Enhanced auditory spatial performance using individualized head-related transfer functions: An event-related potential study

Matthew G. Wisniewski, Griffin D. Romigh, Stephanie M. Kenzig, Nandini Iyer, Brian D. Simpson, and Eric R. Thompson
711th Human Performance Wing, United States Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433, USA
matt.g.wisniewski@gmail.com, griffin.romigh@us.af.mil, stephanie.kenzig.ctr@us.af.mil, nandini.iyer.2@us.af.mil, brian.simpson.4@us.af.mil, eric.thompson.28@us.af.mil

Clayton D. Rothwell
Infoscitex Corporation, Dayton, Ohio 45431, USA
clayton.rothwell.ctr@us.af.mil

Abstract: This study examined event-related potential (ERP) correlates of auditory spatial benefits gained from rendering sounds with individualized head-related transfer functions (HRTFs). Noise bursts with identical virtual elevations (0°–90°) were presented back-to-back in 5–10 burst "runs" in a roving oddball paradigm. Detection of a run's start (i.e., elevation change detection) was enhanced when bursts were rendered with an individualized compared to a non-individualized HRTF. ERPs showed increased P3 amplitudes to first bursts of a run in the individualized HRTF condition. Condition differences in P3 amplitudes and behavior were positively correlated. Data suggests that part of the individualization benefit reflects post-sensory processes.
[http://dx.doi.org/10.1121/1.4972301]
[QJF]
Date Received: July 6, 2016 **Date Accepted:** November 22, 2016

1. Introduction

Head-related transfer functions (HRTFs) capture location-dependent alterations to a sound caused by the physical interaction with a listener's head, shoulders, and outer ears. Sound sources can be filtered with these HRTFs in a manner that allows presentation of sources in virtual space through ear- or headphones. However, because anthropometric features (e.g., head size, outer ear shape) vary across listeners, auditory spatial maps are highly specific to an individual.[1] Virtual sounds are perceived more accurately in space when stimuli are rendered with individualized HRTFs than non-individualized HRTFs (e.g., an HRTF recorded from an acoustic manikin).[1–3] The general belief is that differences in behavioral measures when using individualized and non-individualized HRTFs result from a mismatch of spectral features in a non-individualized HRTF to one's own natural HRTF. For instance, both HRTFs may contain distinct peak and notch features, but those features may appear at different center frequencies and with different bandwidths.[1–3]

The goal of this study was to determine whether or not detectable event-related potential (ERP) correlates of the benefits gained from using individualized HRTFs exist. This question serves practical and theoretical purposes. If an index of the individualized HRTF benefit exists in the ERP, that index could potentially be used as an objective measure of spatial perception when behavioral differences are difficult to observe (e.g., at ceiling level performance). The temporal resolution of ERPs also makes the method useful for characterizing the cognitive processes that underlie auditory spatial perception. For instance, effects may manifest in early components of the ERP (e.g., N1) generated largely within the auditory system itself,[4] or in later components (e.g., P3) associated with domain-general cognitive processing.[5,6]

To our knowledge there have been no ERP studies attempting to derive an index that correlates with benefits gained from using individualized compared to non-individualized HRTFs. Here, listeners were tasked with detecting changes in the elevation of 250-ms noise bursts rendered with their own HRTF, or an HRTF recorded from an acoustic manikin. ERPs elicited by bursts at the onset of an elevation change were examined. Relationships between elevation change detection and ERP amplitudes were also analyzed. It was hypothesized that detection accuracy and the amplitude of

ERP components elicited by elevation change would be enhanced in the individualized HRTF condition.

2. Methods

2.1 Listeners

Thirteen individuals with normal audiometric thresholds [<20 dB hearing level (HL), 0.25–8 kHz] participated in exchange for monetary compensation. All were experienced listeners in psychoacoustic studies, some studies of which involved localizing virtual stimuli. One individual was dropped because of noisy electroencephalogram (EEG) data. The final sample included 12 listeners.

2.2 Stimuli and apparatus

HRTFs were measured for each individual listener and for the Knowles Acoustic Manikin for Acoustic Research (KEMAR) in the Auditory Localization Facility at Wright-Patterson Air Force Base, OH. HRTFs were level normalized so all individual HRTFs and the KEMAR HRTF had the same root-mean-square level at the location directly in front of the listener. Further details on HRTF measurement procedures are described elsewhere.[2,3]

All experimental procedures were executed in MATLAB R2013a. Listeners heard sounds over Etymotic ER-2 earphones in a sound-attenuating audiometric booth. Sounds were presented at a comfortable listening level (<82 dB sound pressure level).

2.3 Task

Figure 1(a) gives a graphical representation of the task. Bursts of white noise (250-ms, 10-ms on- and off-ramps) were filtered with either KEMAR's HRTF or an individual's personal HRTF, rendering sounds at the following virtual elevations: 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, and 90° along the median plane. Bursts were presented back-to-back (onsets separated by 1024 ms) in grouped "runs" having the same elevation. The length of these runs, and the elevation of sounds presented within them, were selected randomly for each run. Similar "roving oddball" paradigms are commonly used in ERP work.[7] In Fig. 1(a), a run of six bursts convolved with an HRTF associated with 60°, is followed by a run of nine bursts similarly convolved to appear at an elevation of 80°, followed by a run of five bursts at 10°. Run lengths varied from 5 to 10 bursts long. The task of a participant was to detect the start of each run (i.e., to detect a change in elevation) with a button press. Sixty runs were presented per block. There were three blocks per HRTF condition. Blocks were pseudorandomly ordered so that no more than two blocks with the same condition could occur back-to-back.

Selecting elevations at random for each run within a block yields a greater amount of small degree separations. A histogram showing the number of changes at each level of degree separation for a representative individual is shown in Fig. 1(b). Critically, this paradigm allows the collection of behavioral and electrophysiological data relating to changes in space without relying on any single degree of change or one particular spatial location. The current paradigm aligns with the classic auditory oddball paradigm in EEG studies, which allows examination of known ERP components (e.g., P3) elicited by changes in stimulus presentation.[4,6,7]

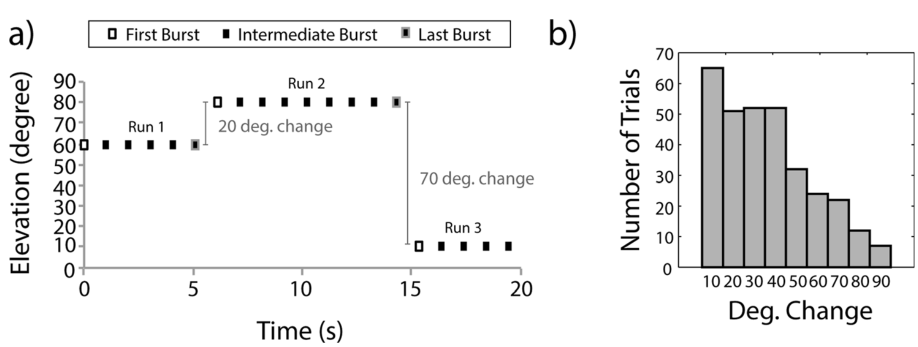

Fig. 1. A depiction of the task used in the experiment (a). Within a block, several runs of bursts from a randomly selected elevation were presented with run lengths varying randomly between 5 and 10 bursts. First, intermediate, and last bursts of a run are labeled. The selection of elevations at random yielded a distribution of degree separations with low degree separations (e.g., 10°) occurring more often. The distribution of degree separations for a representative individual is shown (b).

2.4 Electrophysiology

Data were collected from a 39-channel array of electrodes with a BioSemi Active II system (BioSemi, Amsterdam, Netherlands), at a 2048 Hz sampling rate, and 24-bit A/D resolution. Data were referenced online to the Common-Mode-Sense/Driven-Right-Leg reference of the BioSemi system (see www.biosemi.com). Thirty-two electrodes were fixed within a cap and arranged according to the international 10–20 system. Seven additional electrodes were placed at the mastoids, on lateral sides, and below each eye and the nose tip.

All offline analyses were performed using EEGLAB (Ref. 8) and custom MATLAB functions. The data were offline referenced to the nose tip, resampled at 512 Hz, and digitally bandpass filtered between 0.5 and 100 Hz. Portions of data and channels contaminated by excessive noise or movement artifacts were removed. Remaining data were submitted to independent components analysis. Independent components identified as artifacts (e.g., blinks) were subsequently removed from the channel data.[5,8]

Epochs were extracted from -0.2 to 0.8 s surrounding the onset of the first burst within each run (excluding the first run of a block). First burst epochs corresponding to "hit" trials in which a listener responded within 2 s following the first burst's onset were separated from epochs identified as "miss" trials (i.e., epochs in which the listener did not respond). Only "hit" epochs were formally analyzed. The minimum number of epochs used to compute any ERP was 46 ($M = 74.50$, $SD = 19.32$). After subtracting mean baseline voltages (-0.2 to 0 s), ERPs were created by averaging epoch voltage time-courses.

2.5 Analyses and statistics

Analyses were conducted on mean voltage within time-windows corresponding to N1 (90–110 ms), P2 (150–200 ms), and P3 (300–500 ms) components of the ERP at each of the 32 scalp electrodes. A non-parametric permutation method was used to compare component amplitudes generated under individualized and non-individualized HRTF conditions. A null-hypothesis distribution of t-values was generated by randomly assigning condition labels for each individual's ERP waveforms for 1000 iterations. A t-statistic was calculated for each iteration, creating a distribution of t-values expected under the null hypothesis. The p-values for real data were considered to be the proportion of iterations having more extreme t-values than the real data.[9] At this step, significant differences were identified at an alpha level of 0.05. To control for multiple comparisons, p-values were adjusted using the false-discovery rate procedure[10] separately for each component.

To assess the relationship between ERPs and behavior, ERP component amplitudes showing significance were correlated with detection accuracy. Specifically, the difference between HRTF conditions (individualized–non-individualized) for P3 amplitudes and the A' signal detection measure of behavioral accuracy were submitted to Spearman rank-order correlations. Spearman correlations were used rather than Pearson correlations in order to minimize the influence of outliers in typically non-normally distributed EEG data.[9] In addition to assessing the relationship between behavioral accuracy and P3 amplitude, the relationship between HRTF spectral variability and P3 amplitude was also assessed. Spectral variability in an individual's own HRTF was defined as the average standard deviation of level across the tested elevations. That is, standard deviation was computed across elevations for each frequency. The mean standard deviation across frequencies was then considered spectral variation. The spectral variation measure served to test the hypothesis that differences in ERPs reflect differences in spectral qualities of sources filtered with individualized and non-individualized HRTFs, rather than any perceived spatial qualities. The difference between the generic KEMAR and an individual's own HRTF was also assessed for a relationship with P3 amplitude. Here, we used an average spectral correlation metric. For each tested elevation on the median plane ($0°$–$90°$; $10°$ increments), we computed the Pearson correlation coefficient for the left ear, then averaged over all elevations.[11] This average spectral correlation metric was used to examine a potential relationship between HRTF similarity and ERP differences.

3. Results

3.1 Behavior

Elevation change detection "hit" rates (H) were defined as the proportion of times a listener indicated that they perceived a change in elevation within 2 s of an actual elevation change. "False alarm" rates (F) were defined as the proportion of intermediate bursts [see Fig. 1(a)], excluding those already labeled as hits, in which a change was

Table 1. Detection sensitivity (A'). Numbers in parentheses are standard errors of the mean.

	Non-individualized	Individualized
A'	0.55 (0.01)	0.60 (0.02)

indicated within 2 s. The A' signal detection measure of sensitivity for each condition is shown in Table 1. A' is equal to $0.5 + (\text{H-F})(1 + \text{H-F})/4\text{H}(1\text{-F})$ when $\text{H} \geq \text{F}$, and $0.5 - (\text{F-H})(1 + \text{F-H})/4\text{F}(1\text{-H})$ when $\text{H} \leq \text{F}$.[12] A' varies between 0 and 1, with greater A' indicative of greater sensitivity. Corroborating previous research, the detection of changes in elevation was aided by employing an individualized HRTF, $t(11) = 3.93$, $p = 0.002$, Cohen's $d = 0.91$. All individuals performed best with their own HRTF.

3.2 Electrophysiology

Figure 2 shows ERPs at channel Pz time-locked to the first burst of a run (i.e., changes in elevation), along with scalp maps of condition differences in voltage for N1, P2, and P3 components (individualized minus non-individualized). Readers interested in viewing ERP waveforms at each scalp channel can refer to the supplemental materials.[13] ERPs at Pz show clear N1, P2, and P3 components for both conditions. Though there is a qualitative trend for larger amplitudes of N1 in the individualized HRTF condition, this trend did not reach significance at any electrode. P2 amplitudes were comparable for the two conditions and also showed no significant effects. In contrast, the P3 component appears larger in amplitude when sounds are presented under the individualized HRTF condition than the non-individualized HRTF condition. The differences in voltage between conditions for P3 shows that enhanced P3 amplitude for the individualized HRTF condition is primarily observed at posterior scalp locations. Electrodes showing significant differences are marked with stars ($p < 0.05$, false discovery rate corrected) in Fig. 2.[14]

3.3 ERP/behavior relationship

Figure 3 shows scatterplots of P3 amplitude differences at Pz between HRTF conditions (y-axes) in relation to detection accuracy differences, spectral variability, and the similarity between one's own HRTF and the non-individualized HRTF (x-axes). A significant positive relationship was found for P3 amplitude differences and detection accuracy differences between conditions, Spearman's $\text{Rho}(10) = 0.60$, $p = 0.043$. Individuals who showed a larger behavioral individualization benefit showed a larger difference between HRTFs in P3 amplitudes. Neither one's own individualized HRTF spectral variability nor the similarity of one's own HRTF to KEMAR's HRTF appeared to relate to the observed P3 amplitude differences, $p > 0.20$.

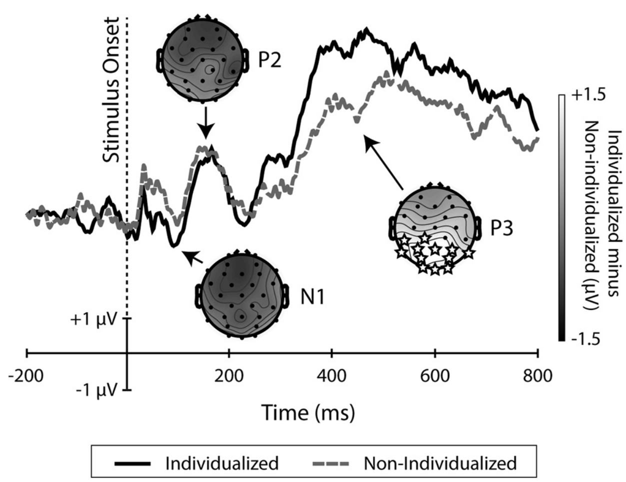

Fig. 2. ERPs time-locked to the first burst of a "run" in the individualized (solid black line) and non-individualized (dashed grey line) HRTF condition at electrode Pz. Scalp maps of differences in voltage in the defined N1 (90–110 ms), P2 (150–200 ms), and P3 (300–500 ms) time-windows are shown. Electrode locations are represented by black dots. Stars mark electrodes showing a statistically significant difference between conditions after false discovery rate correction ($p < 0.05$).

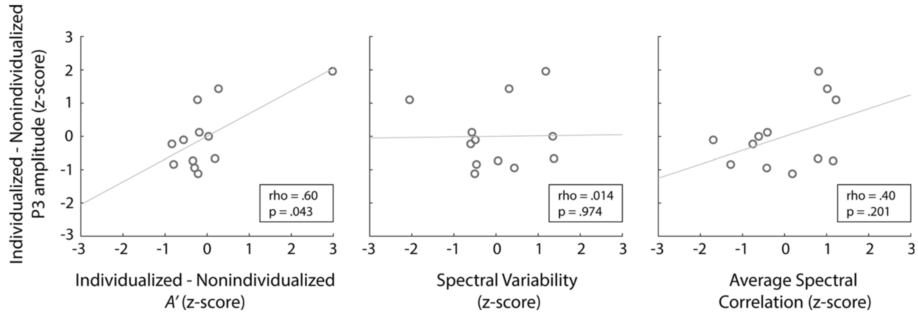

Fig. 3. Scatterplots of HRTF condition differences in P3 amplitudes (y-axes) as a function of differences in A' between conditions, spectral variability of one's own HRTF, and the similarity between one's own HRTF and the nonindividualized HRTF (x-axes). Lines show best linear fits of the data.

4. Discussion

In a simple elevation change detection paradigm, we found that individualized HRTFs enhanced detection sensitivity, corroborating several other behavioral works that have demonstrated benefits of individualized HRTFs over non-individualized HRTFs for localization.[1–3] Though early ERP responses did not appear to be strongly affected by individualization, ERPs elicited by changes in elevation showed enhanced P3 amplitudes to the first burst of a run in the individualized compared to the non-individualized HRTF condition. Also, the degree of difference in P3 amplitudes between conditions was correlated with the degree to which one benefited from an individualized HRTF.

To our knowledge, these results present the first demonstration of an effect of individualized HRTFs on neural responses to elevation changes in humans. Investigations of auditory spatial perception in the horizontal plane have noted similar parallels with behavior. Teder-Sälejärvi and Hillyard[15] found that N1 and P3 amplitudes were largest for sounds presented at attended azimuths. Though both N1 and P3 amplitudes paralleled detection rates, the P3 was more narrowly tuned (i.e., amplitudes decreased sharply with distance from the attended speaker). They concluded that two stages of spatial processing were at play. One early stage related to processing in auditory cortex with a rougher representation of spatial relevance (reflected in the N1), followed by a later selection process in non-auditory brain regions (reflected in the P3). In a follow-up study the same group reported that late-blind individuals showed a localization benefit compared to sighted controls that was paralleled by larger P3 amplitudes.[16] Expanding on the work of Teder-Sälejärvi and Hillyard[15] those authors concluded that late-blind individuals learn to perceive location differences at a late processing stage of auditory space related to selection and decision-making. Following this work on spatial perception in the horizontal plane, our results support a view of HRTF effects and auditory spatial perception that at least partially includes advantages in processing at a post-sensory level (for an extended discussion, see Wisniewski *et al.*[17]).

The results we present are preliminary and do require cautious interpretation for several reasons. First, we have no way to confirm that listeners heard sounds as varying in the elevation dimension. Given that the cues to elevation are mainly available in the high-frequency range of a sound's spectrum, listeners could perform the task using some sort of spectral analysis without ever perceiving sounds at spatially distinct locations. One also needs to consider that we have only sampled a small range of auditory space. It remains to be seen whether similar effects might occur for stimuli in the front-back dimension, or at other locations off the median plane. Nevertheless, this work shows that the combination of ERP and behavioral techniques can inform practical and theoretical research directions in HRTF research. Future experiments that utilize the P3 component will shed further light on the cognitive processes that underlie advantages gained from using individualized HRTFs.

Acknowledgments

M.G.W. was supported by a National Research Council postdoctoral fellowship and an Oak Ridge Institute for Science and Education fellowship at the U.S. Air Force Research Laboratory. S.M.K was supported by an internship provided by the Oak Ridge Institute for Science and Education.

References and links

[1]E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123 (1993).
[2]G. D. Romigh and B. D. Simpson, "Do you hear where I hear?: Isolating the individualized sound localization cues," Front. Neurosci. **8**, 370 (2014).
[3]G. D. Romigh, B. D. Simpson, and N. Iyer, "In ear to out there: A magnitude based parameterization scheme for sound source externalization," in *Proceedings of the 22nd International Conference on Auditory Display*, Canberra, Australia (2016).
[4]R. Näätänen and T. Picton, "The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure," Psychophysiology **24**, 375–425 (1987).
[5]M. G. Wisniewski, E. Mercado III, K. Gramann, and S. Makeig, "Familiarity with speech affects cortical processing of auditory distance cues and increases acuity," PLoS One **7**, e41025 (2012).
[6]J. Polich, "Updating P300: An integrative theory of P3a and P3b," Clin. Neurophysiol. **118**, 2128–2148 (2007).
[7]M. I. Garrido, K. J. Friston, S. J. Kiebel, K. E. Stephan, T. Baldeweg, and J. M. Kilner, "The functional anatomy of the MMN: A DCM study of the roving paradigm," Neuroimage **42**, 936–944 (2008).
[8]A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent components analysis," J. Neurosci. Methods **134**, 9–21 (2004).
[9]M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice* (MIT Press, Cambridge, MA, 2014).
[10]Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. R. Stat. Soc. B **57**, 289–300 (1995).
[11]J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624 (1992).
[12]N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide* (Cambridge University Press, New York, NY, 1991).
[13]See supplemental material at http://dx.doi.org/10.1121/1.4972301 for butterfly plots of ERP data in individualized and non-individualized HRTF conditions.
[14]The difference between conditions in P3 amplitude and the interpretation of significance was not sensitive to the time-window chosen for analysis. For instance, a larger window from 300–800 ms yielded the same effect. Further, an alternative measure of amplitude (maximum amplitude within the P3 time-window) led to the same conclusions. P3 component latencies, not reported here, did not differ significantly between conditions, $p > 0.05$ (false discovery rate corrected).
[15]W. A. Teder-Sälejärvi and S. A. Hillyard, "The gradient of spatial auditory attention in free field: An event-related potential study," Percept. Psychophys. **60**, 1228–1242 (1998).
[16]A. Fieger, B. Röder, W. Teder-Sälejärvi, S. A. Hillyard, and H. J. Neville, "Auditory spatial tuning in late-onset blindness in humans," J. Cogn. Neurosci. **18**, 149–157 (2006).
[17]M. G. Wisniewski, E. Mercado III, B. A. Church, K. Gramann, and S. Makeig, "Brain dynamics that correlate with effects of learning on auditory distance perception," Front. Neurosci. **8**, 396 (2014).