asked to identify “loud but comfortable” levels while ignoring the pitch of different stimuli. Second, small fluctuations in loudness between most comfortable level across electrodes within an ear were removed by sequentially stimulating sets of five adjacent electrodes, from apical to basal, and asking the listeners to report on any difference in loudness between these. Adjustments in level were made until all five electrodes were judged to be equal to the first electrode in the sequence. Then, a new set of electrodes was stimulated, where the first electrode corresponded to the last of the previous set. This procedure was repeated until all available electrodes were loudness balanced, and performed separately for the right and the left arrays.

**General Pitch Comparison Approach.** The general approach was to choose a fixed set of three “reference” electrodes in the right ear, and for each reference electrode to use sequential pitch judgments to determine which “comparison” electrode in the left ear was perceived as producing a comparable pitch, defined as the PSE. For each of the three experiments (discrimination, ranking, and matching), the possible existence of a testing-range bias was evaluated by comparing three different ranges of comparison electrodes. Within a block of trials, the testing range was held fixed, and the comparison electrode varied from trial to trial, either randomly or adaptively. For two of the experiments (discrimination and matching), the reference electrode was also held fixed throughout the block. For the ranking task where all reference and comparison electrodes were ranked relative to one another, the

nature of the task precluded the reference electrode from being held fixed throughout a block, but it was held fixed for a series of adaptive trials within the block (i.e., partial blocking). The reference- and comparison-electrode conditions were as described by Goupell et al. (2019), with three reference electrodes in the right ear (E4, E12, and E20) and the 11 even-numbered comparison electrodes in the left ear (E2, E4, E6, E8, E10, E12, E14, E16, E18, E20, and E22). The testing-range conditions involved subsets of the 11 comparison electrodes in the left ear: apical half (E12 to E22), basal half (E2 to E12), and full range (E2 to E22). The order of experiments was the same for all listeners: first discrimination, then ranking, and finally matching.

In addition to the interaural pitch comparisons that were the main focus of the study, unilateral pitch comparisons were also assessed for the discrimination and ranking experiments. For the discrimination experiment, the purpose of these within-ear pitch comparisons was to evaluate whether any observed testing-range effects were specific to the across-ear nature of the interaural pitch-comparison tasks. The unilateral pitch-comparison tasks were carried out for both ears, with three different reference electrodes per ear (E4, E12, E20; see Goupell et al., 2019). For the ranking experiment, the results were used to order the electrode positions on the virtual dial for the matching task. All active even-numbered electrodes were tested. Therefore, there were a total of 27 bilateral conditions (3 reference electrodes  $\times$  3 testing ranges  $\times$  3 experiments) and 4 unilateral conditions (left and right ear  $\times$  2 experiments).

The unilateral results are omitted from the current report because of the following reasons. First, the main focus was on the interaural pitch comparisons. Second, the unilateral data were relatively straightforward (i.e., generally monotonic functions with an occasional place-pitch reversal, which is expected).

**Experiment 1: Discrimination.** Goupell et al. (2019) evaluated pitch discrimination for seven BI-CI listeners. Data



from that study were compared with data from the same listeners tested in Experiments 2 (ranking) and 3 (matching) described later. In addition to the three reference-electrode and three testing-range conditions that were common to all three experiments, Goupell et al. (2019) also investigated whether presenting the three reference-electrode conditions in separate test blocks or in a randomly interleaved fashion within each test block affected the results and found no significant effect. Only the blocked data were included here, as the two other experiments employed a fully (matching) or partially (ranking) blocked procedure. Thus, the raw pitch-judgment results (i.e., percent higher) and PSE estimates derived from the psychometric functions deviate slightly from the exact values reported by Goupell et al. (2019), because the blocked and randomized results were averaged in that report. Also, data were collected using both left and right ears separately in listener BC14. For this study, we only analyzed the data for the right-ear reference to be consistent with the other listeners.

A two-interval two-alternative forced-choice procedure required listeners to indicate whether the second sound was higher or lower in pitch relative to the first sound while ignoring any loudness differences or ear of presentation. Thirty trials were collected for every condition. For more details, refer to Goupell et al. (2019).

**Experiment 2: Ranking.** In this experiment, listeners were also asked to judge the relative pitches of two sounds presented sequentially to a pair of electrodes. The main difference from the discrimination task (Experiment 1) was that ranking employed an efficient adaptive algorithm used to rank order the perceived pitches associated with a given set of electrodes. In this algorithm, known as mid-point comparison, the electrode pair chosen on a given trial depended on the previous response (Long et al., 2005; Steinhaus, 1950), with the range of possible ranks reduced by half on every subsequent trial. This results in a more efficient approach to infer the required pitch information (Long et al., 2005) than for a full method of constant stimuli (Experiment 1) or for a traditional adaptive algorithm with a fixed step size.

Each test block consisted of a series of pairwise pitch comparisons designed to rank order all the electrodes in question. For the unilateral conditions (results not shown), listeners ranked the 11 even electrodes from one ear. For the interaural pitch comparisons, all 3 reference electrodes and 6 or 11 comparison electrodes (depending on the testing-range condition) were rank ordered. The adaptive algorithm made no distinction between reference and comparison electrodes, treating all 9 or 14 electrodes the same for the purposes of the ranking procedure. This meant that on any given trial, a listener could be asked to compare two electrodes in the same ear or two electrodes from opposite ears.

On the first trial of a test block, two electrodes were selected at random for sequential presentation and ranked based on the listener's response as to which was perceived as higher in pitch. Then, a third electrode was selected and ranked within the existing list with an adaptive series of pairwise comparisons. Subsequent electrodes were then selected at random and ranked within the existing list until all 9 or 14 electrodes were included in the ranking. Unlike the discrimination task, an additional button was provided to allow the listener the option to repeat the stimulus presentation, with the idea that it was desirable to limit errors due to inattention that could have a cascading effect on the ranking of subsequent electrodes in a given block.

The procedure for ranking a new randomly chosen electrode within a list of already ranked electrodes obeyed the following rules. For the first pairwise comparison, the new electrode was paired with the electrode at the midpoint of the current ranked list. If there was an even number of electrodes in the current ranking, then one of the two middle electrodes was chosen at random. Based on the listener's response, the ranked list was then cut in half, and the next electrode to be compared was selected to be the midpoint of the remaining list. This process continued until the new electrode was ranked between two neighboring electrodes in the existing list. This is equivalent to performing a binary search on the electrode space, requiring  $\log_2(N)$  iterations on average (instead of  $N$  for linear search).

**Experiment 3: Matching.** In this experiment, the listener adjusted the comparison electrode to match the pitch of a reference electrode. Like the other experiments, matching involved the listener making pitch comparisons for sequentially presented stimulus pairs; but in this case, the comparison electrode was controlled by listener instead of being selected at random or via an adaptive algorithm. Therefore, we hypothesized that listeners would quickly converge toward the PSE, much like the ranking procedure, and thus experience fewer range effects. Specifically, whether control over the stimulus provides an advantage in reduced range effects or disadvantage from more time spent refining matches was evaluated.

For each test block, the reference electrode was held constant (as in Experiment 1), and the comparison electrode varied from trial to trial as the listener turned a virtual representation of a dial using the computer mouse. Following each presentation of the sequential stimuli, the listener adjusted the dial to shift the comparison electrode by zero, one, or two steps. The listener completed as many trials as they wanted, until they deemed the comparison electrode to be best matched in pitch to the reference electrode, which they indicated

by pressing a virtual button. This initiated a new block with a different randomly chosen reference electrode.

The number of steps shown on the dial was the number of elements for each testing-range condition, either 6 for the apical and basal ranges or 11 for the full range. The elements on the dial were arranged clockwise using the ordering produced by the unilateral results of the ranking experiment. This differs from the conventional approach to measure pitch matching in CIs (e.g., Aronoff et al., 2016), where the stimuli are arranged according to their position in the array (e.g., 1 to 22). The reason to do this was to minimize listener confusion about the perceptual place-pitch scale due to potential pitch reversals, which can occur in place-pitch discrimination with CI users (e.g., Goupell et al., 2019, BCI5 in Figure 2; Macherey & Carlyon, 2012). However, unilateral pitch reversals occurred rarely in practice, meaning that this step was largely unnecessary. Other than subject BCI7 whose data were excluded from the analysis, a unilateral reversal occurred for only one of the six listeners and only involved a single out-of-order electrode.

Eight blocks were completed for each combination of reference electrode and testing-range condition, and the PSE was calculated by averaging across the eight blocks. Two different initial comparison electrodes were used to allow for a *bias check* of whether the PSE was immune to changes in starting point (Carlyon, Macherey et al., 2010). For the first four blocks, the initial comparison electrode (i.e., the starting point on the dial) was set at the most basal end of the testing range, which was either electrode E12 (apical half) or E2 (full range or basal half). For the last four blocks, the starting point was 40% of the distance from the basal to the apical end of the range, which corresponded to electrode E6 (basal half), E16 (apical half), or E10 (full range).

## Results

The methodology for extracting the PSE differed across experiments. For illustration, Figure 1 shows an example of the results for the three experiments for one representative listener (BCI4). In each panel, the horizontal axis represents the range of comparison electrodes (except for ranking, where both the reference and comparison electrodes are represented by the horizontal axis), and the vertical axis represents the response outcome specific to each experiment. Each row of panels represents the results from one experiment, each column represents a different testing range, and each symbol/color combination represents a different reference electrode. The gray-shaded areas represent comparison electrodes that were outside the testing range for a given condition.

### Individual Listener Example

**Discrimination.** For discrimination (Experiment 1), the PSE was extracted from the psychometric function as described by Goupell et al. (2019). The response distribution (where the comparison electrode was judged higher in an electrode pair) was fitted using a maximum likelihood logistic function (Wichmann & Hill, 2001a, b). For every fitted function, the 50% crossover point was computed. There were five cases (7% of the total) where the performance function did not include 50% (i.e., the range of responses was limited to values above or below 50%). For the purpose of the current study, it was critical to have PSEs for the following: apical range and references E20 and E12; full range and references E20, E12, and E4; and basal range and references E12 and E4. For this set of conditions, there was one case (BCI1, apical range, reference 12) where the 50% crossover point was beyond the comparison range. In this single case, the PSE was defined as the 50% crossover point derived after extrapolating the fit outside of the comparison range. Note that this PSE was included in the current study to provide a full set of PSEs for all conditions, in contrast to its omission from the summary report in Goupell et al. (2019, Figure 3).

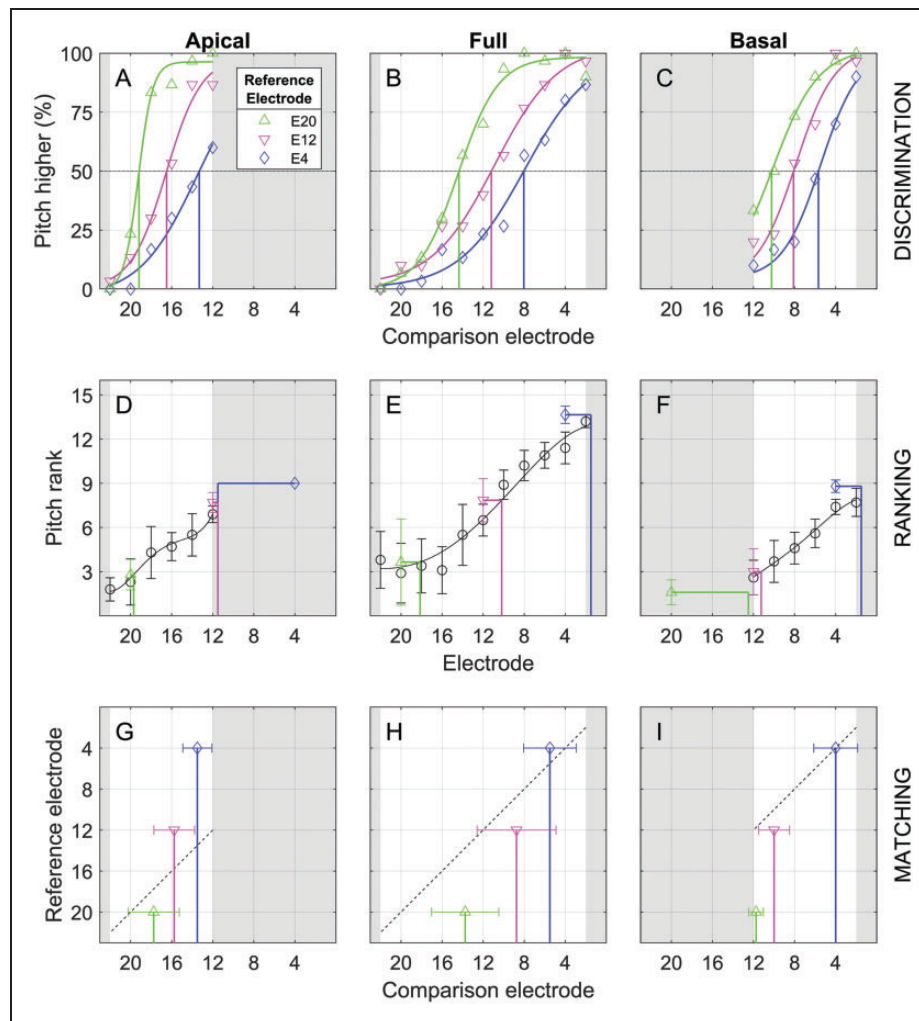
The top row of Figure 1 (Panels A–C) shows the raw discrimination data for example listener BCI4 together with the fitted psychometric functions. The vertical axis represents the percentage of trials for which the listener reported the comparison electrode to be higher in pitch than the reference electrode. The vertical lines in each panel represent the estimated PSE based on the 50% crossover point. These example data show two main trends, which are described in detail in Goupell et al. (2019). First, there was a substantial testing-range effect, as evidenced by the fact that for a given reference electrode, the three testing-range conditions yielded different PSE estimates (compare the same color across the three panels). Second, even though the data show that the PSE was different for each reference electrode (i.e., the three curves in each panel do not overlap), the magnitude of the shift was smaller than expected. An eight-electrode change in the reference electrode (from E20 to E12 or from E12 to E4) led to an average shift of 4.8 electrodes in the estimated PSE.

**Ranking.** For ranking, PSEs were estimated by determining the comparison electrode that, on average, was given the same pitch rank as the reference electrode in question. An illustration of this calculation is shown in the middle row of Figure 1 (Panels D–F). Each panel plots the average rank for each of the three reference electrodes (green, magenta, and blue symbols) and 6 or 11 comparison electrodes (black circles). The comparison-electrode ranks were fit with a fourth-order polynomial

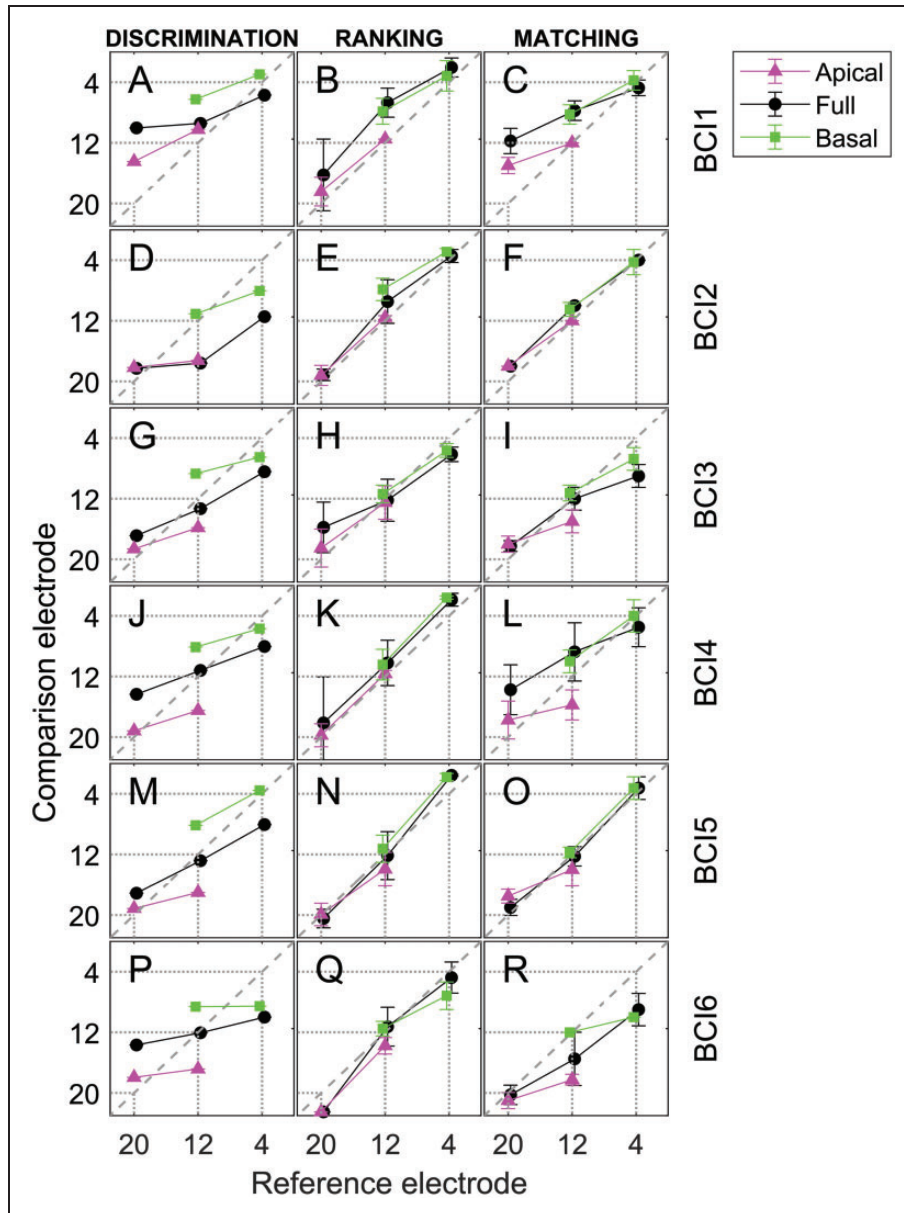
function that was constrained to be monotonic. Based on this fitted function, the PSE (vertical lines) was defined by drawing a horizontal line from the reference-electrode rank to the fitted curve to identify the comparison electrode that yielded an equivalent pitch rank. In cases where the PSE would have fallen outside the available range of comparison electrodes, the PSE was defined as 0.5 electrodes past the end point of the range (e.g., Figure 1D, reference electrode E4, blue diamond, and horizontal blue line). Like the discrimination data, these ranking data show that the

PSE estimate changed as a function of reference electrode. Unlike the discrimination data, however, the magnitude of the change was relatively closer to the expected eight-electrode shift between reference-electrode conditions compared with the discrimination task (see especially Figure 1E).

*Matching.* The bottom row of Figure 1 (Panels G–I) shows the mean and standard deviation of the eight PSE estimates for each reference-electrode and testing-range condition. The diagonal line in each panel



**Figure 1.** Example Pitch-Comparison Results for Listener BCI4. The apical (left column), full (middle column), and basal (right column) testing ranges are shown for the three experiments. The shaded areas indicate comparison electrodes that were outside the testing range. Vertical lines indicate the estimated PSE for each reference electrode. (A–C) Pitch discrimination: percentage of trials where the comparison electrode was reported as higher than the reference electrode. PSE estimates (vertical lines) were derived from the 50% point on the fitted psychometric function (solid curves). (D–F) Pitch ranking: average rank for each reference electrode (colors) and comparison electrode (black circles). PSE estimates (vertical lines) were derived by extracting from the fitted function (curves) the comparison electrode yielding the same average rank (horizontal lines) as a given reference electrode. In cases where the reference rank occurred outside the range of ranks for the comparison range (e.g., reference E4 in all three panels), the PSE was taken to be 0.5 electrodes past the end of the range. Error bars show  $\pm 1$  standard deviation. (G–I) Pitch matching: average pitch-matched comparison electrode number (vertical lines). The diagonal dashed black lines show where the matches would be if each reference electrode were matched to the comparison electrode with the same number. Error bars show  $\pm 1$  standard deviation.



**Figure 2.** PSE Estimates for Individual Listeners. Different experiments (columns), listeners (rows), and testing conditions (symbols and colors) are plotted. In the absence of range effects, pitch comparisons are not affected by the testing range (apical, full, or basal) and yield overlapping PSE estimates (e.g., BCI6 tested on pitch ranking); conversely, vertical offsets in PSE estimates for the same reference electrode are indicative of range bias (e.g., BCI6 tested on pitch discrimination). The dashed diagonal line represents a PSE at the number-matched electrode. Dotted vertical and horizontal lines indicate the positions of the three reference electrodes. Error bars for the ranking and matching experiments indicate  $\pm 1$  standard deviation.

represents the case where the matched comparison electrode equals the number of the reference electrode in the other ear (i.e., no mismatch). The fact that across reference electrodes, the pitch estimates followed a trajectory that was steeper than the diagonal line means that, like the discrimination experiment, the change in the matched comparison electrode (horizontal axis) was smaller than the eight-electrode shift that would be

expected with an eight-electrode change in reference electrode (vertical axis).

### Summary of All Individual Listener Results

Figure 2 shows the PSE estimates for each individual listener (rows) and experiment (columns). Within each panel, the horizontal axis represents the reference

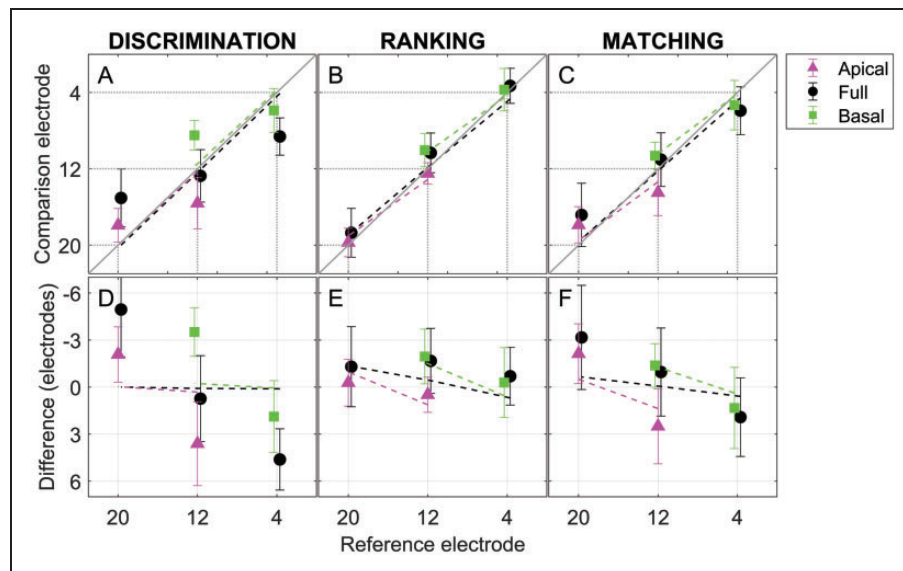


electrode, and the vertical axis represents the comparison electrode yielding the estimated PSE. Each color/symbol combination represents a different testing-range condition, and the diagonal line represents the case where the matched comparison electrode number equals the reference electrode number. Estimates for the two conditions where the PSE was expected to fall far outside the testing range (i.e., the apical half for reference-electrode E4 and the basal half for E20) were not included in the plot or any further analyses because these conditions led to nonsensible results in most cases (e.g., blue diamonds in Figure 1D and G). Figure 3 shows the PSE estimates averaged across the six listeners (the points represent the data, while the dotted lines represent the results of model simulations discussed later). The top row of Figure 3 plots the absolute PSE, while the bottom row plots the difference between the comparison-electrode PSE and reference-electrode number.

This study posed two main and related questions: (a) whether each pitch-comparison approach would be influenced by a testing-range bias, and (b) whether a given change in reference electrode would be reflected by a similar change in the PSE estimates. The fact the PSE estimates differed across testing-range conditions for the discrimination experiment (Figures 2 and 3, left column) and to some extent for the matching experiment (Figures 2 and 3, right column) suggests that these two methods were susceptible to a testing-range bias. The fact that the slopes of the PSE functions were shallow

(slope < 1), especially for the discrimination experiment (Figures 2 and 3, left column) and to some extent for the matching experiment (Figures 2 and 3, right column), suggests that these two methods failed to capture a similar eight-electrode shift in comparison-electrode PSE that might be expected for an eight-electrode change in reference electrode. This can also be easily seen in the bottom row of Figure 3, where the difference between the comparison-electrode PSE and reference electrode was often different from zero (Figures 3D and F). In contrast, the PSE estimates for the ranking experiment (Figures 2 and 3, middle column) were closely (though not perfectly) overlapping and had a slope closer to one for all three testing-range conditions. This suggests that the PSE estimates from the ranking task primarily reflected the pitch associated with the reference electrode independent of testing range.

Repeated-measures mixed-model linear regression analyses were carried out on the PSE estimates (Figure 3, top row) to examine the effect of testing range. These models included testing range, reference electrode, and their interaction as fixed effects and listener as a random effect. There were significant main effects of testing range and reference electrode for all three experiments ( $p < .0001$  for both), and a significant interaction between the two variables for ranking ( $p < .0001$ ) and matching ( $p = .017$ ), but not for discrimination ( $p = .46$ ). To explore these interactions, separate post hoc mixed models were generated for each reference electrode, with Bonferroni corrections applied for three



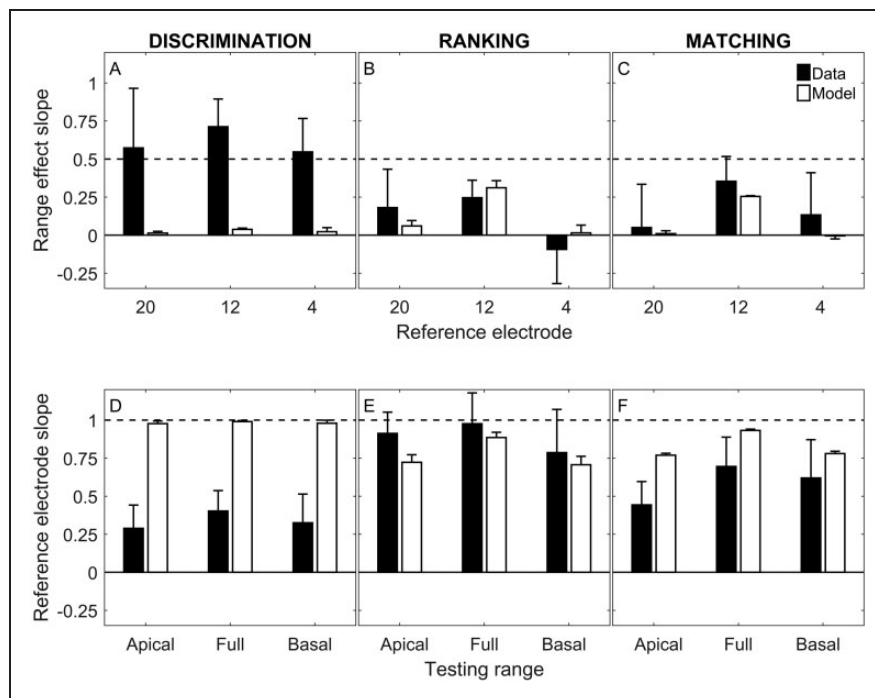
**Figure 3.** Group-Average Pitch-Match Estimates. (A–C) Group-average PSE estimates (symbols) and truncation model simulations (colored dashed lines; see text and Figure 6 for details). The solid diagonal line represents a PSE at the number-matched electrode. Dotted vertical and horizontal lines indicate the positions of the three reference electrodes. (D–F) Group-average differences between the PSE and number-matched comparison electrodes (symbols) and model simulations (colored dashed lines). A value of zero on the vertical axis represents a PSE at the number-matched electrode. Error bars show  $\pm 1$  standard deviation.

comparisons (i.e., criterion  $p = .0167$ ). For the discrimination experiment, all three reference electrodes showed a significant effect of testing range—E4:  $\chi^2(1) = 12.8$ ,  $p = .0004$ ; E12:  $\chi^2(2) = 27.8$ ,  $p < .0001$ ; E20:  $\chi^2(1) = 7.63$ ,  $p = .0057$ ; for the ranking experiment, only E12 showed a significant effect— $\chi^2(2) = 46.0$ ,  $p < .0001$ ; and for the matching experiment, E12— $\chi^2(2) = 84.7$ ,  $p < .0001$ —and E20— $\chi^2(1) = 7.24$ ,  $p = .0071$ —showed a significant effect.

For an interaural pitch-comparison procedure to be acceptable, PSE estimates should reflect changes in the reference electrode and not in the testing range. The black bars in Figure 4 characterize these two effects (the white bars reflect the results of model simulations presented later). The top row (Figure 4A–C) plots the slope of the testing-range effect, defined as the change in PSE compared with the change in the midpoint of the testing range (apical: E7; full: E12; basal: E17). For reference electrode E12, the slope was calculated by fitting a line across the three testing ranges. For reference electrodes E4 and E20, the apical or basal testing range, respectively, was excluded as in Figures 2 and 3. The horizontal dashed lines in Figure 4A to C indicate a criterion slope of 0.5, below which Carlyon, Macherey et al. (2010) argued that a PSE estimate can be

considered as acceptable. The average range effect was generally larger than 0.5 for discrimination (Figure 4A), but smaller than 0.5 for ranking (Figure 4B) and matching (Figure 4C). A linear mixed-model regression analysis conducted on the testing-range slopes found significant main effects of experiment— $\chi^2(2) = 29.3$ ,  $p < .0001$ —and electrode— $\chi^2(2) = 9.05$ ,  $p = .011$ , but no significant interaction between the two variables— $\chi^2(4) = 2.08$ ,  $p = .72$ . Post hoc tests showed that the average range-effect slope was significantly greater than zero for discrimination (average slope = 0.61;  $p < .0001$ ) and for matching (average slope = 0.24;  $p = .0012$ ), but was not significantly different from zero for ranking (average slope = 0.12  $p = .075$ ). The range-effect slope was significantly larger for discrimination than for matching ( $p = .0001$ ) and ranking ( $p < .0001$ ), but was not significantly different between ranking and matching ( $p = .20$ ).

To quantify the effect of changing the reference electrode, the bottom row (Figure 4D–F) plots the relationship between the change in PSE and the change in reference electrode. The horizontal dashed lines in Figure 4D–F indicate a slope of 1 (i.e., the value expected if the PSE perfectly reflected the change in reference electrode). The reference-electrode slope was much smaller than 1 for discrimination (Figure 4D),



**Figure 4.** Effects of Testing Range and Reference Electrode on PSE Estimates. (A–C) The relationship between the change in PSE and the change in the center of the testing range. The horizontal dashed line indicates a criterion slope = 0.5, below which the range effect is proposed to be acceptably small (Carlyon, Macherey, et al., 2010). (D–F) The relationship between the change in PSE and the change in reference electrode. The horizontal dashed line indicates the slope = 1 expected in the case where the PSE in the comparison ear is equal to the reference electrode. Black bars indicate group-average slopes ( $\pm 1$  standard deviation) derived from the data in Figures 2 and 3. White bars indicate the results of the truncation model (see text and Figure 6 for details).

close to 1 for ranking (Figure 4E), and intermediate between the other two experiments for matching (Figure 4F). A linear mixed-model regression analysis conducted on the reference-electrode slopes found significant main effects of experiment— $\chi^2(2) = 51.4$ ,  $p < .0001$ —and testing range— $\chi^2(2) = 6.09$ ,  $p = .048$ , but no significant interaction between the two variables— $\chi^2(4) = 6.98$ ,  $p = .14$ . Post hoc tests showed that the average reference-electrode slope was significantly smaller than 1 for discrimination (average slope = 0.34;  $p < .0001$ ) and for matching (average slope = 0.59;  $p < .0001$ ), but was not significantly different from 1 for ranking (average slope = 0.89;  $p = .059$ ). The reference-electrode slope was significantly smaller for discrimination than for matching ( $p = .0002$ ) and ranking ( $p < .0001$ ), and significantly smaller for matching than for ranking ( $p < .0001$ ).

In summary, there were considerable differences in procedural bias across the three pitch-comparison experiments. Discrimination was most susceptible to biases, showing a reference-electrode slope much smaller than 1 and a testing-range effect slope greater than 0.5. Matching showed a smaller bias than discrimination. Ranking showed no significant bias.

### Analysis of Starting Point

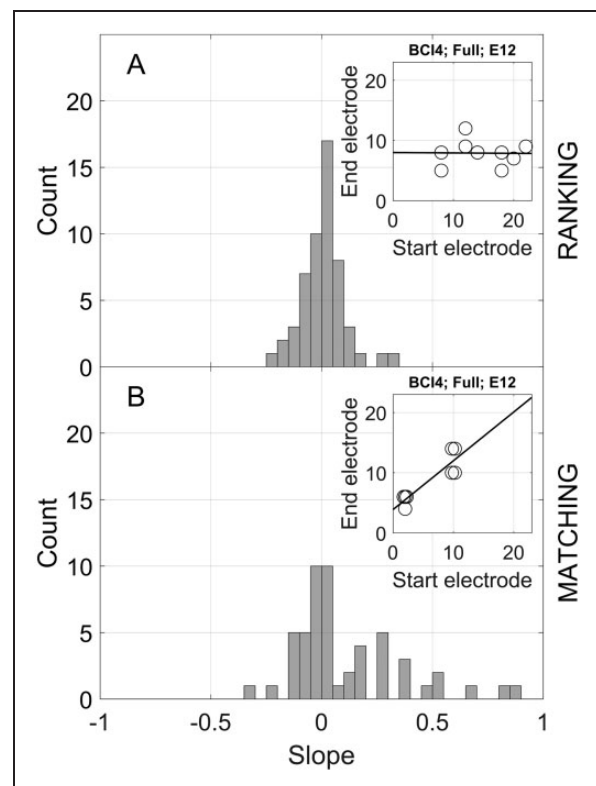
In addition to the testing-range effect described earlier, Carlyon, Macherey, et al. (2010) proposed a second bias check for adaptive pitch-comparison paradigms that involves the starting point of the procedure. On one hand, if two different comparison-electrode starting points yield the same PSE estimate, the experimenter can conclude that this estimate is reliably providing information regarding the perceived pitch of the reference electrode. On the other hand, if the PSE depends on the starting point, this can be taken as evidence that the estimate is contaminated by the characteristics of the order of presentation of comparison electrodes. Therefore, starting-point effects were examined for both the ranking and matching methods, both of which involved adaptive tracking.

Starting-point effects were calculated for each listener, electrode, and experiment by fitting a scatterplot of the PSE estimates versus starting comparison electrode with a straight line (see inset examples for listener BCI4, full range, reference electrode E12 in both panels of Figure 5). For the ranking experiment (Figure 5, top panel), the starting comparison electrode was random. For the matching experiment (Figure 5, bottom panel), there were two possible starting comparison electrodes. Figure 5 plots histograms of the slopes, expressed as the change in PSE relative to the change in the starting comparison electrode. Across the six listeners, three reference electrodes, and three testing-range conditions, the

starting-point slope was not significantly different from zero for ranking, mean = 0.009,  $SD = 0.097$ ,  $t(53) = 0.68$ ,  $p = .50$ . However, the mean slope was significantly different from zero for matching, mean = 0.11,  $SD = 0.25$ ,  $t(53) = 3.2$ ,  $p = .002$ . Carlyon, Macherey, et al. (2010) suggested a starting-point slope criterion of 0.5, above which the starting point could be interpreted as having a meaningful influence on the PSE. Figure 5 shows that for the vast majority of cases (100% of electrodes tested for ranking and 91% for matching), the estimated slopes were less than 0.5.

### Range Bias or Truncation Effect?

The results shown in Figures 2, 3, and 4 are suggestive of a testing-range bias that was largest for the discrimination experiment, smaller but still significant for the matching experiment, and nearly nonexistent for the ranking experiment. While these results suggest that the testing range influenced the response distribution, another possibility is that the trend might instead be a result of an edge effect, also called a distribution truncation effect. The idea is that with a limited range of



**Figure 5.** Starting Point Analysis. The distribution of the slopes of the regressions of the PSE estimates as a function of starting comparison electrodes pooled across all listeners and conditions for (A) pitch ranking and (B) pitch matching. The inset in each panel shows an example of the slope derivation for listener BCI4, the full testing range, and reference electrode E12.

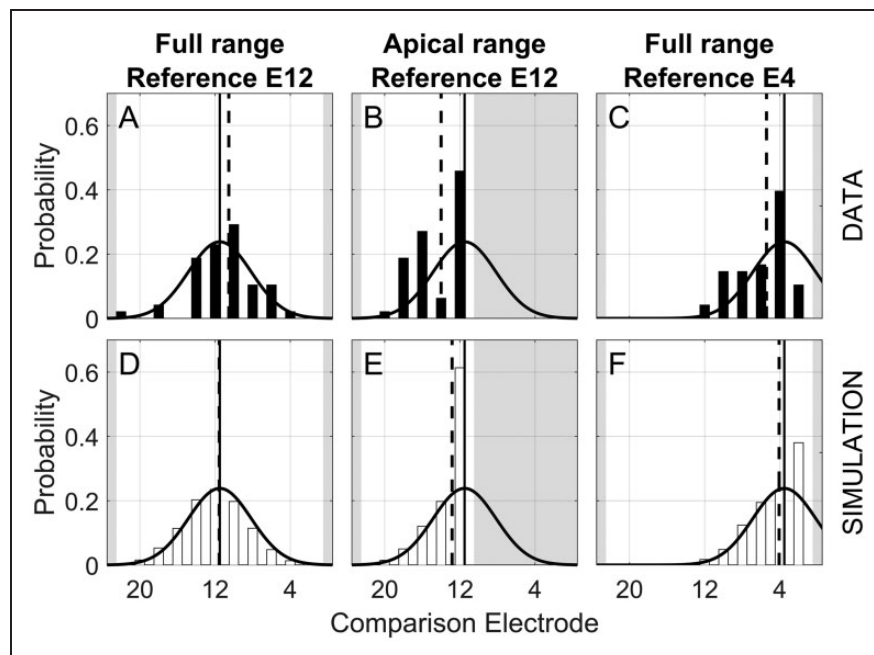
possible comparison electrodes, there is no possibility that a listener could provide a response outside this range. This would skew the distribution of responses away from the edge of the comparison range.

The influence of truncation is illustrated in the top row of Figure 6, which shows example distributions of PSEs for the matching experiment, pooled across listeners, for three example combinations of reference electrode and testing range. Figure 6A represents the condition where the full range of comparison electrodes was available and the reference electrode (E12) was in the middle of the array. Here, the distribution of responses was not constrained by the lack of response options and the distribution followed a Gaussian form extending over most of the testing range. The other two panels illustrate cases with truncated response distributions. In Figure 6B, the testing range has been cut in half. In Figure 6C, the reference electrode (E4) is near the apical end of the array. In these cases, the simple arithmetic mean of the truncated response distribution yields a PSE estimate (dashed vertical line) that deviates from the mean (solid vertical line) of the underlying (now truncated) Gaussian distribution (solid curve).

There are at least two possible explanations for observed shifts in the PSE associated with a change in

comparison-electrode range. According to the *testing-range bias* interpretation, rather than responding based on the percepts associated with the comparison and reference stimuli on a given trial, the listener's response is biased by their knowledge of the available range of only the comparison stimuli. According to the *truncation-effect* interpretation, the *actual* mean response is unaffected by the change in testing range, but the measured PSE is shifted by the truncation of the response distribution, resulting from the limited possible response range. Therefore, distinguishing a truncation effect from a true range bias requires a comparison of PSE distributions.

We modeled the response distributions to determine whether truncation could account for any apparent testing-range effects observed in the empirical data. Ideally, this analysis would fit an underlying truncated Gaussian function to the observed response distributions. In practice, this was not possible due to several factors, including a limitation in the amount of data available, the fact that a broad range of truncated Gaussian functions could reasonably fit a given distribution, and complications regarding how out-of-range matches should be reflected in the assumed distributions. Instead, Monte Carlo simulations of the experimental procedures were developed to generate empirical



**Figure 6.** Truncation-Model Simulations of Response Distributions. Distributions of PSE estimates pooled across all listeners for three example combinations of reference electrode and testing range (columns) for pitch matching. The top row shows experimental data; the bottom row shows model simulations. The shaded areas indicate comparison electrodes that were outside the testing range. Gaussian curves fit to the data, representing the underlying nontruncated distributions, are identical for each data-simulation pair. The truncation bias is reflected in the difference between the mean of the fitted distributions (solid vertical lines) and the arithmetic mean of the data (dashed vertical lines).



distributions based on a parameterization of the assumed underlying Gaussian response distributions. Each electrode was assumed to generate an underlying Gaussian distribution of pitch percepts (defined in terms of electrode number). A given simulated trial involved drawing pitch percepts from the underlying Gaussian distributions for the reference and comparison electrodes. The simulated response was based on which of these two percepts was higher, thereby generating a distribution of modeled PSEs the same way as in the actual experiment. Importantly, the simulations assumed no testing-range bias: the mean of the underlying Gaussian distribution was assumed to be the same regardless of the testing-range condition and was fixed at a value equal to the reference electrode number. Thus, this approach allowed us to determine whether the apparent differences in PSE across testing-range conditions (Figures 2 and 3) could be explained by a truncation. Note that this approach did not simulate any interaural frequency mismatch between the cochlear places of stimulation in the two ears. Because the interaural frequency mismatch is unknown, we cannot distinguish between a true interaural frequency mismatch and a bias in the mean response. However, it is likely that on average across listeners, any mismatch would have been close to zero because mismatch for a given listener would have been equally likely to occur in either direction.

The model output was fit to the data (minimum least-squares difference in PSE) using only a single free parameter to model internal Gaussian noise in the pitch percept. While it is possible that this variance could change across electrodes and ears, we elected to use the same standard deviation for all electrodes in both ears to limit model complexity. Furthermore, the model variance was constrained within the 95% confidence interval of the variance in the data for reference E12 with a full testing range (i.e., the condition where edge effects were least likely to affect the response distribution). While we also used this approach to fit the data separately for each listener (not shown), the model results that are shown in the following were produced by pooling the data across all listeners, which allowed for a better characterization of the shape of the response distributions given the limited number of trials.

Examples of the model-simulated response distributions are shown in the bottom row of Figure 6. The Gaussian functions in each panel of Figure 6 represent the results of the fitted model and are identical in the corresponding panels in the top and bottom rows. The simulated response distributions (bars in Figure 6D–F) showed similar trends to the actual response distributions (bars in Figure 6A–C). The distributions tended to be most truncated in the second column where the available response range was the most severely limited. As result, the fitted underlying Gaussian functions in

Figure 6E and F had a different mean (dotted vertical lines) than the arithmetic average of the individual PSE estimates (solid vertical lines).

The results of the model simulations for each experiment, reference electrode, and testing range are shown in Figure 3 (dashed lines) and Figure 4 (white bars) in comparison to the experimental data (Figure 3, solid lines; Figure 4, black bars). For the discrimination experiment, the PSE estimates reflect the 50% point of the psychometric functions fit to the Monte-Carlo-simulated experiment. Here, the model produced a poor fit and failed to capture the range effects observed in the experimental data (Figure 3A). In Figure 3D, whereas the experimental data showed large differences between the PSE and the reference electrode in many conditions (points), the model predicted zero differences (dashed lines), and the root-mean-square (rms) difference between the data and model simulations was large (3.29 electrodes). Consequently, the model (Figure 4A, white bars) was unable to account for the large range effect and shallow reference-electrode slope observed in the experimental results (black bars).

For the ranking experiment, the model showed a close correspondence to the PSE data (Figure 3B) and mostly captured the small deviations (Figure 3E) between the PSE estimates and the reference electrode number for the apical and basal ranges. The rms difference between the data and model simulations was small (0.98 electrodes). The model simulations also showed estimates of the magnitude of the range effect (white bars, Figure 4B) and reference electrode slope (white bars, Figure 4E) that were similar to the experimental data (black bars). In summary, any small (albeit nonsignificant) deviations of a range-effect slope above the ideal value of zero and a reference-electrode slope below the ideal value of 1 were accounted for by the truncation effect.

For the matching experiment, the model captured some of the deviation (Figure 3F) between the PSE estimates and the reference-electrode number for the apical and basal ranges (green and magenta symbols and dashed lines). However, the simulation did not account for these deviations in the full-range condition (black circles), instead predicting a PSE that was equal to the reference electrode number (black dashed lines, Figure 3C and F). As a result, the rms difference between the data and model simulations was 1.43 electrodes, slightly larger than for the ranking experiment. While the model accounted for the small range effect (Figure 4C, compare black and white bars), it predicted a reference-electrode slope that was steeper (i.e., closer to the ideal value of 1) than was observed in the experimental data (Figure 4F). In summary, the matching experiment showed a small range effect that was accounted for by truncation but also a bias in the mean of the response distribution

toward the center of the range that could not be accounted for by truncation.

Overall, the model results shown in Figures 3 and 4 demonstrate very different effects for the three experiments. Because the simulations assumed only a truncation effect, this allowed us to distinguish truncation from a true testing-range bias. Discrimination showed large range-effect and reference-electrode biases. According to the discrepancy between data and model, these biases were interpreted to reflect true range effects. Ranking did not show any significant range-effect or reference-electrode biases. Any small, nonsignificant trend toward a bias was accounted for by truncation and therefore was interpreted to not indicate a true range effect. Matching showed a range-effect bias that was accounted for by truncation, but a reference-electrode bias that was not accounted for by the truncation model and therefore was interpreted to indicate a true, albeit small, range bias. Thus, only the ranking experiment was immune to range biases independent of truncation.

### *Duration of Testing*

Another important factor in deciding which pitch-comparison method to use is the amount of time required to complete the measurements. Listeners in this study completed many more trials in the discrimination experiment, and as a result, the experiment was considerably longer in this case. For discrimination, each listener completed an average of 2,040 trials at an average rate of 6.6 s/trial, for a total testing time of 3 hr 48 min. For ranking, listeners completed an average of 929 trials at an average rate of 5.9 s/trial, for a total testing time of 1 hr 31 min. For matching, listeners completed an average of 750 trials at an average rate of 8.0 s/trial, for a total testing time of 1 hr 40 min. Thus, the lower number of total trials needed in the matching experiment was offset by the longer trial duration relative to the ranking experiment. Note that these time estimates are based on the full combination of three electrodes and three testing ranges. If only the full testing range had been examined, the total testing time would have been approximately halved. Also note that each experiment was designed to produce reliable data based on previous experiments of these types with the intention of having roughly comparable variability in results across tasks; obviously, testing time could have been altered by different design choices. Finally, listeners were given sufficient breaks (time and duration chosen by the listener) to avoid fatigue during the experiments. Although we cannot rule out the possibility that fatigue could have influenced the results, it is unlikely to have differentially affected the range effects in the individual experiments. Even though the discrimination experiment was longer than the other two experiments, an analysis

conducted on only the first half of the discrimination data collected (not shown) showed the same pattern of results.

## **General Discussion**

All sensory measurements are subject to unintended biases (Poulton, 1979; Stevens & Greenbaum, 1966). Pitch-based assessments of interaural place of stimulation in CI users seem particularly prone to such methodological biases, which include testing-range and possibly starting-point effects (Adel et al., 2019; Carlyon, Macherey, et al., 2010; Goupell et al., 2019). It was recently shown that for BI-CI listeners, a discrimination task was highly susceptible to methodological biases, such that the PSE in the comparison ear was more heavily dependent on the testing range than on which electrode was stimulated in the reference ear (Goupell et al., 2019). The question posed in this study was whether these methodological biases would be pervasive across three pitch-comparison methods that have been used in the literature: discrimination, ranking, and matching.

### *Summary of the Experiments*

Figures 2, 3, and 4 show that procedural biases varied considerably across methods. The top row of Figure 4 shows how adjusting the comparison range altered the PSE for each task. The range effect was largest for discrimination, relatively smaller for matching, and almost nonexistent (and not significant) for ranking. It is important to point out that any observed testing-range effects cannot be explained by a large mismatch between the PSE and the available testing range. Conditions where the reference electrode did not overlap with the contralateral testing electrode range (e.g., reference electrode E4 and apical comparison range) were intentionally excluded in Figures 2, 3, and 4.

The bottom row of Figure 4 illustrates the relationship between the PSE and changes in reference electrode. If we assume that, on average, the BI-CI listeners had little interaural mismatch in the cochlear place of stimulation for a given electrode pair, then we would expect that a change in reference electrode (e.g., from E20 to E12) would yield an equivalent (i.e., eight-electrode) change in the PSE. In other words, we would expect a reference-electrode slope equal to 1. For discrimination, this was clearly not the case, with an average reference-electrode slope of only about 0.3 electrodes/electrode. For matching, the slope was about 0.6 electrodes/electrode, but still significantly <1. Only ranking yielded a slope close to and not significantly different from 1. This means that the ranking experiment was the only case where any biases were small enough to not significantly shift the PSE away from the expected value.

Model simulations showed that the nature of the testing-range effect differed across the experiments. Through Monte Carlo simulations, we were able to distinguish between what constituted a *true* range effect, where the listener response is weighted based on knowledge of the available range of comparison stimuli from previous trials, and a *truncation-based* edge effect, where the response is based only on the relationship between the comparison and reference stimulus on a given trial. For a true testing-range effect, the center of the Gaussian distribution of responses actually shifts as the testing range is varied. For a truncation-based edge effect, the Gaussian response distribution falling near the edge of the range is truncated. In this case, the listener's mean PSE is not affected by the testing range but only appears to shift when taking the mean of the truncated distribution (see Figure 6).

For discrimination, truncation could not account for any of the observed range effects (Figures 3A, D and 4A, D). In this experiment, the PSE was derived by fitting a psychometric function to a set of data points. This approach is not affected by truncation because a similar psychometric fit is obtained even when some of the range is missing. Therefore, we conclude that the testing-range effects observed for the discrimination experiment reflect a true testing-range bias. For ranking, any nonsignificant trend toward a small range effect could likely be explained by truncation (Figures 3B, E and 4B, E). For matching, the testing-range effects were explained by truncation (Figure 4C). The tendency for the PSE to shift toward the center of the testing range (black circles in Figure 3F; shallow reference-electrode slope in Figure 4F) was only partially explained by truncation and therefore likely includes a true testing-range bias. In conclusion, of the three methods tested, only ranking was immune to testing biases that could not be explained by truncation.

Besides biases, both ranking and matching were more efficient tasks to achieve PSEs. On average, it took about 1.5 to 2 hr to perform the ranking and matching tasks compared with the 3.5 to 4 hr to perform the discrimination task.

### Comparison of Pitch Tasks

**Discrimination.** The theoretical strength of pitch discrimination is that it provides a mathematically trackable approach to estimate both the PSE (50% crossover of the psychometric function) and pitch discriminability (slope of the psychometric function; see Figure 1A–C). This, however, comes at the expense of a relatively longer testing time needed to reduce the variance for each condition represented on the psychometric function. In practice, any theoretical advantages were superseded by massive disadvantages. Besides the long testing

time, discrimination was susceptible to large persistent testing-range effects (Figures 3A, D, and 4A) and generated PSEs that deviated considerably from the expected value (Figure 4D). While we cannot rule out that listeners in this study had a real interaural place mismatch, there was a consistent bias in the results, with an 8-electrode change in reference electrode yielding only an average 4.8-electrode change in the PSE. These effects were a result of the perceptual biases introduced by specific psychophysical procedures (Poulton, 1979), rather than a truncation effect, which was negligible for this experiment. Evidence to strongly suggest a true bias includes the data from Goupell et al. (2019, Figure 3) that shows that listener BCI6 had the same pattern of biases regardless of whether the right or left ear served as a reference. It was not the case that one ear showed compression of PSEs and the other ear showed expansion, which would be an indication of a true interaural place mismatch. In summary, our interpretation of these data leaves us with the opinion that the discrimination method should be avoided because, of the three methods tested here, it is the most susceptible to methodological biases and the slowest.

**Ranking.** The ranking and discrimination tasks share many commonalities and are effectively the same task from the point of view of the listener who, in both cases, is presented with pairs of stimuli and is asked to express a relative judgment of pitch. The key difference is that ranking does not require the complete set of electrode pairs to define a psychometric function and therefore can be completed with fewer trials. This approach thus moves toward the suggestion by Poulton (1979) to minimize testing-range biases by minimizing the number of trials per listener. The midpoint comparison ranking procedure used in Experiment 2 was highly efficient, requiring 38% fewer trials than discrimination (on average, per listener) and took about half the time to complete.

Ranking was robust to testing-range biases (Figure 4B), produced reference-electrode slopes near 1 (Figure 4E), and was significantly better than discrimination and matching for these metrics. The ranking procedure was also mostly immune to changes in starting values (Figure 5A). Furthermore, any small apparent biases in the results are likely explained by truncation (Figures 3E, 4B, and 4E) rather than a true testing-range bias. Support for the relative success of ranking can be found in Adel et al. (2019). They compared PSE estimates for pitch in SSD-CI listeners using a binary search procedure, which has similarities to the midpoint-comparison ranking procedure, but is even more efficient. They verified no effect of starting point, similar to our results; however, they did not explicitly



investigate range and reference-electrode slope effects as was done in this study.

Why is ranking immune to these procedural biases? One possibility is that the ranking approach involves a combination of within- and across-ear pitch comparisons. This is unlike the discrimination and matching approaches, which involve interaural comparisons on every trial. In these other tasks, the listener may have a tendency to attend to just the comparison ear and judge pitch relative to the available range (particularly for discrimination, as suggested in Goupell et al., 2019). On any given trial for ranking, the listener does not know a priori which kind of comparison will be presented and therefore may better attend to and compare to the two pitches.

One potential weakness of the ranking procedure is that the various stimuli are not presented equally often, with electrodes at the extremes of the array being chosen less frequently. This can lead to a nonlinear *stimulus frequency bias* (Poulton, 1979), which can increase the chances of the listener responding based on expected equal probability of stimuli. However, we did not observe any significant biases in the results (Figure 4B and E). Another potential issue with this procedure is that any listener error in the response process, especially if it occurs at the beginning of a ranking trial, will affect the rank of all the following elements, as each electrode pair is compared only once for every run. In other words, an error in pitch judgment early in the track would lead to a relatively more inaccurate PSE. To reduce these sources of error, we repeated each ranking procedure 10 times and added a repeat option in our design for each trial (alternatively or in addition, an undo option could be used). This difference in procedure may explain some of the difference in the discrimination and ranking results. It is possible that the repeated presentation for ranking allowed for a pitch judgment to be less affected by bias. However, listeners only repeated sounds on 14% of trials over the entire experiment. The number of repeated presentations within a run was, on average, 4.0 ranging from 0 to 31. The runs with the largest number of use of the repeat option occurred primarily for one listener (BCI3) in their initial five blocks. It may have been they were learning the task. Given the small percentage of trials where the repeat option was used, this would argue for a small or negligible effect of the repeat option contributing to the difference in discrimination and ranking results.

Taken together, the ranking task appears to be best suited to estimates of relative interaural pitch, particularly for CI users, because it is relatively absent of procedural biases compared with discrimination and matching. Furthermore, any observed biases can be attributed to truncation. This means that any small effects can be statistically modeled and adjusted for if

there is sufficient data to characterize the shape of the response distribution. Pitch ranking via midpoint comparison also offers an efficient way to obtain interaural PSE estimates for all electrodes on both sides in a single run (or a few runs, for increased robustness), which is of clinical interest.

**Matching.** The matching task had some advantages. First, it is a direct measurement of the PSE and therefore has the best face validity of the three pitch-comparison methods. Second, it is relatively fast, taking approximately as long as the ranking task and half as long as the discrimination task. Third, it was robust to some procedural biases. Although there were a few cases where the starting point appeared to influence the PSE, and on average the starting point slope of 0.11 was significantly different from zero, the starting point slope was less than the proposed criterion value of 0.5 (Carlyon, Macherey, et al., 2010) for 91% of cases tested (Figure 5B). This method showed testing-range effects that were significantly smaller than for discrimination and not significantly different than for ranking (compare black bars across Figure 4A–C). Furthermore, these range effects were accounted for by truncation (compare black and white bars in Figure 4C).

The matching task also had some disadvantages. In particular, the reference-electrode slope was significantly  $<1$  (Figure 4F), demonstrating a bias away from edges of the testing range. This was significantly worse than the ranking task but significantly better than the discrimination task and was only partially explained by truncation (Figure 4F, compare black and white bars), which suggests that matching was susceptible to procedural biases reviewed in Poulton (1979). In addition, there is an unknown psychometric function unlike the discrimination task. While the matching task does not directly measure discriminability, it could be derived from variance of the response distribution with enough trials. However, such information works against the relatively short duration of the task because it would require many more trials per condition than the eight that were collected for this study. Furthermore, in the matching task, there is the assumption that there is a monotonic place-pitch function. In reality, nonmonotonic functions can occur and possibly produce the wrong interaural electrode matching, increased confusion, or variability of responses. We explicitly addressed this issue in the current study by ordering our electrodes by using ranks from ranking experiment; while such an approach comes at the expense of testing duration, the very small number of reversals made this step effectively unnecessary.

In summary, matching appears to be a relatively sound and rigorous approach, and arguably the second-best pitch task, depending on the experimenter's



priorities. It is relatively fast and has the best face validity to determine a PSE because this is the only one of the three experiments where the listener is directly instructed to match the pitches in the two ears. However, it does suffer from greater biases than ranking, especially a bias in the PSE away from the edges of the testing range (Figure 4F).

### Comparison to Previous Literature

The most important observation in this study is that the PSE was highly susceptible to differences in testing methodology. These results add to a growing literature that has demonstrated the susceptibility of pitch-comparison tasks to the specific methodological details of the experiment. While Carlyon, Macherey, et al. (2010) pointed out the existence of testing-range biases for SSD-CI listeners in pitch-comparison tasks that employ a method of constant stimuli, Goupell et al. (2019) extended these findings to BI-CI listeners who have more symmetric hearing. In addition, Adel et al. (2019) examined the influence of stimulus properties on the PSE using a ranking task in SSD-CI listeners. They found that the characteristics of the acoustic stimulus—pure-tones, narrow-band noise, or a band-limited harmonic complex tone—had a substantial effect on the mean and variance of the PSE; a robust measurement of PSEs should provide the same average answer relatively independent of the exact spectral-temporal nature of the narrowband acoustic stimulus.

A second key observation of this study is that adhering to the guidelines proposed by Carlyon, Macherey, et al. (2010) is necessary but not sufficient for identifying pitch-comparison biases. Carlyon, Macherey, et al. (2010) proposed two bias checks: (a) that the PSE should be relatively independent of the testing range and (b) that the PSE should be relatively independent of the starting point of an adaptive track. The proposed maximum allowable effect size on the PSE was 50% of the change in testing-range midpoint or adaptive-track starting point. Importantly, it was proposed that only one of these checks is required for a given measurement: the testing-range check proposed for methods involving constant stimuli (i.e., discrimination) and the starting-point check proposed for adaptive methods (i.e., ranking or matching).

In this study, we applied both checks to the two adaptive methods. We also added a third check, which was to examine whether the PSE shifted as expected with a given change in reference electrode. Both experiments consistently passed the starting-point check (there were few cases of a starting-point effect  $>0.5$ , Figure 5) as well as the range-effect check (slopes consistently  $<0.5$ , Figure 4B and C). However, the reference-electrode slope was consistently and significantly  $<1$  for the

matching experiment (average = 0.6), which means that this method yielded a PSE that was biased toward the center of the array. Therefore, we argue that these data demonstrate that conducting only a starting-point analysis check for an adaptive task is not sufficient; checks of the range effect and reference-electrode slope are also required for the measurement to be deemed valid.

One major problem with the reference-electrode slope check is the difficulty in applying it on an individual-listener basis. This check was applied to group-average BI-CI data under the assumption that, on average, the arrays in each ear were either interaurally matched, or at least that any mismatch was constant across the array, and could therefore be used to validate or invalidate a particular pitch-comparison task. However, this check cannot be applied as a check for an individual listener. Here, interaural mismatch is the very quantity being measured and could be affected by individual differences in array type (i.e., different interelectrode spacing), physical electrode location (Stakhovskaya et al., 2007), or pitch plasticity (Reiss et al., 2007, 2014). To avoid the need for this assumption, the pitch-comparison measurement could be repeated with each ear serving as the reference, and the resulting reference-electrode slopes could be compared. The two slopes would need to be mutually consistent for the measurement to be deemed free of bias. Such an approach was done for one BI-CI listener in Goupell et al. (2019) using discrimination. The PSEs were compressed in both ears, suggesting a methodological bias rather than an interaural frequency mismatch or plasticity.

The fact that changing only the properties of the stimulus or the methodological details of the pitch-comparison measurement can have a large effect on the experimental outcome strongly questions the utility of using pitch-comparison methods to evaluate interaural place-of-stimulation mismatch for CI users. Numerous studies have used interaural place-pitch comparisons to estimate the relative place of electrical stimulation for BI-CI listeners. In many cases, those matches were assumed to be the point of best binaural alignment, and then binaural processing was evaluated (e.g., Churchill et al., 2014; Goupell, 2015; Kan et al., 2013, 2015, 2019; Laback & Majdak, 2008; Litovsky et al., 2010; Majdak et al., 2006; van Hoesel, 2007). However, the current experiments combined with the results of Goupell et al. (2019) suggest that their equivalence is suspect, especially in those cases that used discrimination, or to some extent, a matching approach. Furthermore, the growing numbers of CI users with usable acoustic hearing in the contralateral ear (i.e., bimodal and SSD-CI listeners) have led to similar questions regarding optimizing interaural frequency alignment (Schatzer et al., 2014; Wess et al., 2017) and whether the auditory system demonstrates plasticity to

such alignment (Svirsky et al., 2015). Many of these studies have used interaural place-pitch comparisons to address these questions. Our results in BI-CI listeners suggest that conclusions regarding interaural place mismatch for bimodal and SSD-CI listeners may be heavily influenced by biases unless methodological steps are taken to reduce them.

Given these issues, it is an open question whether it is sensible to conduct pitch-comparison experiments in CI listeners for the purpose of identifying the relative cochlear place of stimulation in the two ears. Our response is that it depends on the goal. If the goal is to understand the percept of pitch and how it is affected by implantation, hearing loss, or plasticity (Reiss et al., 2007, 2014), then pitch comparisons cannot be avoided. In these cases, we recommend a ranking approach, given the low susceptibility to range biases, and the fact that any biases could be explained by a simple model involving internal noise in the pitch task. However, in many cases, the goal is not pitch-specific—for example, to improve binaural outcomes through precise binaural matching (Bernstein et al., 2018; Hu & Dietz, 2015). In these cases, methods that more directly target these spatial-hearing mechanisms—such as binaural tuning curves (Bernstein et al., 2018; Dirks et al., 2020; Francart et al., 2018; Hu & Dietz, 2015; Kan et al., 2013, 2015) or CT scans (Canfarotta et al., 2020)—are recommended.

### *Limitations and Future Directions*

A primary limitation of this type of work is that there is no ground-truth measurement for interaural pitch comparison in CI users (Adel et al., 2019). In other words, it is unclear what the true PSE should be in this population. Such a problem does not occur in NH listeners, where one can assume the ears to be copies of each other and that auditory processing and pitch perception will be nearly perfectly equivalent between the two ears. Besides not having a ground truth to compare against, such measurements in CI listeners are made more difficult by issues of potential plasticity (e.g., Reiss et al., 2014), which is why objective place measures such as X-rays or CT scans cannot be fully trusted as a ground truth. Binaural tuning measures also do not necessarily align with pitch measures because of potential differential plasticity across the two systems (as suggested in Goupell et al., 2019). Given this, one should carefully attend to any potential biases in their pitch measurements, as we have tried to carefully document in this study.

Another limitation is that the BI-CI listeners in this study had extensive practice judging interaural place pitch by the end of this protocol. This potential practice effect may have been compounded by the fact that the

listeners performed the types of pitch tasks in the same order. Therefore, order effects may have contributed to our results. Naïve listeners from the general population of BI-CI users could perform differently from those tested here.

Our modeling of the truncation effect consisted of a simple one-parameter model that assumed no interaural mismatch. Such an assumption is likely valid given any small mismatches in interaural insertion depth would average to approximately zero over multiple listeners and that informal observation of CT scans from the listeners in this study did not show large mismatches in most of them. However, future work more closely examining truncation could include a second parameter of mismatch, informed by CT or binaural tuning curve information.

Finally, while we demonstrated that ranking may be the superior interaural pitch comparison method because of its robustness to biases in BI-CI listeners, there is the possibility that this may not be the case for listeners with acoustic and electric hearing, such as bimodal and SSD-CI listeners. Future studies examining biases in these populations would be necessary to truly extrapolate the results here. Goupell et al. (2019) found that the magnitude of the range effects for SSD-CI listeners in a discrimination experiment was similar to that observed for BI-CI listeners; however, that study only included three SSD-CI listeners, so that comparison should be interpreted cautiously.

### **Summary and Conclusions**

Interaural pitch comparisons are commonly used in spatial-hearing-related experiments in CI users; however, the consistency of these measurements across measurement paradigms has yet to be rigorously evaluated. This study evaluated three pitch-comparison methods in BI-CI listeners for procedural biases such as testing-range effects that can influence the measurement of the PSE for a given reference electrode. Discrimination was the slowest method and most susceptible to biases; we advise against using this method. Matching was faster and less prone to biases than discrimination, but was still susceptible to a reference-electrode slope bias that shifted the PSE toward the center of the testing range. Ranking was also faster than discrimination and, critically, was mostly unaffected to the procedural biases investigated here. The results also suggested that a single bias check was not sufficient to rule out procedural bias in a pitch-comparison task. While both the matching and ranking experiments passed the starting-point check, the matching procedure was nevertheless found to exhibit some bias in terms of the slope of the relationship between the PSE and the reference electrode.

Overall, these results led to two main conclusions. First, to rule out the existence of bias, pitch-comparison measurements must be subjected to not just one, but to three checks: testing-range effects, starting-point effects (for matching and ranking tasks), and a reference-electrode slope bias. Second, to avoid procedural biases, we recommend that studies of relative interaural pitch for CI listeners employ a ranking approach.

### Acknowledgments

We thank Cochlear Ltd. for providing the testing equipment and technical support. We thank Danielle King who helped collect data for this study. We thank the Department of Hearing and Speech Sciences at University of Maryland-College Park (Dr. Nicole Nguyen), the University of Maryland Medical School (Dr. David Eisenman and Dr. Ronna Hertzano), the Cochlear Implant Center at Greater Baltimore Medical Center (Dr. Regina Presley), and Walter Reed National Military Medical Center (Dr. Gerald Schuchman) for their assistance with recruiting listeners. We thank Ginny Alexander for managerial help.

This work was partially presented at the Conference on Implantable Auditory Prostheses in Lake Tahoe, CA, USA, July 2017 and July 2019.

The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the authors, Department of Defense, or any component agency. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.


### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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 R01 DC015798 (M. J. G. and J. G. W. B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### ORCID iD

Matthew J. Goupell  <https://orcid.org/0000-0003-2662-3057>

### References

- Adel, Y., Nagel, S., Weissgerber, T., Baumann, U., & Macherey, O. (2019). Pitch matching in cochlear implant users with single-sided deafness: Effects of electrode position and acoustic stimulus type. *Frontiers in Neuroscience, 13*, 1119. <https://doi.org/10.3389/fnins.2019.01119>
- Aronoff, J. M., Padilla, M., Stelmach, J., & Landsberger, D. M. (2016). Clinically paired electrodes are often not perceived as pitch matched. *Trends in Hearing, 20*, 1–9. <https://doi.org/10.1177/2331216516668302>
- Bernstein, J. G. W., Stakhovskaya, O. A., Schuchman, G. I., Jensen, K. K., & Goupell, M. J. (2018). Interaural-time-difference discrimination as a measure of place of stimulation for cochlear-implant users with single-sided deafness. *Trends in Hearing, 22*, 1–19. <https://doi.org/10.1177/2331216518765514>
- Bierer, J. A., & Faulkner, K. F. (2010). Identifying cochlear implant channels with poor electrode-neuron interface: Partial tripolar, single-channel thresholds and psychophysical tuning curves. *Ear and Hearing, 31*(2), 247–258. <https://doi.org/10.1097/AUD.0b013e3181c7daf4>
- Blanks, D. A., Roberts, J. M., Buss, E., Hall, J. W., & Fitzpatrick, D. C. (2007). Neural and behavioral sensitivity to interaural time differences using amplitude modulated tones with mismatched carrier frequencies. *Journal of the Association for Research in Otolaryngology, 8*(3), 393–408. <https://doi.org/10.1007/s10162-007-0088-5>
- Canfarotta, M. W., Dillon, M. T., Buss, E., Pillsbury, H. C., Brown, K. D., & O'Connell, B. P. (2020). Frequency-to-place mismatch: Characterizing variability and the influence on speech perception outcomes in cochlear implant recipients. *Ear and Hearing, 41*(5), 1349–1361. <https://doi.org/10.1097/AUD.0000000000000864>
- Carlyon, R. P., Deeks, J. M., & Macherey, O. (2013). Polarity effects on place pitch and loudness for three cochlear-implant designs and at different cochlear sites. *Journal of the Acoustical Society of America, 134*(1), 503–509. <https://doi.org/10.1121/1.4807900>
- Carlyon, R. P., Lynch, C., & Deeks, J. M. (2010). Effect of stimulus level and place of stimulation on temporal pitch perception by cochlear implant users. *Journal of the Acoustical Society of America, 127*(5), 2997–3008. <https://doi.org/10.1121/1.3372711>
- Carlyon, R. P., Macherey, O., Frijns, J. H., Axon, P. R., Kalkman, R. K., Boyle, P., & Dauman, R. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *Journal of the Association for Research in Otolaryngology, 11*(4), 625–640. <https://doi.org/10.1007/s10162-010-0222-7>
- Churchill, T., Kan, A., Goupell, M. J., & Litovsky, R. Y. (2014). Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners. *Journal of the Acoustical Society of America, 136*(3), 1246–1256. <https://doi.org/10.1121/1.4892764>
- Collins, L. M., Zwolan, T. A., & Wakefield, G. H. (1997). Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects. *Journal of the Acoustical Society of America, 101*(1), 440–455. <https://doi.org/10.1121/1.417989>



- Dirks, C. E., Nelson, P. B., Winn, M. B., & Oxenham, A. J. (2020). Sensitivity to binaural temporal-envelope beats with single-sided deafness and a cochlear implant as a measure of tonotopic match (L). *Journal of the Acoustical Society of America*, *147*(5), 3626–3630. <https://doi.org/10.1121/10.0001305>
- Ehlers, E., Goupell, M. J., Zheng, Y., Godar, S. P., & Litovsky, R. Y. (2017). Binaural sensitivity in children who use bilateral cochlear implants. *Journal of the Acoustical Society of America*, *141*(6), 4264–4277. <https://doi.org/10.1121/1.4983824>
- Firszt, J. B., Reeder, R. M., Holden, L. K., & Dwyer, N. Y. (2018). Results in adult cochlear implant recipients with varied asymmetric hearing: A prospective longitudinal study of speech recognition, localization, and participant report. *Ear and Hearing*, *39*(5), 845–862. <https://doi.org/10.1097/AUD.0000000000000548>
- Fischer, B. J., & Peña, J. L. (2009). Bilateral matching of frequency tuning in neural cross-correlators of the owl. *Biological Cybernetics*, *100*(6), 521–531. <https://doi.org/10.1007/s00422-009-0312-y>
- Francart, T., Wiebe, K., & Wesarg, T. (2018). Interaural time difference perception with a cochlear implant and a normal ear. *Journal of the Association for Research in Otolaryngology*, *19*(6), 703–715. <https://doi.org/10.1007/s10162-018-00697-w>
- Goupell, M. J. (2015). Interaural correlation-change discrimination in bilateral cochlear-implant users: Effects of interaural frequency mismatch, centering, and age of onset of deafness. *Journal of the Acoustical Society of America*, *137*(3), 1282–1297. <https://doi.org/10.1121/1.4908221>
- Goupell, M. J., Cosentino, S., Stakhovskaya, O. A., & Bernstein, J. G. W. (2019). Interaural pitch-discrimination range effects for bilateral and single-sided deafness cochlear-implant users. *Journal of the Association for Research in Otolaryngology*, *20*(2), 187–203. <https://doi.org/10.1007/s10162-018-00707-x>
- Harris, J. D. (1948). Discrimination of pitch; suggestions toward method and procedure. *American Journal of Psychology*, *61*(3), 309–322. <https://doi.org/10.2307/1417151>
- Hu, H., & Dietz, M. (2015). Comparison of interaural electrode pairing methods for bilateral cochlear implants. *Trends in Hearing*, *19*, 1–22. <https://doi.org/10.1177/2331216515617143>
- Kan, A., Goupell, M. J., & Litovsky, R. Y. (2019). Effect of channel separation and interaural mismatch on fusion and lateralization in normal-hearing and cochlear-implant listeners. *Journal of the Acoustical Society of America*, *146*(2), 1448–1463. <https://doi.org/10.1121/1.5123464>
- Kan, A., & Litovsky, R. Y. (2015). Binaural hearing with electrical stimulation. *Hearing Research*, *322*, 127–137. <https://doi.org/10.1016/j.heares.2014.08.005>
- Kan, A., Litovsky, R. Y., & Goupell, M. J. (2015). Effects of interaural pitch-matching and auditory image centering on binaural sensitivity in cochlear-implant users. *Ear and Hearing*, *36*(3), e62–e68. <https://doi.org/10.1097/AUD.0000000000000135>
- Kan, A., Stoelb, C., Litovsky, R. Y., & Goupell, M. J. (2013). Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. *Journal of the Acoustical Society of America*, *134*(4), 2923–2936. <https://doi.org/10.1121/1.4820889>
- Kenway, B., Tam, Y. C., Vanat, Z., Harris, F., Gray, R., Birchall, J., & Axon, P. (2015). Pitch discrimination: An independent factor in cochlear implant performance outcomes. *Otology & Neurotology*, *36*(9), 1472–1479. <https://doi.org/10.1097/MAO.0000000000000845>
- Laback, B., & Majdak, P. (2008). Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(2), 814–817. <https://doi.org/10.1073/pnas.0709199105>
- Landsberger, D. M., Vermeire, K., Claes, A., Van Rompaey, V., & Van de Heyning, P. (2016). Qualities of single electrode stimulation as a function of rate and place of stimulation with a cochlear implant. *Ear and Hearing*, *37*(3), e149–e159. <https://doi.org/10.1097/AUD.0000000000000250>
- Litovsky, R. Y., Goupell, M. J., Godar, S., Grieco-Calub, T., Jones, G. L., Garadat, S. N., & 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>
- Litovsky, R. Y., Jones, G. L., Agrawal, S., & van Hoesel, R. (2010). Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *Journal of the Acoustical Society of America*, *127*(1), 400–414. <https://doi.org/10.1121/1.3257546>
- Long, C. J., Eddington, D. K., Colburn, H. S., & Rabinowitz, W. M. (2003). Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. *Journal of the Acoustical Society of America*, *114*(3), 1565–1574. <https://doi.org/10.1121/1.1603765>
- Long, C. J., & Nimmo-Smith, I., Baguley, D. M., O'Driscoll, M., Ramsden, R., Otto, S. R., Carlyon, R. P. (2005). Optimizing the clinical fit of auditory brain stem implants. *Ear and Hearing*, *26*(3), 251–262. <https://doi.org/10.1097/00003446-200506000-00002>
- Macherey, O., & Carlyon, R. P. (2012). Place-pitch manipulations with cochlear implants. *Journal of the Acoustical Society of America*, *131*(3), 2225–2236. <https://doi.org/10.1121/1.3677260>
- Majdak, P., Laback, B., & Baumgartner, W. D. (2006). Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing. *Journal of the Acoustical Society of America*, *120*(4), 2190–2201. <https://doi.org/10.1121/1.2258390>
- McKay, C. M., O'Brien, A., & James, C. J. (1999). Effect of current level on electrode discrimination in electrical stimulation. *Hearing Research*, *136*(1–2), 159–164. [https://doi.org/10.1016/S0378-5955\(99\)00121-5](https://doi.org/10.1016/S0378-5955(99)00121-5)
- Noel, V. A., & Eddington, D. K. (2013). Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences. *Journal of the Acoustical Society of*



- America*, 133(4), 2314–2328. <https://doi.org/10.1121/1.4794372>
- Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, 86(4), 777–803. <https://doi.org/10.1037/0033-2909.86.4.777>
- Reda, F. A., McRackan, T. R., Labadie, R. F., Dawant, B. M., & Noble, J. H. (2014). Automatic segmentation of intra-cochlear anatomy in post-implantation CT of unilateral cochlear implant recipients. *Medical Image Analysis*, 18(3), 605–615. <https://doi.org/10.1016/j.media.2014.02.001>
- Reiss, L. A., Lowder, M. W., Karsten, S. A., Turner, C. W., & Gantz, B. J. (2011). Effects of extreme tonotopic mismatches between bilateral cochlear implants on electric pitch perception: A case study. *Ear and Hearing*, 32(4), 536–540. <https://doi.org/10.1097/AUD.0b013e31820c81b0>
- Reiss, L. A., Turner, C. W., Erenberg, S. R., & Gantz, B. J. (2007). Changes in pitch with a cochlear implant over time. *Journal of the Association for Research in Otolaryngology*, 8(2), 241–257. <https://doi.org/10.1007/s10162-007-0077-8>
- Reiss, L. A., Turner, C. W., Karsten, S. A., & Gantz, B. J. (2014). Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation. *Neuroscience*, 256, 43–52. <https://doi.org/10.1016/j.neurosci.2013.10.024>
- Schatzer, R., Vermeire, K., Visser, D., Krenmayr, A., Kals, M., Voormolen, M., & Zierhofer, C. (2014). Electric-acoustic pitch comparisons in single-sided-deaf cochlear implant users: Frequency-place functions and rate pitch. *Hearing Research*, 309, 26–35. <https://doi.org/10.1016/j.heares.2013.11.003>
- Stakhovskaya, O., Sridhar, D., Bonham, B. H., & Leake, P. A. (2007). Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *Journal of the Association for Research in Otolaryngology*, 8(2), 220–233. <https://doi.org/10.1007/s10162-007-0076-9>
- Steinhaus, H. (1950). *Mathematical Snapshots*. Oxford University Press.
- Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1(5), 439–446. <https://doi.org/10.3758/bf03207424>
- Svirsky, M. A., Fitzgerald, M. B., Sagi, E., & Glassman, E. K. (2015). Bilateral cochlear implants with large asymmetries in electrode insertion depth: Implications for the study of auditory plasticity. *Acta Oto-Laryngologica*, 135(4), 354–363. <https://doi.org/10.3109/00016489.2014.1002052>
- Tan, C. T., Martin, B., & Svirsky, M. A. (2017). Pitch matching between electrical stimulation of a cochlear implant and acoustic stimuli presented to a contralateral ear with residual hearing. *Journal of the American Academy of Audiology*, 28(3), 187–199. <https://doi.org/10.3766/jaaa.15063>
- van Hoesel, R. J. M. (2007). Sensitivity to binaural timing in bilateral cochlear implant users. *Journal of the Acoustical Society of America*, 121(4), 2192–2206. <https://doi.org/10.1121/1.2537300>
- van Hoesel, R. J. M., Jones, G. L., & Litovsky, R. Y. (2009). Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse rate, modulation rate, and place of stimulation. *Journal of the Association for Research in Otolaryngology*, 10(4), 557–567. <https://doi.org/10.1007/s10162-009-0175-x>
- Vickers, D. A., Degun, A., Canas, A., Stainsby, T., & Vanpoucke, F. (2016). Deactivating cochlear implant electrodes based on pitch information for users of the ACE strategy. In P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, & C. Lanting (Eds.), *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 115–123). Springer International Publishing. doi: 10.1007/978-3-319-25474-6\_13
- Wess, J. M., Brungart, D. S., & Bernstein, J. G. W. (2017). The effect of interaural mismatches on contralateral unmasking with single-sided vocoders. *Ear and Hearing*, 38(3), 374–386. <https://doi.org/10.1097/AUD.0000000000000374>
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313. <https://doi.org/10.3758/bf03194544>
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8), 1314–1329. <https://doi.org/10.3758/bf03194545>
- Zhao, Y., Dawant, B. M., Labadie, R. F., & Noble, J. H. (2014). Automatic localization of cochlear implant electrodes in CT. *Medical Image Computing and Computer-Assisted Intervention*, 17(Pt 1), 331–338. [https://doi.org/10.1007/978-3-319-10404-1\\_42](https://doi.org/10.1007/978-3-319-10404-1_42)