Acuity of spatial stream segregation along the horizontal azimuth with non-individualized head-related transfer functions

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Acuity of spatial stream segregation along the horizontal azimuth with non-individualized head-related transfer functions

Martha M. SHIELL[1]; Elia FORMISANO[1]

[1] Maastricht University, Netherlands

ABSTRACT

Auditory spatial cues help the nervous system segregate features of a soundwave into distinct streams. To study this process, we must allot for changes in auditory spatial acuity along the horizontal azimuth, and some evidence suggests that this relationship differs for concurrent versus consecutive sounds. Here, we developed a paradigm to measure the change in spatial stream segregation along the horizontal azimuth and validate the effectiveness of non-individualized head-related transfer functions, the most easily-accessed form of auditory spatialization, in this procedure. We tested 18 normal-hearing adults using anthropometrically-matched non-individualized head-related transfer functions. We applied a spatial stream segregation task where participants identified the rhythm of a target stream presented with a spatially-separated masker. Spatial separation varied according to an adaptive staircase procedure, and thresholds were calculated both near the midline and in the far left periphery. This work will assist neuroscientists and others in the design of stimuli for eliciting brain activity related to spatial stream segregation.

Keywords: Spatial stream segregation, auditory spatial acuity

1. INTRODUCTION

Typically when we think of auditory spatial information we think of how it helps us understand where sounds originate from in our environment – that is, the process of localization. But auditory spatial information has a second, often overlooked role in perception. Concurrent sounds from the environment mix together and hit our eardrums as a single soundwave, and the auditory system needs to dissect this messy signal to make sense of it. Multiple acoustic cues contribute to this scene analysis, one of which is spatial information: The further apart two sounds are, the more likely they are to be heard as two separate sources, a process known as spatial stream segregation (SSS).

Research with neuropsychology patients indicates that SSS and localization behaviours may be disociable: Some brain-lesioned patients show preserved SSS with impaired localization while others show preserved localization with impaired SSS (1). While the neural mechanisms supporting each behaviour are not fully understood, the few neuroscience studies that have examined neural responses to concurrent spatially-separated sounds confirm that the brain activity patterns elicited by these stimuli are not merely combinations of the patterns elicited by single sound locations (2-5). A better understanding of the neural mechanisms may have clinical impact by highlighting potential factors that can be optimized in hearing aids and cochlear-implants.

To manipulate SSS for experimental study, neuroscientists can vary the spatial separation between concurrent sounds. However, every change in spatial separation has a corresponding change in spatial location that may trigger localization processing, and thus raises a potential confound in the interpretation of experimental measurements. This confound can be controlled for with conditions where the locations of concurrent sounds change but SSS is held constant. To do so requires careful monitoring though of the relationship between SSS and horizontal azimuth: Since listeners have better spatial acuity for sounds in front of them than sounds in their periphery, greater spatial separations are required in the latter to achieve the same behaviour (6-9).

This difference in acuity across the azimuth can largely be explained by physical differences in the soundwave as it reaches each ear (i.e. binaural cues vary at a faster rate around the midline than in the periphery). However, there is evidence that our perception of spatial separation between concurrent sounds does not merely follow the physics of binaural cues, but has a psychophysical component as well. Evidence for this psychophysical component comes from the finding that there is a "sweet-spot" for sensitivity to changes in spatial separation between concurrent sounds: We are best at judging changes that occur around six degrees of separation, irrespective of the location along the horizontal azimuth (10). Note that this sensitivity cannot be explained by the physics of binaural cues. This dissociation between SSS and binaural cues justifies measuring the SSS-azimuth relationship directly, rather than inferring it from measures of binaural cues.

In the current study, we developed a paradigm to characterize the relationship between SSS and horizontal azimuth for the purpose of stimulus design in neuroscience experiments. To do so, we adopted the spatial rhythmic-masking release task from Middlebrooks and Onsan (2012). This task is a direct proxy of SSS. It uses streams of broadband noise bursts, which are ideal for brain imaging because they will elicit brain activity that (1) is independent of ecological associations (unlike, for example, semantic information in language), and (2) avoid interactions with frequency.

Whereas Middlebrooks and Onsan (2012) report on thresholds in highly-practiced participants, we evaluated performance in typical inexperienced listeners, whom are more readily available for neuroscience experiments. Furthermore, since brain imaging experiments often require sound delivery through headphones, we generated virtual auditory space through head-related transfer functions rather than free-field. Because facilities and resources for measuring these functions are not always available or feasible for neuroscience experiments, we used non-individualized head-related transfer functions (NI-HRTF) from a publicly available library with corresponding anthrompometry data to match with participants. Finally, we selected different locations than those that were previously studied: In order to maximize potential differences in corresponding brain activity measurements, we measured SSS at the most peripheral location available in our library of NI-HRTFs (80 degrees). It was also important to limit the locations to a single hemifield, in order to simplify the interpretation of potential hemispheric interactions in brain activity. Because concurrent spatially-separated sounds can push apart the perceived locations of one another (11), we measured SSS at a location that was offset from the midline (10 degrees) to mitigate the risk of the perception of the target's location being pushed across the midline into the contralateral hemifield.

2. METHODS

2.1 Participants

The study was approved by the Ethical Review Committee for the Faculty of Psychology and Neuroscience at Maastricht University. Participants were recruited from the Maastricht University undergraduate psychology program. All gave informed written consent and were compensated for their time either by course credits or gift vouchers. Eighteen healthy adults with self-reported normal hearing participated in the experiment (4 males and 14 females, age in years old: mean = 22.5, range = 18 - 33). All participants were right-handed but one female who was ambidextrous.

2.2 Materials

All stimuli were created with Matlab (Mathworks Inc.). NI-HRTFs were taken from CIPIC repository (12,13). Stimuli locations that were mid-HRTF measurements were interpolated as needed. Training and measurements were completed with Psychopy (14). All stimuli were presented with Sensimetric earbuds (model S14) at a comfortable sound level in a sound-attenduated booth.

2.3 General Procedure

Over two days that were 1-7 days apart, participants completed two 60-90 minute-long sessions of testing. In addition to the experimental task, the first session included NI-HRTF selection (~5 minutes) and brief task training consisting of instructions, stimulus examples, and a practice run of the experimental task (~3 minutes).

2.4 Selection of Non-individualized Head-Related Transfer Function (NI-HRTF)

For each participant, we manually measured head width and depth with a carpenter's square and a ruler. These two measures were chosen based on their relevance for spatial hearing: head width is highly-correlated with interaural timing delays (ITD) (12,15), a feature that may be essential to SSS (9), and head-depth is the best predictor of a listener's preferred NI-HRTF (16). ITD predicted from these variables share 76% of variance with actual ITD (12).

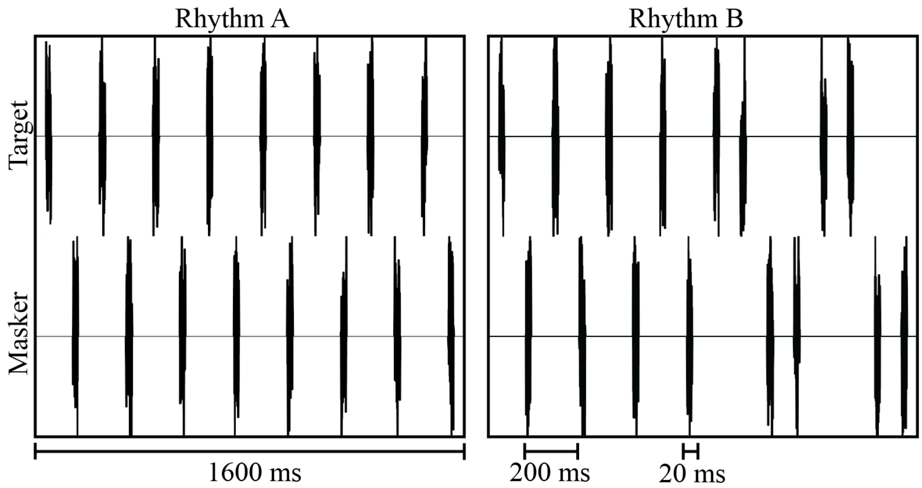

Figure 1 - Examples of target and masker streams for rhythms A and B in rhythmic-masking release task.

NI-HRTFs from the CIPIC repository (12,13) were ordered based on the Euclidean distance of their corresponding anthropometry from the participant's measures of head width and depth. The five NI-HRTFs with the closest measures were selected. We then asked participants to subjectively evaluate which of these five NI-HRTFs provided the most realistic representation of auditory space, using a procedure adapted from Seeber and Fastl (17). In this procedure, participants were instructed to listen to five examples of auditory space, in any order and as many times as needed. Each example consisted of five 200 ms noisebursts, presented every 400 ms sequentially to the left hemifield at -80, -60, -40, -20, and 0 degrees (i.e. at the midline). In each example, these locations were simulated via one of the NI-HRTFs. The participants were asked to base their evaluation on the following criteria: The sound (1) moved from the left to the midline, (2) moved in equally-spaced steps, (3) moved at a constant elevation, (4) and a constant distance, (5) far away from the participant.

2.5 Measurement of Spatial Stream Segregation (SSS)

SSS was measured using the spatial rhythmic masking release task (RMR) (9). This measurement was conducted under two conditions, termed peripheral and midline, according to where the participant was instructed to direct their spatial attention.

In each trial, a target stream of noise bursts was presented at the attended location in one of two temporal patterns, termed rhythms A and B, illustrated in figure 1.The rhythm was randomized between trials. The leading stream (the "Target" in the case of Figure 1) was alternated between target and masker locations in a random order to ensure that its' onset could not be used as a cue for the task. The noise bursts were 20 ms in duration, with 5 ms ramps on and off. Combined, the two streams produce a rate of 10 noise-bursts per second. The rhythm was repeated for a maximum of 6400 ms. The participant was asked to identify the rhythm by button-press and could respond at any moment after stream onset, which terminated the trial. Feedback was displayed for 500 ms between trials in order to maintain participant motivation.

Concurrent to the target stream, a masker stream was presented (Figure 1). Because both the masker and target streams consisted of unfrozen noise bursts, if the two streams are co-located the participant will hear a single stream, and it is impossible to discriminate between rhythms A and B. Likewise the task becomes easier as the spatial separation between the streams increases. Thus the spatial separation between masker and target was varied according to an adaptive staircase procedure in order to determine a threshold level for effective stream segregation. For each correct response, in the following trial the spatial separation was decreased by one step, and for each incorrect response, it was increased by three steps. The masker was always to the right of the distractor such that both streams remained within the same hemifield. Hence in the peripheral condition the target was fixed at -80 degrees, and in the midline condition the masker was fixed at -10 degrees (to the left, where midline is 0 degrees). Although this introduced a confound of a static versus changing location of the target in the peripheral versus midline conditions, participants reported during pilot testing that this was easier than switching attention between the left and right streams in the two conditions. Futhermore, it allowed the target to maintain the maximum potential benefit from the head-shadow effects relative to the masker in both conditions.

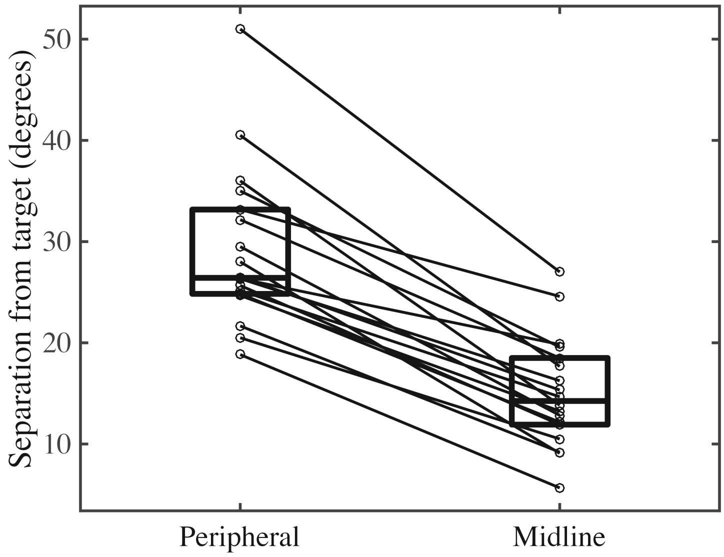

Figure 2 – Spatial separation thresholds between target and masker in the peripheral and midline conditions of the rhythmic-masking release task: Individual (circles and diagonal lines), group median (middle horizontal lines), and interquartile range (rectangles).

The one-up one-down staircase with 3:1 ratio in step size provides two advantages: (1) It allows the staircase to vary at a faster rate than a typical transformed staircase (e.g. following a 1-up-3-down rule, (18)), thereby sampling space more efficiently, and (2) it quickly removes participants from below threshold levels, which may be presented by chance (19). Although the threshold where this staircase procedure converges is not established (20), we were not primarily concerned with the threshold detection level, but rather with its replicability across sessions (see analysis section below), which is essential for neuroimaging experiments involving measurements on different days.

Additionally, we were concerned that the convergence of the staircase was consistent between the midline and peripheral conditions. The step size of a staircase can affect the calculated threshold, which presents a challenge in the current paradigm because the resolution of spatial cues is not equal between the peripheral and midline conditions. Thus, a step size in the midline condition will potentially have more binaural change than the same absolute step size in the periphery, leading to differences in thresholds between the two conditions that are merely due to the sampling of testing points. Short of measuring binaural cues for each participant across the azimuth, the required difference in step size between the periphery and midline can only be estimated. In the original RMR task, Middlebrooks and Onsan (2012) used step size of 2.5 and 5 degrees at 0 and 45 degree locations, respectively. We used the same values in our paradigm, and evaluated this choice *post hoc* (see analysis section below).

The staircase terminated after 15 reversals, which took four minutes on average. Participants were told to take short rests as needed between runs, and forced to exit the testing booth for a longer break after every three runs. Runs of the midline and peripheral conditions were alternated, and participants completed 6 runs of each in each session, for a total of 12 runs in each condition.

3. ANALYSIS AND RESULTS

We calculated thresholds for each run as the mean spatial separation between target and masker across all fifteen reversals. We calculated thresholds for each participant as their median threshold across twelve runs for each of the peripheral and midline conditions. Participants' individual thresholds, group median and interquartile range are shown in figure 2. In the peripheral condition, the median threshold (26.4 degrees) was approximately 1.9 times that of the midline (14.3 degrees), but participants varied in this relationship.

To assess the correspondence of the chosen step sizes in the peripheral and midline conditions, we

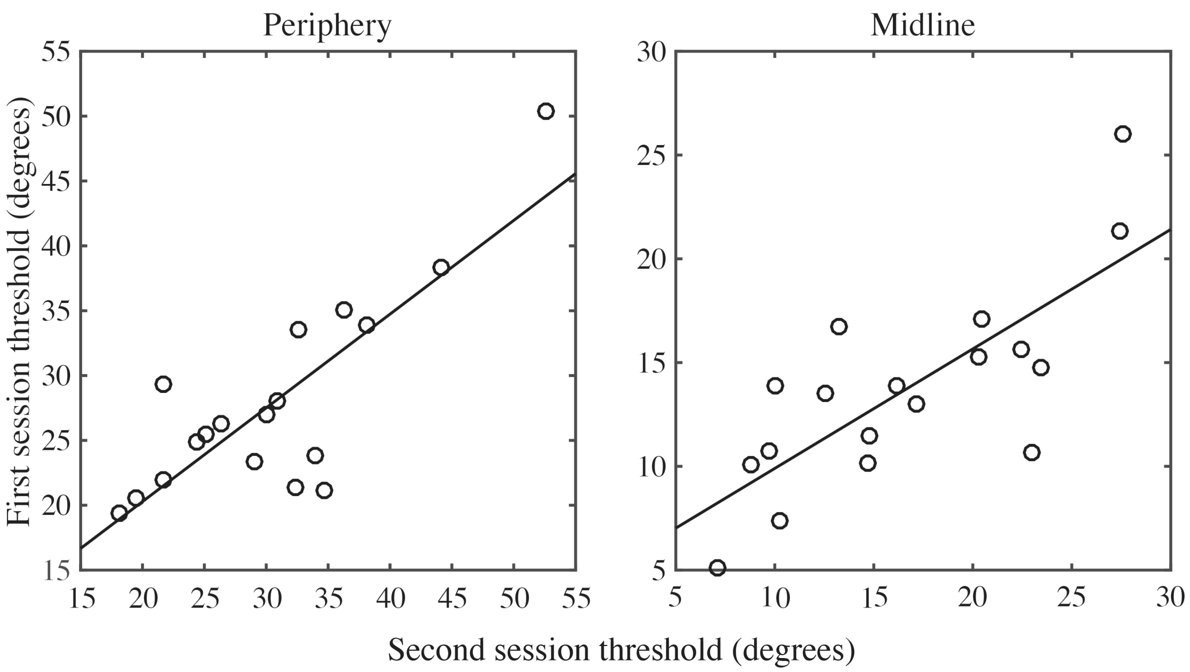

Figure 3 – Reliability of spatial separation thresholds in the rhythmic-masking release task across two sessions, in the peripherial and midline conditions.

used a Wilcoxon signed-rank test to compare participants actual accuracy scores on the staircase testing level nearest to their calculated threshold. If the 5 and 2.5 degree step sizes in the peripheral and midline conditions respectively are not sufficiently matched, then these accuracy scores will differ. Nearest-threshold accuracy scores were similar between peripheral and midline conditions (group median degrees: peripheral = 0.81, midline = 0.80) and we found no evidence for a difference between them ($W = 77$, $p = 0.64$).

To assess the replicability of the staircase over sessions, we measured the Pearson correlation between thresholds from the first and second sessions, shown in Figure 3. Thresholds correlated across sessions (periphery: $R = 0.82$, $p = 0.0002$; midline: $R = 0.76$, $p = 0.0002$), indicating that the task was replicable. To test if there were effects of learning, we compared group median scores across sessions with a Wilcoxon-signed rank test. Thresholds in the first session were higher than those in the second session (group medians in degrees for session 2, 1: periphery = 25.9, 30.4; midline = 13.7, 15.5; Wilcoxon-signed rank: peripheral: $W = 118$, $p = 0.049$; midline: $W = 144$, $p = 0.011$) indicating that there may be a learning effect.

4. SUMMARY

We describe here a reliable paradigm for measuring the relationship between SSS and horizontal azimuth to be used for the design of stimuli in the study of the neural correlates of SSS. We present data from a typical group of healthy inexperienced participants. The staircase procedure produces a threshold for 0.805 accuracy, which is consistent across the tested locations in the midline and the periphery. Participants showed improved performance on the task on the second day of testing, indicating a learning effect that should be taken into account in future applications of this paradigm.

ACKNOWLEDGEMENTS

This research was supported by grant NWO VICI 453-12-002 to EF. MS was supported by a postdoctoral fellowship from the Canadian Institutes of Health Research.

REFERENCES

1. Duffour-Nikolov C, Tardif E, Maeder P, Thiran AB, Bloch J, Frischknecht R, et al. Auditory spatial deficits following hemispheric lesions: Dissociation of explicit and implicit processing. Neuropsychol Rehabil. Psychology Press; 2012 Sep 20;22(5):674–96.

2. Shiell MM, Hausfeld L, Formisano E. Activity in Human Auditory Cortex Represents Spatial Separation Between Concurrent Sounds. J Neurosci. 2018 May 23;38(21):4977–84.
3. Middlebrooks JC, Bremen P. Spatial stream segregation by auditory cortical neurons. J Neuro. 2013 Jul 3;33(27):10986–1001.
4. Maddox RK, Billimoria CP, Ben P Perrone, Shinn-Cunningham BG, Sen K. Competing Sound Sources Reveal Spatial Effects in Cortical Processing. Plos Biol. Public Library of Science; 2012;10(5):e1001319.
5. Zatorre RJ, Bouffard M, Ahad PA, Belin P. Where is "where" in the human auditory cortex? Nat Neuro. 2002 Aug 19;5(9):905–9.
6. Best VA, van Schaik A, Carlile S. Separation of concurrent broadband sound sources by human listeners. J Acoust Soc Am. 2004 Jan;115(1):324–36.
7. Perrott DR. Concurrent minimum audible angle: A re‐examination of the concept of auditory spatial acuity. J Acoust Soc Am. Acoustical Society of America; 1984 Jun 4;75(4):1201–6.
8. Divenyi PL, Oliver SK. Resolution of steady‐state sounds in simulated auditory space. J Acoust Soc Am. Acoustical Society of America; 1998 Jun 4;85(5):2042–52.
9. Middlebrooks JC, Onsan ZA. Stream segregation with high spatial acuity. J Acoust Soc Am. Acoustical Society of America; 2012;132(6):3896–911.
10. Carlile S, Fox A, Orchard-Mills E, Leung J, Alais D. Six Degrees of Auditory Spatial Separation. J Assoc Res Otolaryngol. Springer; 2016 Jun 1;17(3):209–21.
11. Lee AKC, Deane-Pratt A, Shinn-Cunningham BG. Localization interference between components in an auditory scene. J Acoust Soc Am. Acoustical Society of America; 2009 Nov;126(5):2543–55.
12. Algazi VR, Duda RO, the DTWO, 2001. The CIPIC HRTF database - IEEE Conference Publication. New Paltz, USA; 2001.
13. Andreopoulou A, Roginska A. Towards the Creation of a Standardized HRTF Repository. New York; 2011. pp. 1–7.
14. Peirce JW. PsychoPy—Psychophysics software in Python. J Neurosci Methods. 2007 May;162(1-2):8–13.
15. Middlebrooks JC. Individual differences in external-ear transfer functions reduced by scaling in frequency. J Acoust Soc Am. Acoustical Society of America; 1999 Aug 23;106(3):1480–92.
16. Schonstein D, Katz BFG. HRTF selection for binaural synthesis from a database using morphological parameters. Sydney; 2010.
17. Seeber BU, Fastl H. Subjective selection of non-individual head-related transfer functions. Georgia Institute of Technology; 2003 Jul 1.
18. Levitt H. Transformed Up‐Down Methods in Psychoacoustics. J Acoust Soc Am. Acoustical Society of America; 2005 Aug 3;49(2B):467–77.
19. Kaernbach C. Simple adaptive testing with the weighted up-down method. Perception, & Psychophysics. Springer; 1991;49(3):227–9.
20. García-Pérez MA. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. Vision Res. Elsevier; 1998;38(12):1861–81.