

Untrained listeners experience difficulty detecting interaural correlation changes in narrowband noises

Matthew J. Goupell^{a)} and Mary E. Barrett

Department of Hearing and Speech Sciences, University of Maryland, College Park,
Maryland 20742, USA
goupell@umd.edu, mbarrett@umd.edu

Abstract: Interaural correlation change detection was measured in untrained normal-hearing listeners. Narrowband (10-Hz) noises were varied by center frequency (CF; 500 or 4000 Hz) and diotic level roving (absent or present). For the 500-Hz CF, 96% of listeners could achieve threshold (79.4% correct at the easiest testing level) if roving was absent, but only 36% of listeners could if level roving was present. No one could achieve threshold at the 4000-Hz CF, unlike trained listeners in the literature. The results raise questions about how individual differences affect learning and generalization of monaural and binaural cues related to interaural correlation detection.

© 2015 Acoustical Society of America

[QJF]

Date Received: March 18, 2015 Date Accepted: June 15, 2015

1. Introduction

Human psychoacoustical experiments often utilize trained listeners, meaning the listeners have been given explicit instructions on how to perform a particular auditory task and they have been exposed in some way to the task before data collection (i.e., they have had some explicit and beneficial practice). In many cases, listeners may practice the task for hours. Such practice often includes correct answer feedback and continues until there is an apparent saturation in listener performance. A major reason to include such training is to avoid a substantial within-subject performance change during data collection that may obscure across-condition effects. The purpose of this work was to investigate the performance of untrained listeners in a binaural task that shows highly variable performance, namely, detection of interaural correlation changes (ICC) in narrowband noises. Of particular interest was to obtain a better understanding of the detection cues listeners might be using when performing ICC detection, and how performance differs between untrained and trained listeners.

In some psychoacoustical tasks, it is unnecessary for listeners to receive explicit training because they do not improve over time. For example, Trahiotis *et al.* (1990) showed that listeners had stable and unchanging thresholds over 25 sessions when detecting an interaurally in-phase or out-of-phase 500-Hz tone embedded in a 2900-Hz bandwidth diotic noise (called NoSo and NoS π detection, respectively). In other cases, training is necessary as significant improvements in performance can be observed. For example, Wright and Fitzgerald (2001) showed that untrained listeners had 500-Hz tone interaural time difference (ITD) discrimination thresholds of about 40–60 μ s, and these thresholds improved to 20 μ s over two weeks of training (specifically, nine days of training with one hour of training/day; however, most of the improvement occurred over the first 30 min of training). Likewise, these listeners had 4000-Hz tone interaural level difference (ILD) discrimination thresholds of about 4 dB, and thresholds improved to 2 dB after training (the time course of improvement was relatively longer than for the ITD discrimination task). While these two example studies refer to different tasks, they have some commonalities: they were presented over headphones to precisely control the properties of the stimuli, they used non-ecologically valid stimuli that are rarely experienced outside of the laboratory, and they tested binaural processing abilities (meaning the detection cues were accessed through an interaural comparison of the signals).

Of interest in this study is sensitivity to ICC, which is related to the binaural masking level difference (e.g., Durlach *et al.*, 1986; Goupell and Litovsky, 2014) and

^{a)} Author to whom correspondence should be addressed.

how speech is experienced and understood in background noise and reverberant rooms (e.g., [Lavandier and Culling, 2010](#)). Listeners can be highly sensitive to ICC changes in noises ([Gabriel and Colburn, 1981](#)), where $\rho = 1$ for a perfectly correlated noise and $\rho < 1$ for a decorrelated noise. As ρ decreases, larger fluctuations in the ITDs and ILDs are introduced ([Goupell, 2010](#)), and the salient perceptual change is a widening or blurring of intracranial image (for larger bandwidth signals; [Whitmer et al., 2012](#)) or a moving intracranial location (for small bandwidths of about 10 Hz or less; [Gabriel and Colburn, 1981](#); [Goupell and Litovsky, 2014](#)).

ICC sensitivity can be highly variable across experienced listeners when presented narrowband noises ([Koehnke et al., 1986](#); [Goupell, 2012](#)), but it is unclear why this variability exists. One reason could be that some listeners are inherently more sensitive to binaural cues ([Koehnke et al., 1986](#)). Another reason could be that listeners use different cues to perform the task. For example, one listener could rely more on fluctuating ITDs and another on fluctuating ILDs ([Goupell and Hartmann, 2007](#); [Goupell, 2010](#); [Mao and Carney, 2014](#)). Others may ignore the spatial percepts and attempt to use an increase in loudness for the dichotic target compared to the diotic non-target stimuli ([Edmonds and Culling, 2009](#)), which would imply that the performance of these listeners would be particularly susceptible to diotic level roving where the loudness cue would be made unreliable. Or perhaps some listeners confuse monaural envelope fluctuations (i.e., roughness) with the binaural fluctuations ([Goupell and Hartmann, 2006](#)). Therefore, one goal of this work is to examine what cues untrained listeners rely on to detect ICC by varying stimulus center frequency (CF) and the absence or presence of diotic level roving. If we can determine what cues listeners are using, such an approach could help explain the relatively large inter-individual variability observed in ICC detection (e.g., [Koehnke et al., 1986](#); [Goupell, 2012](#)), and could improve binaural models' ability to explain ICC and binaural unmasking performance ([Goupell and Hartmann, 2007](#); [van der Heijden and Joris, 2009](#); [Goupell, 2010](#); [Mao and Carney, 2014](#)).

It is also unclear what the time course of training-induced improvement is for ICC detection as listeners gain experience with this task. Therefore, another goal of this work was to characterize the initial untrained performance and improvement of listeners in ICC detection if they were provided correct answer feedback. There is evidence that the improvement and saturation in ITD sensitivity can occur within 30 min of training using 500-Hz tones ([Wright and Fitzgerald, 2001](#)).

We hypothesized that untrained listeners would be worse at ICC detection than what has been previously reported in experienced listeners because the untrained listeners might ignore the binaural fluctuations and attempt to use other potentially confusing cues. We also hypothesized that there would be rapid improvement in ICC thresholds at 500 Hz, but not 4000 Hz ([Wright and Fitzgerald, 2001](#)). This is because ICC sensitivity at 500 Hz is thought to be dominated by fluctuating ITDs ([van der Heijden and Joris, 2009](#)), whereas ICC sensitivity at 4000 Hz is thought to be dominated by fluctuating ILDs ([Goupell, 2012](#)).

2. Methods

2.1 Listeners and equipment

Fifty-nine listeners participated in this study, all of whom were considered untrained listeners because they had no experience in detecting ICC in psychoacoustical headphone experiments. The listeners (age range = 18–42 years; mean age = 20.0 years; 49 females) had normal audiometric thresholds (≤ 20 dB hearing level at octave frequencies between 250 and 8000 Hz) and no appreciable interaural asymmetries in hearing thresholds (< 10 dB at any tested frequency). Most of them were college undergraduates and were compensated with class credit or a small payment.

The stimuli were generated on a personal computer in MATLAB (Mathworks, Natick, MA), delivered by a sound card (Edirol UA-25EX, Roland Corporation, Japan) to a power amplifier (D-75A, Crown Audio, Elkart, IN) and then to open-backed circumaural headphones (HD650, Sennheiser Corporation, Germany). The listeners were seated in a double-walled sound attenuating booth (IAC, Bronx, NY) for the testing.

2.2 Stimuli

The stimuli were 10-Hz bandwidth noises with a 500- or 4000-Hz CF. The rationale for using 10-Hz bandwidth noises was listeners may demonstrate greater idiosyncratic weighting of the detection cues, namely, the weighting of fluctuating ITDs and ILDs ([Goupell and Hartmann, 2007](#); [van der Heijden and Joris, 2009](#); [Goupell, 2010](#); [Mao and Carney, 2014](#)). The stimuli had a duration of 300 ms and were shaped by a Tukey

Table 1. Proportion of untrained listeners who could achieve threshold performance (i.e., $\geq 79.4\%$ correct) for the starting point of the adaptive track (i.e., reference $\rho = 1$ and target $\rho = 0$).

CF (Hz)	Roving	n	Achieved Threshold (%)
500	Absent	25	96
500	Present	14	36
4000	Absent	10	0
4000	Present	10	0

window with a 10-ms rise-fall time. They were presented at 65 dB-A, unless there was diotic level roving where the level was randomly varied over a 10-dB range (± 5 dB of rove chosen from a rectangular distribution). The stimulus interaural correlation was precisely controlled using an orthogonalization procedure (Culling *et al.*, 2001). The number of listeners tested in each condition is reported in Table 1.

2.3 Procedure

Listeners performed a three-interval, two-alternative forced-choice task in a three-down, one-up adaptive procedure to obtain a threshold that targeted 79.4% correct (Levitt, 1971). Difficulty was varied by changing the interaural correlation of the noise ($\Delta\rho$) and followed the adaptation rules in Goupell and Litovsky (2014). The only major difference was if listeners could not reliably detect ICC at the easiest value (target $\rho = 0$), the adaptive procedure did not terminate early. The procedure continued to present target $\rho = 0$ trials until there were three correct answers in a row or until the completion of all of the trials. There were five simultaneous adaptive tracks of the same condition and, on a given trial, the track was randomly chosen. Each track consisted of 50 trials. Therefore, each listener experienced the same number of trials, 250 per block.

In a single trial, listeners were presented three stimuli that were separated with a 300-ms interstimulus interval. The first stimulus was always interaurally correlated ($\rho = 1$). The other intervals contained an interaurally correlated non-target and decorrelated ($\rho < 1$) target, where the order was randomized on each trial. Correct answer feedback was provided after each trial. If there was diotic level roving, the level was randomly varied across the three intervals in a trial.

Listeners performed three separate blocks of the same condition, which took approximately 45 min to complete. Since each listener only performed one CF and roving condition, there was no randomization across blocks. Thresholds were calculated by averaging the reversals that occurred in all five adaptive tracks.

3. Results

Table 1 shows the proportion of untrained listeners who could achieve threshold performance (i.e., 79.4% correct for target $\rho = 0$) for at least one of three testing blocks. Of the listeners who performed the 500-Hz roving-absent condition, most ($24/25 = 96\%$) achieved threshold performance. This is in contrast to the listeners who performed the 500-Hz roving-present condition, where only $5/14 = 36\%$ listeners achieved threshold performance. None of the listeners achieved threshold performance for either of the 4000-Hz conditions.

Figure 1 shows the individual and average ICC thresholds for the three testing blocks for the 500-Hz CF conditions. Clearly, performance was highly variable across listeners, where some could barely achieve threshold performance and some performed nearly as well as trained listeners from previous studies (shaded area or dashed line). On average, the untrained listeners in our study performed approximately a factor of 10 worse than the previously reported data in trained listeners.

A two-way analysis of variance (ANOVA) was performed on the data with factors CF and roving.¹ Thresholds did not significantly change with CF or roving ($p > 0.05$). There was a significant CF \times roving interaction [$F(1,165) = 960$, $p = 0.001$, $\eta_p^2 = 1$], which resulted from none of the listeners achieving threshold performance for either 4000-Hz condition.

To assess if rapid improvement of ICC detection occurred within the three blocks, the difference in threshold between the first and third testing block can be observed for the 29 listeners plotted in Fig. 1. Ten listeners were found to have a threshold improvement $\Delta\alpha \leq -0.05$.² However, eight listeners had almost no threshold change ($-0.05 < \Delta\alpha \leq 0$), and 11 listeners had a threshold increase ($\Delta\alpha > 0$). Therefore,

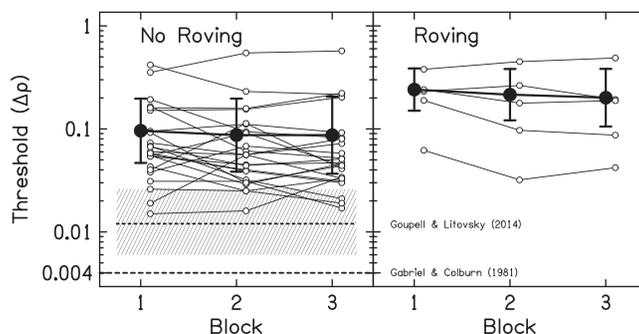


Fig. 1. ICC thresholds for the 500-Hz CF conditions for three testing blocks. The left panel shows performance when diotic level roving was absent and the right panel shows when roving was present. The average (solid circles) and individual (open circles) thresholds are shown. The dashed line represents average performance from two experienced listeners in [Gabriel and Colburn \(1981\)](#). The shaded region shows the average ± 1 standard deviation from nine listeners in [Goupell and Litovsky \(2014\)](#).

there was no significant change in threshold with testing block and there was no block \times roving interaction (two-way ANOVA with factors block and roving; $p > 0.05$ for both).

4. Discussion

This work aimed to be a starting point to characterize stimulus and listener factors that lead to individual variability in binaural tasks. The experiment demonstrated that untrained listeners exhibit a wide range of ICC sensitivity, which depends on the CF and whether diotic level roving was introduced (Table 1 and Fig. 1). Individual variability is commonly seen in some binaural experiments ([McFadden *et al.*, 1973](#); [Koehnke *et al.*, 1986](#); [Wright and Fitzgerald, 2001](#)), but not others ([Trahiotis *et al.*, 1990](#)).

It may be that some of the untrained listeners attempted to use loudness cues to perform this task because only 36% of the listeners could achieve threshold performance with level roving, in contrast to the 96% who could achieve threshold performance when level roving was absent. Furthermore, untrained listeners seem to be less able to initially use fluctuating ILDs to detect ICC because no listener could perform the task at the 4000-Hz CF where lack of phase locking to the carriers would make fluctuating ITDs inaccessible. The results of this study are in contrast to the previous literature that demonstrate that listeners are very sensitive to ICC for 10-Hz bandwidth noises at 500-Hz CF, and sometimes at 4000-Hz CF ([Goupell, 2012](#)). This difference appears to be primarily due to the experience of the listeners. The listeners in [Goupell \(2012\)](#) were tested only after several hours of training and practice, until the apparent saturation in performance had occurred. In addition, they were explicitly told to ignore monaural cues, like loudness, and rather attend to binaural cues, like image width. In contrast, the listeners of the present study were given minimal instruction on which cues to attend to during the task. However, note that interesting individual patterns in ICC performance occur across CF (see Fig. 2 in [Goupell, 2012](#)), which suggests that listeners might weight the detection cues in different ways ([Goupell and Hartmann, 2007](#); [Goupell, 2010](#); [Mao and Carney, 2014](#)).

The data from this study are also interesting when considering the percentage of listeners who could not achieve threshold performance and the notably poor performance of some listeners. [Wright and Fitzgerald \(2001\)](#), who measured the ability to detect static ITDs and ILDs, did not have listeners who could not achieve threshold performance. Therefore, there seems to be something unique about ICC detection and fluctuating ITDs and ILDs that distinguishes itself from static ITD and ILD detection. The average threshold of the untrained listeners who could achieve threshold performance in this study was a factor of 10 worse than studies that used trained listeners ([Gabriel and Colburn, 1981](#); [Goupell and Litovsky, 2014](#)). There are at least three possible reasons for this. First, the listeners in this study might all achieve thresholds comparable to the previous literature with sufficient training (longer than 45 min and over multiple days to allow for consolidation of learning). Second, the performance reported in the previous ICC detection literature is taken from listener samples that are not representative of the greater population. In other words, those listeners may have been selected, either intentionally or unintentionally, from a group of exceptionally sensitive listeners. Other reports highlight individual variability in binaural tasks ([McFadden](#)

et al., 1973; Koehnke *et al.*, 1986). It is worth noting that the listeners in Wright and Fitzgerald (2001) had thresholds after training that were noticeable higher than those in other studies (approximately 60 μ s at start and 30 μ s at end, as compared to 10 μ s). Only the best two listeners in Wright and Fitzgerald (2001) performed at levels commonly reported in the literature (e.g., Brughera *et al.*, 2013). Third, our listeners may have had interaural asymmetries that we were not aware of. An alternative explanation for some of the variability in the data may not be related to how cues are being utilized and weighted, but rather how the cues are encoded. For normal-hearing listeners, it is assumed that they have the same loudness growth and temporal modulation transfer functions across the ears. However, if differences exist, these asymmetries may cause overall poorer performance. Since the binaural system is acutely sensitive to interaural differences, it may be that seemingly small differences in monaural performance could have a large impact on binaural performance. Considering such factors may also explain the variability in performance seen in trained listeners (e.g., Goupell, 2012). The second and third explanations are also not mutually exclusive; it may be that exceptionally sensitive listeners have relatively more interaural symmetry in their monaural auditory processing.

There are a number of possible cues to perform ICC detection, some discussed in this work and likely many not discussed. Future work should focus on understanding the binaural and monaural cues that untrained listeners attend to when learning to detect ICC in narrowband noises. It is possible that our poorly performing untrained listeners confused monaural envelope fluctuations with interaural fluctuations despite correct answer feedback. When presented diotic stimuli, listeners tend to choose stimuli that have more monaural envelope fluctuations when they are asked to choose the interaurally decorrelated stimulus (Goupell and Hartmann, 2006). It is also possible after sufficient training (likely over longer time scales than the testing in this study), listeners would learn to ignore the monaural cues if they were harmful for ICC detection.

Our data showed that a subset of listeners could rapidly improve at ICC detection at 500 Hz (Fig. 1), consistent with our hypothesis that was based on the results of Wright and Fitzgerald (2001). However, other listeners showed no change or worse performance over time, therefore resulting in no improvement over the entire group. Other studies have also reported groups of non-learners (Zhang and Wright, 2009), consistent with our results. Fatigue effects or frustration may have affected the non-learners. Or it is possible that the non-learners needed rest and time for consolidation when learning to perform a new auditory task (Wright and Fitzgerald, 2001; Ortiz and Wright, 2009, 2010). The changes in ICC detection thresholds for narrowband noises are also in great contrast to the NoS π detection thresholds of a 500-Hz tone in a relatively wideband noise (Trahiotis *et al.*, 1990), even though NoS π and ICC detection are both thought to rely on changes in ρ (e.g., Durlach *et al.*, 1986; Goupell and Litovsky, 2014). In Trahiotis *et al.* (1990), listeners showed absolutely no change in performance over presumably many hours of testing, suggesting that when the bandwidth of the stimuli is large enough such that across-channel comparisons can be performed, listeners can utilize a set of detection cues that require little to no training or learning to access. Or it could be that the slow fluctuations that occur in a 10-Hz narrowband noise (Goupell and Litovsky, 2014) might initially confuse people, thus making them attend to monaural envelope fluctuations or loudness.

The data from the present study are also interesting because we know very little about transfer effects and generalization of learning ICC detection from one CF to another. Generalization must occur as none of the untrained listeners in this study could perform the ICC detection at 4000 Hz (Table 1), but listeners trained at low frequencies (e.g., 500-Hz CF) in other studies can detect ICC at 4000 Hz, sometimes exceedingly well (Goupell, 2012). For other binaural tasks like static ITD or ILD discrimination, there seems to be minimal transfer or generalization of learning from 500- to 4000-Hz CFs (Zhang and Wright, 2007; Wright and Zhang, 2009; Zhang and Wright, 2009), which would be in contrast to what the ICC data from this and other studies suggest.

In conclusion, untrained listeners demonstrated much higher thresholds than trained listeners reported in the literature; however, there was great variability in performance with some listeners near trained performance and many who could not perform the task. This work provides new insight on the cues used in ICC detection and the weighting that may occur with them. Further understanding of ICC detection could be gained from a formal multi-day ICC detection training experiment.

Acknowledgments

We would like to thank Matthew Fitzgerald for fruitful and enlightening discussions about auditory training. Matthew Fitzgerald, Ann Todd, and Steve Colburn gave helpful feedback on previous versions of this work. This work was supported by NIH Grant R00-DC010206 (M.J.G.) and P30-DC004664 (Center of Comparative Evolutionary Biology of Hearing core grant). Portions of this study were presented at the 167th Acoustical Society of America Meeting in Providence, RI.

References and links

- ¹As in Goupell and Litovsky (2014), statistical analyses were performed using interaural dissimilarity ($\Delta\alpha$ where $\alpha = \sqrt{1 - \rho^2}$), not interaural similarity ($\Delta\rho$). However, the data were transformed and reported in $\Delta\rho$ for an easier comparison to the broader literature.
- ²Here we used $\Delta\alpha$, rather than $\Delta\rho$, because it is unclear what is the best numerical scale to relate to ICC sensitivity. While not empirically shown, our previous research leads us to believe that α provides a better scale than ρ to describe perception of ICC, which is why we use α -based step sizes in our adaptive procedure (Goupell and Litovsky, 2014). Using $\Delta\alpha$ results in a more conservative estimate of threshold improvements between blocks if the first block has a relatively large threshold.
- Brughera, A., Dunai, L., and Hartmann, W. M. (2013). "Human interaural time difference thresholds for sine tones: The high-frequency limit," *J. Acoust. Soc. Am.* **133**, 2839–2855.
- Culling, J. F., Colburn, H. S., and Spurchise, M. (2001). "Interaural correlation sensitivity," *J. Acoust. Soc. Am.* **110**, 1020–1029.
- Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (1986). "Interaural correlation discrimination: II. Relation to binaural unmasking," *J. Acoust. Soc. Am.* **79**, 1548–1557.
- Edmonds, B. A., and Culling, J. F. (2009). "Interaural correlation and the binaural summation of loudness," *J. Acoust. Soc. Am.* **125**, 3865–3870.
- Gabriel, K. J., and Colburn, H. S. (1981). "Interaural correlation discrimination: I. Bandwidth and level dependence," *J. Acoust. Soc. Am.* **69**, 1394–1401.
- Goupell, M. J. (2010). "Interaural fluctuations and the detection of interaural incoherence. IV. The effect of compression on stimulus statistics," *J. Acoust. Soc. Am.* **128**, 3691–3702.
- Goupell, M. J. (2012). "The role of envelope statistics in detecting changes in interaural correlation," *J. Acoust. Soc. Am.* **132**, 1561–1572.
- Goupell, M. J., and Hartmann, W. M. (2006). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," *J. Acoust. Soc. Am.* **119**, 3971–3986.
- Goupell, M. J., and Hartmann, W. M. (2007). "Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models," *J. Acoust. Soc. Am.* **122**, 1029–1045.
- Goupell, M. J., and Litovsky, R. Y. (2014). "The effect of interaural fluctuation rate on correlation change discrimination," *J. Assoc. Res. Otolaryngol.* **15**, 115–129.
- Koehnke, J., Colburn, H. S., and Durlach, N. I. (1986). "Performance in several binaural-interaction experiments," *J. Acoust. Soc. Am.* **79**, 1558–1562.
- Lavandier, M., and Culling, J. F. (2010). "Prediction of binaural speech intelligibility against noise in rooms," *J. Acoust. Soc. Am.* **127**, 387–399.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Mao, J., and Carney, L. H. (2014). "Binaural detection with narrowband and wideband reproducible noise maskers. IV. Models using interaural time, level, and envelope differences," *J. Acoust. Soc. Am.* **135**, 824–837.
- McFadden, D., Jeffress, L. A., and Russell, W. E. (1973). "Individual differences in sensitivity to interaural differences in time and level," *Perceptual Motor Skills* **37**, 755–761.
- Ortiz, J. A., and Wright, B. A. (2009). "Contributions of procedure and stimulus learning to early, rapid perceptual improvements," *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 188–194.
- Ortiz, J. A., and Wright, B. A. (2010). "Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill," *Exp. Brain Res.* **201**, 441–451.
- Trahiotis, C., Bernstein, L. R., Buell, T. N., and Spektor, Z. (1990). "On the use of adaptive procedures in binaural experiments," *J. Acoust. Soc. Am.* **87**, 1359–1361.
- van der Heijden, M., and Joris, P. X. (2010). "Interaural correlation fails to account for detection in a classic binaural task: Dynamic ITDs dominate N0S π detection," *J. Assoc. Res. Otolaryngol.* **11**, 113–131.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2012). "Apparent auditory source width insensitivity in older hearing-impaired individuals," *J. Acoust. Soc. Am.* **132**, 369–379.
- Wright, B. A., and Fitzgerald, M. B. (2001). "Different patterns of human discrimination learning for two interaural cues to sound-source location," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12307–12312.
- Wright, B. A., and Zhang, Y. (2009). "A review of the generalization of auditory learning," *Philos. Trans. R. Soc. London Ser. B* **364**, 301–311.
- Zhang, Y., and Wright, B. A. (2007). "Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies," *J. Acoust. Soc. Am.* **121**, 2207–2216.
- Zhang, Y., and Wright, B. A. (2009). "An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults," *J. Acoust. Soc. Am.* **126**, 1349–1358.