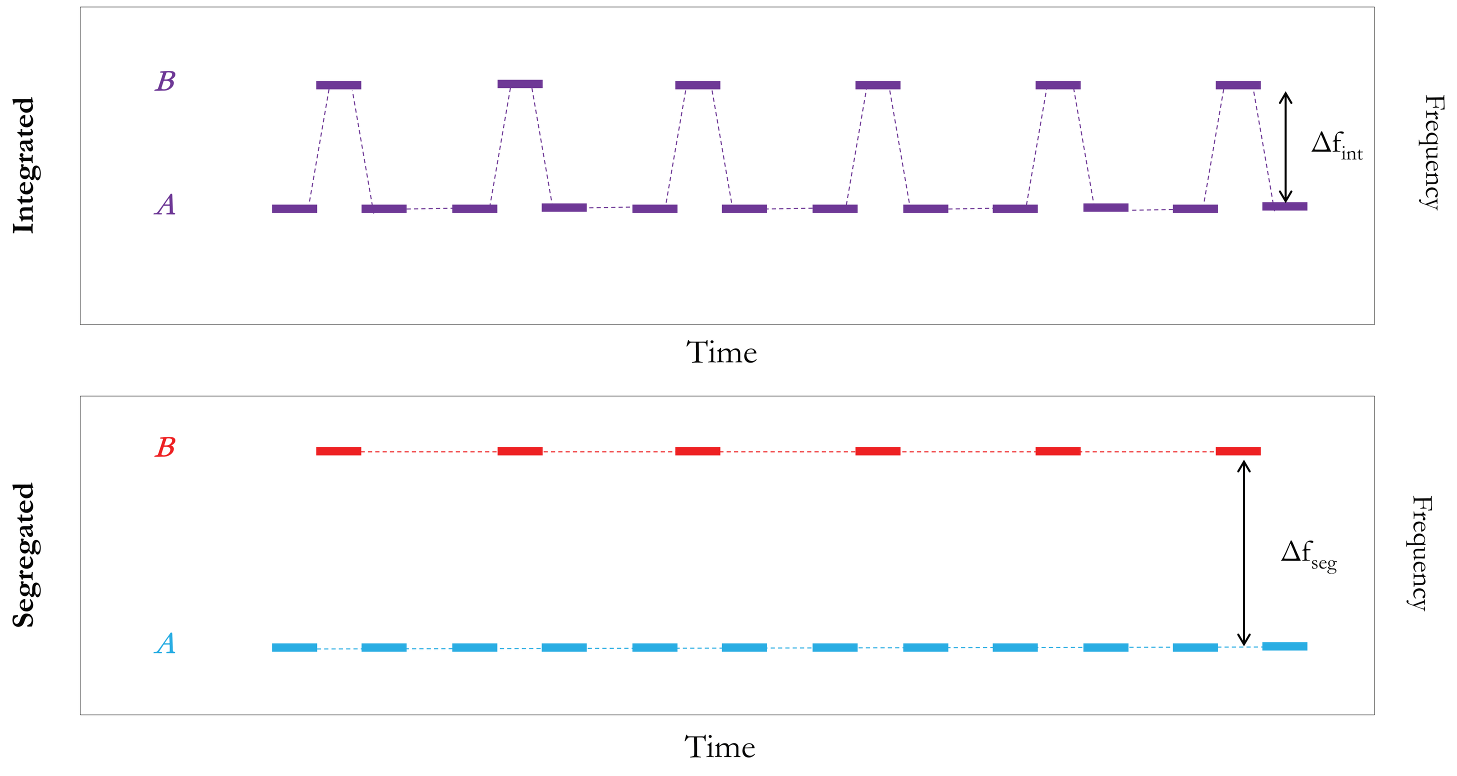
The Interactions of Spatial and Pitch Cues in Auditory Scene Analysis

Nima Zargarnezhad[1, 2], Ezgi Coskun[3], Ingrid Johnsrude[1, 2, 4]

[1]Brain and Mind, Western University; [2]Department of Psychology, Western University; [3]Undergraduate Program in Neuroscience, Schulich School of Medicine & Dentistry, Western University; [4]School of Communication Sciences and Disorders, Western University

Background

Auditory Scene Analysis (ASA) is the process of segregating and grouping sound components in a mixture of sounds to perceptually organize them into meaningful auditory streams. Different physical and perceptual properties of sound contribute to ASA. Two families of these properties are spatial and pitch cues. Sequential streaming[1] is a type of ASA that follows the perceived auditory streams over time. This type of streaming is often studied with van Noorden's ABA triplets[2, 3] also known as the "horse-morse" paradigm.



Integrated
B
A
Δf_{int}
Frequency
Time
Segregated
B
A
Δf_{seg}
Frequency
Time

Figure 1. Van Noorden's ABA paradigm[2]. ***Top.*** When the frequency difference between the A and B tones (Δf) is small the sequence is perceived as a coherent stream sounding like a galloping rhythm (Integrated percept). ***Bottom.*** As the Δf increases the sequences are perceived as two distinct streams of sounds (Segregated percept). Both percepts are accessible at intermediate Δf values[4].

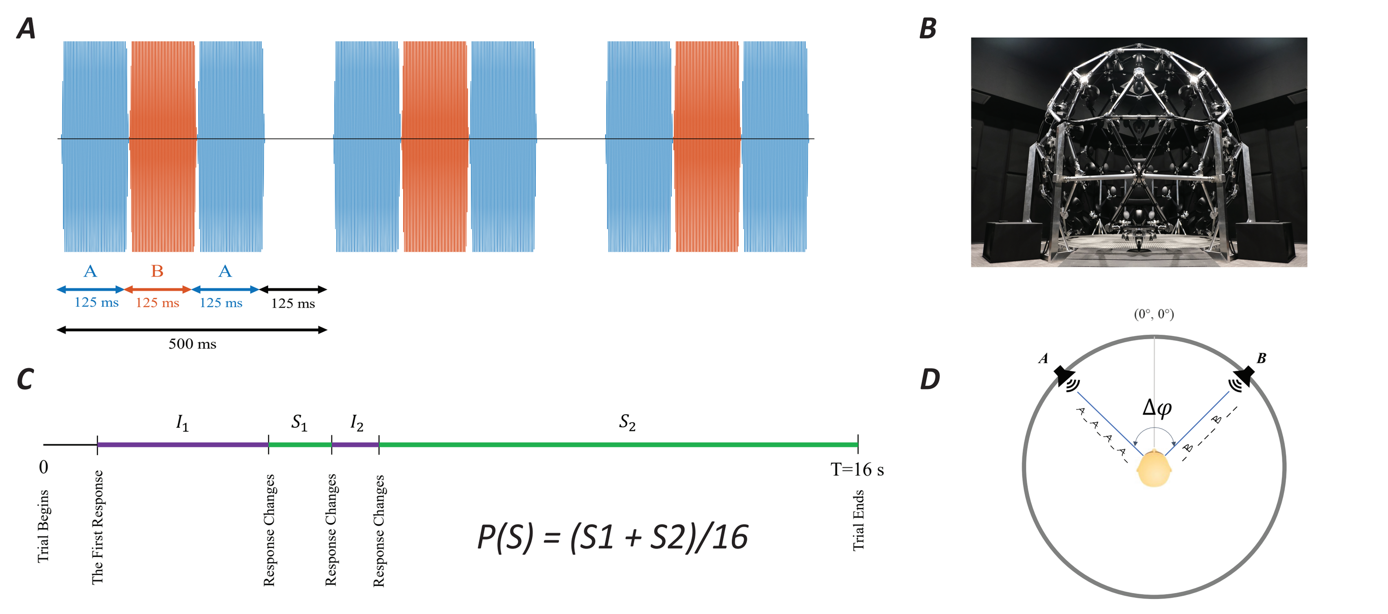
Increasing pitch difference between A and B tones in this paradigm increases the probability of perceptually segregating the two sequences[2, 3, 5]. A and B sounds are usually colocated, and when they are presented to different ears, perceptual segregation is also promoted[6].

Objectives

We aimed to characterize the effects of spatial cues in sequential streaming and the trade-off between spatial and pitch differences in perceptual organization. To accomplish this goal, A and B sources were presented from different locations on the horizontal plane (in a virtual auditory space) and the perceptual weights of spatial and pitch differences were compared.

Materials and Methods

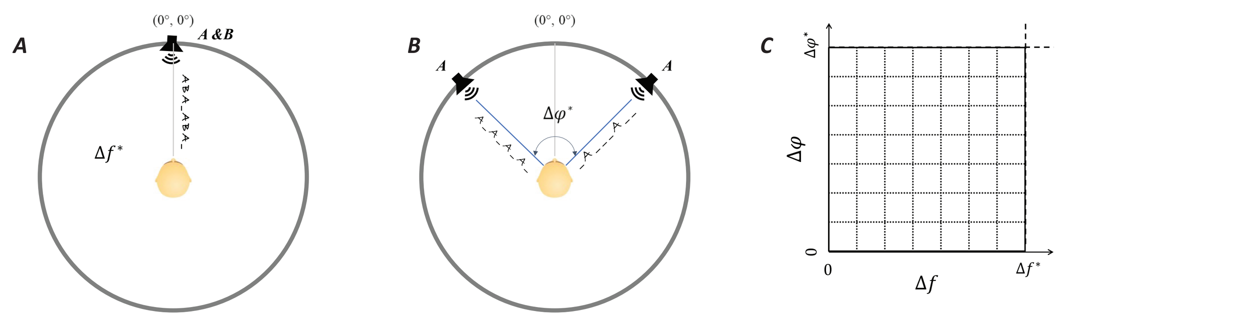
Paradigm and Apparatus



A
B
A
B
A
125 ms
125 ms
125 ms
125 ms
500 ms
C
I_1
S_1
I_2
S_2
0
T=16 s
Trial Begins
The First Response
Response Changes
Response Changes
Response Changes
Trial Ends
P(S) = (S1 + S2)/16
D
(0°, 0°)
Δφ

Figure 2. *A.* Participants listened to 32 repetitions (16 sec) of ABA triplets at every trial with f_A always set to 400 Hz and f_B set to a value from 0 to 12 semitones above f_A[7]. ***B.*** The A and B sequences were presented from virtual locations in a 9th order ambisonic loudspeaker array. *(more information about this system presented at this meeting: poster S 161)*. ***C.*** Participants reported their perception (**I**ntegrated/**S**egregated) throughout the trial via keypress[7]. They were allowed to change their response as they wished. The segregation probability *P(S)* was estimated. ***D.*** A and B sounds presented with the angular difference of Δφ on the horizontal plane with participants' head at the center.

Two Dimensional Psychophysical Space



A
(0°, 0°)
A &B
Δf^*
B
(0°, 0°)
$\Delta\varphi^*$
C
$\Delta\varphi^*$
$\Delta\varphi$
0
0
Δf
Δf^*

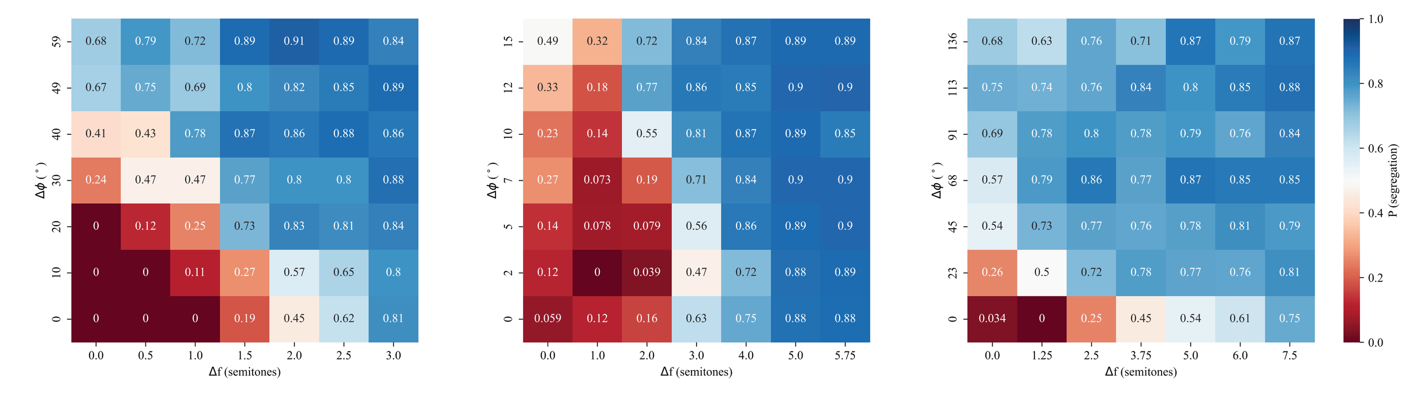
Figure 3. At the beginning of the experiment, two interleaved transformed Up-Down procedures[8] were used to estimate ***A.*** the frequency difference (Δf*) and ***B.*** the angular difference (Δφ*) values with the segregation probability of 0.84 when the other parameter was zero. ***C.*** The 7 values in the range between 0 and the estimated values (inclusive) combined to make an individual specific two-dimensional psychophysical space. Each of the 49 coordinates of this space was presented 10 times to the participants.

A Two-Dimensional Psychophysical Model

$$\underbrace{\psi}_{\text{segregation probability}}(\Delta f,\Delta\varphi)=\left(1+\exp\left(-\left(\underbrace{\beta_0}_{\text{bias term}}+\overbrace{\beta_f}^{\text{frequency difference coefficient}}(\Delta f)+\overbrace{\beta_\varphi}^{\text{angular difference coefficient}}(\Delta\varphi)+\underbrace{\beta_{f\varphi}}_{\text{multiplication (interaction) coefficient}}(\Delta f)(\Delta\varphi)\right)\right)\right)^{-1}$$

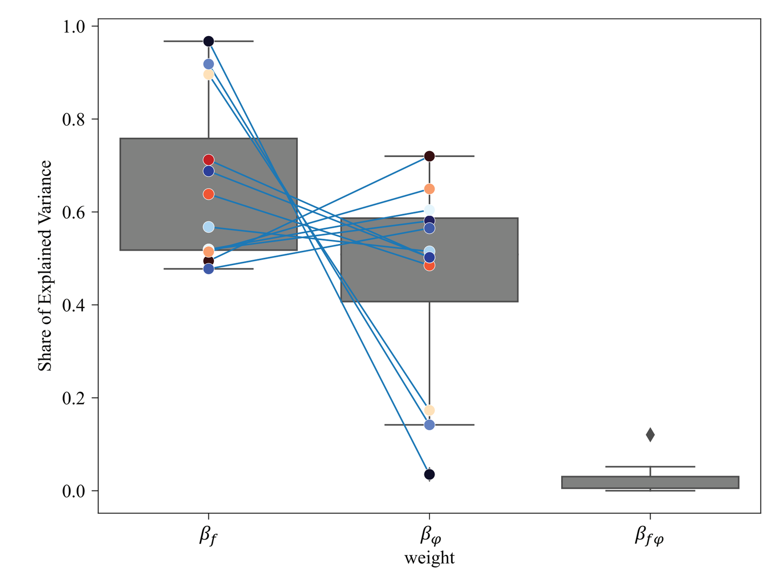
Results

Individual Variability of Coefficient Contributions to Explained Variance



Δφ (°)
Δf (semitones)
P (segregation)

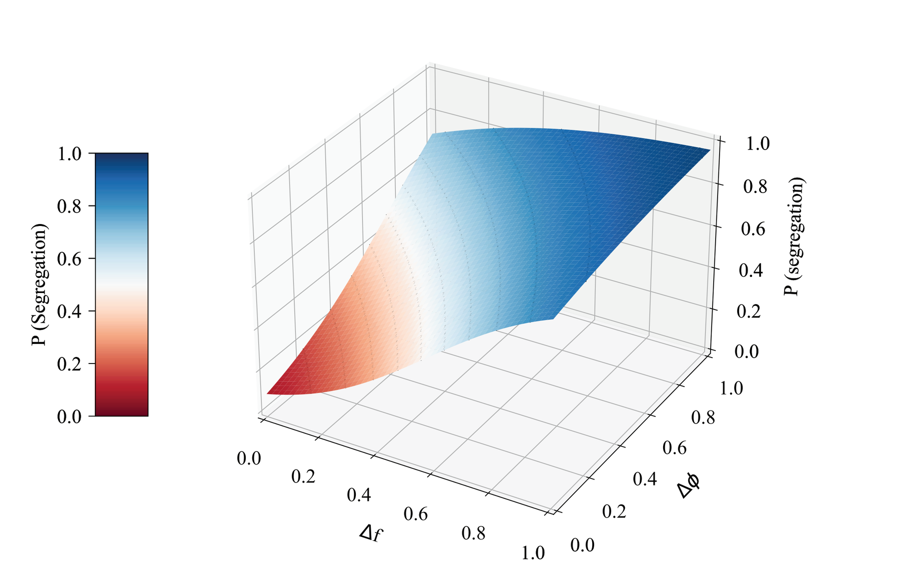
Figure 4. Average segregation probability at the two dimensional psychophysical space coordinates for three sample listeners illustrating a wide range of Δφ* and Δf* values and individual variability in perceptual weights of spatial and pitch cues.



Share of Explained Variance
β_f
β_φ
weight
$\beta_{f\varphi}$

Figure 5. For each listener, the model coefficients and the proportion of their contribution to the total explained variance were estimated (assuming an R^2-inspired measure for explained variance). Besides the individual differences, the frequency coefficient has a higher contribution to explained variance than the angular difference coefficient (*p = 0.05007*).

Frequency Differences Dominate Angular Differences in a Group Model



P (Segregation)
P (segregation)
Δf
Δφ

Coefficient	Value	Contribution to the explained variance proportion
β_0	-2.32	-
β_f	4.01	0.67
β_φ	3.15	0.40
$\beta_{f\varphi}$	-1.57	0.01

Figure 6. All participants (*n=12*) coordinate values were normalized to their Δφ* and Δf* values and the coefficients were estimated for the pooled data. The frequency difference coefficient contribution to explained variance is higher than the angular difference coefficient. The two dimensional psychometric surface is shown on the left and the coefficients and their values are listed in the table on the right.

Conclusions

Summary

- Van Noorden's ABA triplets[2] were presented to listeners with displaced A and B tone sources on the horizontal plane (symmetrical around the median plane) with a range of frequency differences.
- A two-dimensional psychometric model was introduced to quantify perceptual weights of spatial and pitch cues and their interactions.
- Both sets of cues and their interaction were found to contribute to perceptual organization.
- Despite that the average and group level perceptual weights of pitch cues (frequency differences) are higher than the perceptual weights of spatial cues, there is high variability between the participants.

Discussion and Future Steps

The wide range of 84% threshold values for segregation based on angular differences (Δφ*) and the observed individual variability in the trade-off between the spatial and pitch cues could be explained by the individual strategies and preferences when responding to this subjective paradigm. Additionally, some listeners (n = 4, excluded from the analysis) were biased to always report segregation even when there was no difference between the A and B sequences (segregation probabilities higher than 0.49 for the colocated AAA sequences i.e., the (0, 0) coordinate of the psychophysical space).

In a follow-up experiment, we will use an objective measure of perceptual organization in the ABA paradigm[9, 10] to test if the inidvidual variability is reduced.

References

[1] C. J. Darwin, "Auditory grouping," *Trends Cogn. Sci.*, vol. 1, no. 9, pp. 327–333, Dec. 1997.
[2] L. P. A. S. van Noorden, "Temporal coherence in the perception of tone sequences (Unpublished doctoral thesis)," *Eindhoven University of Technology*, 1975.
[3] A. S. Bregman, "Auditory scene analysis: The perceptual organization of sound," vol. 773, 1990.
[4] B. C. J. Moore and H. E. Gockel, "Properties of auditory stream formation," *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 367, no. 1591, pp. 919–931, Apr. 2012.
[5] R. P. Carlyon, "How the brain separates sounds," *Trends Cogn. Sci.*, vol. 8, no. 10, pp. 465–471, Oct. 2004.
[6] S. E. Boehnke and D. P. Phillips, "The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences," *Percept. Psychophys.*, vol. 67, no. 6, pp. 1088–1101, Aug. 2005.
[7] R. Cusack, J. Deeks, G. Aikman, and R. P. Carlyon, "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 30, no. 4, pp. 643–656, Aug. 2004.
[8] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, no. 2B, pp. 467–477, Feb. 1971.
[9] C. Micheyl, R. P. Carlyon, R. Cusack, and B. C. J. Moore, "Performance measures of auditory organization," *Auditory Signal Processing.* pp. 202–210, 2005.
[10] S. K. Thompson, R. P. Carlyon, and R. Cusack, "An objective measurement of the build-up of auditory streaming and of its modulation by attention," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 37, no. 4, pp. 1253–1262, Aug. 2011.

Contact

nzargarn@uwo.ca

CONCH Cognitive Neuroscience of Communication and Hearing
INNOVATION Canada Foundation for Innovation Fondation canadienne pour l'innovation
Western BrainsCAN Transforming brain research.
NSERC CRSNG