

Variable Properties of Auditory Image Analysis: A Case Study of Selected Musical Works

Adam Rosiński (a.rosinski@amfn.pl)

Faculty of Composition, Theory of Music and Sound Engineering, The Feliks Nowowiejski Academy of Music in Bydgoszcz, Juliusza Słowackiego 7 Street, 85-008 Bydgoszcz, Poland

Abstract

The aim of this study is to analyse the auditory scene of musical works and to demonstrate that different compositions may prompt the emergence of distinctly interpreted perceptual streams in the listener's mind. The research focuses on selected excerpts from works by Alexandre Guilmant, Ludwig van Beethoven, and Antonio Vivaldi, which, due to their unique characteristics, elicit diverse auditory impressions. By combining score analysis with auditory scene analysis, this paper seeks to explain how different interpretations of the same sounds result in dissimilar auditory impressions. The auditory scene analysis presented here provides deeper insight into the process of stream formation and its implications for musical performance and aesthetic perception. The findings indicate that perceptual stream formation in music is considerably more complex and context-dependent than previously assumed, with implications for how listeners interpret auditory scenes.

Keywords: perception; music; analysis; art and cognition; auditory scene analysis

Introduction

Combining different fields of knowledge such as psychology, neurology, linguistics and philosophy of mind, contemporary cognitive science offers a unique approach to the analysis of perceptual processes. One of the fundamental issues studied by cognitive scholars is how the human mind organises sensory input to form coherent impressions. With regard to auditory perception, auditory scene analysis plays a paramount role as it enables one to separate and integrate sounds into diverse perceptual streams. This ability is crucial for understanding the mechanisms behind auditory information processing and elaborating on theories that explain the cognitive processes of music perception.

In the cognitivist approach to auditory scene analysis, one explores how attentional and memory processes affect how sounds are processed. Neurocognitive research shows that both lower-level structures (such as the brainstem and auditory cortex) and higher cortical areas work together to enable the interpretation of acoustic stimuli, e.g. in the form of perceptual streams. Understanding how the brain creates coherent auditory images from fragmented information is crucial for developing theories of perception and cognition. Analysing these complex mechanisms helps clarify how the

mind organises fragmented acoustic input into coherent structures — or, conversely, how it disassembles such structures into discrete sound elements, as in musical analysis. Thus, research and analyses in this field contribute to the overall knowledge of the nature of human cognition, elucidating how we hear, process and interpret sounds and determining what affects the perception of music and other sound forms.

This study follows a qualitative case study framework rooted in expert listening and introspective analysis. Although not based on experimental group data, it draws on the author's prior psychoacoustic research and employs detailed analytic procedures in the spirit of phenomenological and cognitive interpretive methods. Such approaches are increasingly acknowledged in cognitive science when investigating complex, subjective phenomena such as music perception (see e.g. Jensen & Mathewson, 2011). This approach aligns with recent defences of introspective case studies in cognitive science (see also: Overgaard, 2021).

Music perception is an exceedingly complex process that involves both the essential senses and more advanced cognitive functions, enabling the interpretation of elaborate sound structures (Hausfeld, Disbergen & Valente, 2021). The perception of music is not just passive listening to sounds but requires the mind to be actively involved in recognising patterns, rhythms, melodies and harmonies, simultaneously identifying the emotions that the music evokes (Szalárdy, Bendixen, Böhm, Davies, Denham & Winkler, 2014; Timmers, Arthurs & Crook, 2020). Perceptual streams represent mental groupings of sound sequences, enabling listeners to separate auditory sources and organise them into coherent, structured wholes (Moore & Gockel, 2012; Snyder, Lee, Carter, Hannon & Alain, 2008). Consequently, it is possible to apprehend that various musical elements (such as parts of different instruments) are distinct components of one larger whole rather than random sets of sounds.

Listening Equipment Overview, Procedure, Sessions, and Audio Material

First The equipment employed during the evaluation sessions consisted of:

headphones: *Beyerdynamic DT 770Pro*, 80 ohms; audio interface: *Steinberg UR824*; sample rate: 44.1/16bit; peak loudness: 95 phones; laptop: *HP Pavilion Aero 13*.

All musical works analysed in this study were personally reviewed by the author under strictly controlled acoustic conditions, minimising disturbances caused by playback quality or room acoustics. To ensure consistency, the material was played using headphones with a neutral sound profile in an acoustically isolated environment. This method eliminated potential interference, enhancing the reliability of observations and conclusions. The focus was placed on music with complex textures and rich sonic layers, which cannot be fully appreciated in laboratory settings that employ test tones exclusively.

Given the wide range of musical, non-musical, and psychological influences on how listeners interpret sound into perceptual streams, the author sought to reduce variables that could interfere with stream formation, aiming to ensure the most accurate auditory image analysis possible (see also: Hermes, 2023).

Due to the inherent variability in musicians' interpretations and the distinct qualities of recordings, excerpts from several different recordings were chosen. This step was critical to standardise and equalise the audio content on a sonic level, both technically and with respect to:

- the varying acoustic properties of different concert venues (see also: Griesinger, 2011, 2015; Haapaniemi, 2018);
- the distinct tonal characteristics of individual instruments (Fischer, Soden, Thoret, Montegry & McAdams, 2021; Fischer & McAdams, 2024; Marozeau, Innes-Brown & Blamey, 2013);
- performance and interpretational differences in the musical pieces (Rosiński, 2024a).

Listening sessions were conducted in October 2024, with each session lasting no longer than two hours. To ensure the listener's concentration remained high and minimise any potential perceptual errors caused by fatigue, only one session per day was held (Jain & Nataraja, 2022; Jain, Nataraja & Narne, 2022).

A key methodological aspect of this study was a comprehensive analysis of musical works, combining investigations of recordings and their scores. The recordings provided insights into the actual acoustic realisation of the pieces, encompassing timbre, dynamics, and expressivity, while the scores offered structural and formal interpretative frameworks that complemented the listening experience. This two-tiered approach revealed the multidimensional nature of musical perception, illustrating how identical acoustic stimuli can be interpreted differently depending on prior knowledge, cultural context, and situational factors such as mood or listening focus.

This analysis draws on the author's extensive expertise in auditory stream analysis, both as a music theorist and a practitioner with advanced analytical skills in auditory perception. The decision not to replicate this experience with respondents stems from the author's prior psychoacoustic experiments conducted on groups of musicians and non-

musicians, which yielded extensive data on how these groups integrate and segregate sounds. These findings have provided a clearer understanding of the perceptual mechanisms influencing the formation of mental figures. It was thus determined that further similar studies would not significantly contribute new insights, as the existing data are sufficient to formulate credible and detailed conclusions (Rosiński, 2021; 2023; 2024a; 2024b; 2024c; 2024d; 2025).

The selected compositions were chosen due to their high degree of contrapuntal complexity, overlapping timbral textures, and layered temporal structures. These musical characteristics are known to elicit perceptual ambiguity and promote the emergence of multiple simultaneous auditory streams (cf. Bregman, 1990; McAdams, Goodchild & Soden, 2022). The following compositions were chosen for the auditory scene analysis, specifically in the context of the emergence of perceptual streams:

- Alexandre Guilmant, *Organ sonata: Finale*, op. 42, no. 1, *Allegro assai*, m. 6–8, 93–95, 105–110, performed by: Olivier Vernet, from the record: *Alexandre Guilmant – Sonate no 1 opus 42, Sonate no 5, opus 80*, 03700, record label: Ligia Digital, 2013;
- Ludwig van Beethoven, *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 14–15, 69–71, 180–181, performed by: Mikhail Pletnev, from the record: *Beethoven: Piano Sonatas, 14 'Moonlight', 21 'Waldstein' & 23 'Appassionata'*, record label: EMI Records/Virgin Classics, 2006;
- Antonio Vivaldi, *Concerto for two violins and string orchestra*, op. 8, no. 3, RV522, *Allegro*, m. 55–61, performed by Takashi Baba and I Solisti Italiani, from the record: *Vivaldi 5 Concerti – I Solisti Italiani*, 33C37-7401, record label: Denon/Nippon Columbia Japan, 1985.

Own research – sheet music analysis, proving the occurrence of multivariantism of auditory perceptions

Figure 1 shows a three-measure long excerpt from the *Organ Sonata: Finale*, Op. 42, no. 1, *Allegro assai* by Alexandre Guilmant. The written sounds, through various classifications made by the perceptor, will be arranged into different mental figures (Thurman, 2010):

- Variant No. 1: sounds performed in the part of the right-hand form the first perceptual stream, while the ones in the left hand – the second (Shin, 2012);
- Variant No. 2: Through the interval proximity of sounds written in a similar pitch scale, the interpretation of a given stream is not uniform. Depending on the focus of the listener's attention, the perceptor may direct his interpretation of the stimuli to melodic (left-hand part) and/or harmonic (right-hand part) elements. In this case, the perceptor may also perceive the melody and the harmony interchangeably, depending on how he interprets the incoming sounds;
- Variant No. 3: each highest sound from the harmonic vertical played by the instrumentalist with his right hand is

treated by the listener as the melody and assigned to the first stream (Benderius, 2015). The second stream is formed by the remaining sounds from the triads (making them dyads), which are interpreted as belonging to the harmony. The third stream is built on the basis of the sounds performed with the instrumentalist's left hand;

- Variant No. 4: this variant is identical to variant No. 3 with one difference – the notes performed with the left hand split into two streams. One of them is formed from the repeated E4 notes (every other sound), while the other consists of all the remaining notes performed by the instrumentalist with his left hand.



Figure 1: Alexandre Guilmant, *Organ sonata: Finale*, op. 42, no. 1, *Allegro assai*, m. 6–8 and 93–95 (both fragments are identical).

Source:

<https://imslp.org/wiki/Special:ImagefromIndex/290298/cy28> [accessed on 20 January 2025].

Figure 2 presents another excerpt from the same movement of Alexandre Guilmant's piece, measures 105 to 110. Different perspectives on the incoming sounds may result in a wide range of mental representations:

- Variant No. 1: All sounds performed in the right-hand part are treated as belonging to the first stream. In this case, the listener hears a descending arpeggio. The sounds written in the part of the left hand form the second stream, and those in the pedal part – the third one;
- Variant No. 2: The proximity of the pitch scales of sounds written in the left- and right-hand parts in measures 105 and 106 makes the sounds merge into one integral perceptual stream. Subsequent measures, as well as the sounds of the pedal, are interpreted according to the description of variant No. 1;
- Variant No. 3: The highest notes in the right-hand part (quarter notes) form the first – melodic perceptual stream. The second stream is composed of the sixteenth notes written in the part of the right hand and all the notes in the part of the left hand. The third and final perceptual stream is formed from the notes performed in the pedal part;
- Variant No. 4: As in variant No. 3, the highest notes are performed in the part of the right hand (quarter notes), which form the first melodic perceptual stream. The second stream is built from the sixteenth notes written in the part of the right hand. Mainly from measures 107 to 110, sounds performed in the part of the left hand, through a significantly increased pitch scale compared to the sounds written in the part of the right hand, are segregated and become the basis for the formation of the third stream. The

sounds performed by the instrumentalist in the pedal form the last, fourth perceptual stream;

- Variant No. 5: This variant can emerge as a result of the listeners using different perceptual segregation processes in individual measures. The structure of the mentioned stream is not homogeneous, and it can be based on the combination of all four previously described variants, as well as the creation of new sub-variants. For example, in measures 109 and 110, the quarter note passage performed by the instrumentalist with his left hand can be integrated into a single perceptual stream with the sounds of the pedal since the applied rhythm is identical and the closeness of the tones to each other in the pitch scale causes the occurrence of the phenomenon of masking (because of the same attack time of those sounds) (Wang & Sogin, 1990).



Figure 2: Alexandre Guilmant, *Organ sonata: Finale*, op. 42, no. 1, *Allegro assai*, m. 105–110.

Source:

<https://imslp.org/wiki/Special:ImagefromIndex/290298/cy28> [accessed on 20 January 2025].

The fragment of Ludvig van Beethoven's piece *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 14–15 and 180–181, presented in Figures 3 and 4, is very interesting since the presented measures may be interpreted by the listener in a perceptually non-uniform way:

- Variant No. 1: Sounds played in the part of the right hand form the first perceptual stream, while those played in the left hand build the second one;
- Variant No. 2: Sounds are combined into harmonic pairs. One of the streams consists of the lower sounds written in the part of the right hand (C4 and D4 – Figure 3, or D4 and E4 – Figure 4) and the higher sounds performed in the part of the instrumentalist's left hand (G3 and A3 – Figure 3, or A3 and B3 – Figure 4). The second perceptual stream is formed by higher sounds performed in the part of the musician's right hand (E4 and F#4 – Figure 3, or F4 and G#4 – Figure 4) together with lower sounds written in the part of the left hand (C3 – Figure 3, or D3 – Figure 4). This way of combining sounds by the perceptor causes the impression of hearing two intertwined streams, thanks to which the listener alternately perceives two harmonic positions: closed and open;
- Variant No. 3: A repeated occurrence of the same or very similar sounds, which can lead to segregation and enable the listener to perceive four separate and independent streams. The first stream is built from the sounds C4 and D4 – Figure 3, or D4 and E4 – Figure 4, the second stream includes the notes E4 and F#4 – Figure 3, or F4 and G#4 – Figure 4, the third stream consists of the repeated tones G3

and A3 – Figure 3, or A3 and B3 – Figure 4, while the last stream, the fourth one, is formed solely from the note C3 – Figure 3, or D3 – Figure 4.



Figure 3: Ludwig van Beethoven, *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 14–15.

Source:

<https://imslp.org/wiki/Special:ImagefromIndex/51748/cy28> [accessed on 28 January 2025].



Figure 4: Ludwig van Beethoven, *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 180–181.

Source:

<https://imslp.org/wiki/Special:ImagefromIndex/51748/cy28> [accessed on 28 January 2025].

In a further part of *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 69–71, it is possible to observe the musical material that is very similar to the one presented earlier. However, the use of different phrases in the right and left-hand parts causes the images obtained as a result of the perceptual analysis to be interpreted in multiple different ways (Huron, 2001):

- Variant No. 1: The sounds performed in the part of the right hand will be the basis for the creation of the first perceptual stream, while the notes of the left-hand part will form the second one (Goebel, 2003) – which is in accordance with the score.
- Variant No. 2: The highest notes written in the right-hand part form the first perceptual stream (every other note in measure 69 and every first note of each sixteenth note group in measures 70 and 71 – excluding the first group, as there is a sixteenth note pause written). The other (lower) notes of the right-hand part are the basis for the formation of the second stream (every other lower note in measure 69 and three subsequent notes omitting the first, highest note of each sixteenth note group in measures 70 and 71). The third stream consists of the notes performed in the instrumentalist’s left-hand part.
- Variant No. 3: The sounds written in the right-hand part are interpreted following variant 1 or 2 described earlier. The sounds written in the left-hand part, because of being performed in a staccato articulation technique, make it possible to create two additional perceptual streams (McAdams, Depalle & Clarke, 2004), where the higher sounds played with the left hand will build the second or third perceptual stream, while the lower sounds will form

the third or fourth perceptual stream (depending on which perceptual variant the listener will follow when analysing the sounds written in the right-hand part);

- Variant No. 4: Occurs when the listener strongly focuses his attention on the higher register sounds. The sounds played with the right hand can mask those played with the left hand (Bregman, 2002). The described way of perceiving sounds incoming to the listener is caused by the note transients occurring at the same time in both left and right-hand parts, as well as by the staccato articulation technique (Kallman & Massaro, 1979; Massaro, 1975; Rakowski, 1983), which results in a much shorter sustain of the sounds written in the part of the left hand. Thus, the sounds performed with the right hand may be more audible.



Figure 5: Ludwig van Beethoven, *Piano sonata C Major (Waldstein)*, op. 53, no. 21, *Allegro con brio*, m. 69–71.

Source:

<https://imslp.org/wiki/Special:ImagefromIndex/51748/cy28> [accessed on 28 January 2025].

Figures 6 and 7 show an excerpt from Antonio Vivaldi’s piece, *Concerto for two violins and string orchestra*, Op. 8, no. 3, RV522, *Allegro*, m. 55 to 61. This piece also can be an example of music that can be identified by the perceptor in different ways, creating various auditory images using perceptual analysis (McAdams, Goodchild & Soden, 2022):

- Variant No. 1: the violin I part creates the first perceptual stream, the violin II part builds the second stream, the other violins and violas are responsible for creating the third perceptual stream, while the single quarter note written in the cello and violone e cembalo parts results in a sound that cannot be considered a fully-fledged stream, as it appears and after a moment fades away;
- Variant No. 2: violin I, through constant repetition of the same tones (the highest notes written in a given measure), forms the first perceptual stream, while the remaining lower notes of the same instrument create the second stream (Bosi, 2003; Massaro & Burke, 1991; Massaro & Idson, 1977). The third perceptual stream is formed based on the melodic line written in the part of violin II. The fourth stream is built by the remaining instrumental parts or any very brief sounds that are difficult to consider in the context of forming the perceptual stream (the sounds of cello and violone e cembalo);
- Variant No. 3: this variant is very similar to variant No. 2. However, in this case, the part of violin II is masked by the part of violin I (Nishihara & Hidaka, 2012). By focusing on the segregation of sounds occurring in the part of violin I, the perceptor overlooks the sounds of violin II and does not notice them (Foyle & Watson, 1984; Kallman, Hirtle & Davidson, 1986; Massaro & Idson, 1978). The remaining

instrument parts are interpreted in the same manner as described for variant No. 2.



Figure 6: Antonio Vivaldi, *Concerto for 2 violins and string orchestra*, op. 8, no. 3, RV522, *Allegro*, m. 55–57.

Source:

https://vmirror.imslp.org/files/imglnks/usimg/4/4d/IMSLP572176-PMLP126413-L_Estro_Armonico_--_conc_8.pdf [accessed on 03 January 2025].



Figure 7: Antonio Vivaldi, *Concerto for two violins and string orchestra*, op. 8, no. 3, RV522, *Allegro*, m. 58–61.

Source:

https://vmirror.imslp.org/files/imglnks/usimg/4/4d/IMSLP572176-PMLP126413-L_Estro_Armonico_--_conc_8.pdf [accessed on 03 January 2025].

Discussion

The results of analyses concerned with music perception and the emergence of perceptual streams confirm that this process is much more complex than it has been asserted in classical psychoacoustic research. Conclusions from the author's own experiments and analyses prompt new questions and suggest directions for future research, which may be divided according to various criteria:

- the individual nature of music perception—music perception is a subjective experience, dependent on many factors, including the individual experience of the listener (Dibben, Coutinho & Vilar, 2018; Lynch, Eilers & Oller, 1990), their emotions, level of attention (Shamma, Elhilali & Micheyl, 2011; Cusack, Deeks, Aikman & Carlyon, 2004) and psychophysical state. Consequently, the same

piece of music can lead to distinct listening experiences in different people, even if exposure takes place in the same acoustic conditions. This raises the question of whether and how it is possible to create musical works that will be received in a similar fashion by a broad audience. Understanding individual differences in music reception may also contribute to more in-depth investigations into how cultural context, life experiences or familiarity with musical styles affect the perception of sounds;

- the role of performer variability in perception—the impact of a performer's interpretation of music perception remains a matter of debate. The analyses conducted to date show that different delivery does indeed give rise to different perceptual streams (Heng & McAdams, 2024; McAdams, 1987). Nevertheless, the degree to which a performer's interpretation can influence the listener's final perception of a piece is yet to be established;
- perceptual variability and listener fatigue—an important factor that applies to every listener is fatigue. It follows from laboratory studies that fatigue affects the quality of perception and stimulus processing (Jain & Nataraja, 2022; Jain, Nataraja & Narne, 2022). Thus, another question that one should ask is whether fatigue impacts perceptual variability to the same degree in different people. Are there individuals who are more or less susceptible to perceptual variation when listening to musical material in a fatigued state?

Certain limitations of the conducted analyses are also worth noting: although the study yielded important information on the influence of acoustic variables on the perception of auditory images, it did not allow for the diverse cultural backgrounds of people from different parts of the world, which may significantly affect the interpretation of the results of sound segregation and integration (Stevens, 2012; Wang, Wei, Heng & McAdams, 2021).

Beyond fatigue, individual differences play a critical role in the perceptual structuring of sound. The author's prior experimental research (anonymised) has shown that trained musicians tend to prioritise hierarchical tonal and rhythmic structures, while non-musicians are more likely to rely on surface features such as timbre and rhythmic salience. These tendencies result in diverging stream segmentation even under controlled conditions.

Additionally, research into cross-cultural music perception (Stevens, 2012; Wang, Wei, Heng & McAdams, 2021) highlights that enculturation affects how listeners organise sound sequences, particularly in terms of timbre salience, pitch hierarchy, and tuning systems. The variability documented in this study should therefore be interpreted in light of both cultural background and individual listening strategies.

Conclusion

The analyses conducted in the current study demonstrate that the formation of perceptual streams in music is a process contingent on many variables, such as tonal distance, articulation, rhythm and tempo. These particular elements

shape how listeners segregate sounds and organise them into coherent perceptual images. It may be noted that the presented findings have some relevant, practical implications for music professionals from a variety of fields:

– for performers — the findings have direct implications for musical performance. Musicians may consciously guide the perception of their listeners by manipulating elements such as rhythm, dynamics, or articulation. Understanding perceptual mechanisms enables them to better anticipate how their interpretations might be perceived by different audience groups.

– for music theorists — research into perceptual streams opens up new possibilities, offering a clearer picture of the correlation between compositional structure and listener reception. Theorists can analyse musical works not only in terms of formal constructs such as harmony or counterpoint, but also from the perspective of sound perception. Taking this dimension into account may lead to more nuanced and insightful analyses, capable of explaining complex phenomena such as multi-level listening and the perceptual segregation or integration of sound layers. This, in turn, can contribute significantly to the development of music perception theory.

– for composers — the discovery of new perspectives and creative possibilities can be of considerable importance, as it enables the deliberate construction of works that reflect the perceptual tendencies of listeners. Knowledge of perceptual processes allows composers to design multidimensional pieces in which individual sound layers are shaped to achieve specific perceptual outcomes (Bregman, 1990). Moreover, composers may experiment with techniques that influence which streams are likely to dominate the listener's experience.

– for sound engineers — the obtained findings may prove particularly valuable when creating the sound layer of recordings. Understanding how recorded music interacts with the listeners' perception of sounds may aid key decisions about mixing, mastering and production. Sound engineers can manipulate acoustic space, dynamic contrasts and instrument placement in the mix to highlight specific perceptual streams and direct the listener's attention to selected elements of the composition (Kaya & Elhilali, 2017).

While the present findings offer potential insights for music theorists and composers, the broader implications for performers and sound engineers remain more speculative. These groups often engage with music through different perceptual strategies and professional priorities. The current study, based on expert introspection and theoretical analysis, provides heuristic suggestions rather than empirically validated recommendations.

Nevertheless, this approach may inspire further research that systematically explores how compositional structures, acoustic conditions, and interpretive choices shape the formation of perceptual streams in various listener populations. The qualitative richness of the presented analysis lays a foundation for interdisciplinary inquiry,

integrating cognitive models of perception with real-world artistic practices.

The findings of this study challenge the predictive power of canonical auditory scene analysis models that rely primarily on bottom-up cues such as pitch separation and temporal proximity (Bregman, 1990; Moore & Gockel, 2012). In contrast, the present analysis reveals the decisive influence of top-down mechanisms, including attention modulation, stylistic familiarity, and structural expectations.

The observed variants also conceptually align with structural grouping principles proposed in the Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1985). However, while GTTM posits a rule-based parsing of tonal material, this study demonstrates the co-existence of multiple plausible auditory organisations arising from a single notated source — depending on attentional focus, interpretative mode, or listener intention.

Building on these findings, research into music perception and the creation of perceptual streams sheds new light on the complex processes of sound processing. The analysis of scores and the application of psychoacoustic methods provide a better understanding of how various factors—the performer's interpretation, the individual experience of the listeners and the acoustic conditions—affect the final aesthetic experience when listening to music (Hochgraf, 2019). Given the complexity and multifaceted nature of the music perception process, further research could focus on how a variety of listening conditions—including concert hall acoustics, the quality of sound equipment and the listener's environment—affect the formation of perceptual streams (Lokki, 2016; Nishihara & Hidaka, 2012; Pressnitzer & McAdams, 2000). The studies of various genres, especially contemporary and experimental music, may bring new insights regarding the influence of innovative-sounding sound structures on listeners' perception (Song, Ellermeier & Hald, 2011), and explain the relationship between sounds and the mind, leading to an even better understanding of the mechanisms of music perception.

Music perception analysis demonstrates once again that the human mind actively processes sounds, organising the perceptual streams that emerge and recognising patterns, which is crucial for understanding the principles of perception and the cognitive integration of information derived from the auditory analysis of musical material. For this reason, research on music and sound perception makes a valuable contribution to the broader understanding of cognitive mechanisms that underlie the perception of complex sensory information.

References

- Benderius, L. (2015). *The Future Role of the French Harmonium*. Master of Fine Arts Thesis, University of Gothenburg, Gothenburg. https://gupea.ub.gu.se/bitstream/handle/2077/39517/gupea_2077_39517_1.pdf?sequence=1 [accessed on 26 September 2024].

- Bosi, M. (2003). *Audio coding: Basic Principles and Recent Developments*. “6th HC International Conference”, Aizu. https://cgi.cse.unsw.edu.au/~cs4121/MP3_1.pdf [accessed on 28 September 2024].
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. The MIT Press. <https://doi.org/10.7551/mitpress/1486.001.0001>
- Bregman, A. S. (2002). The auditory scene. In D. J. Levitin, (Ed.), *Foundations of cognitive psychology: Core readings*.
- Bregman, A. S. (1987). The meaning of duplex perception: sounds as transparent objects. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception*. https://doi.org/10.1007/978-94-009-3629-4_6.
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 643–656. <https://doi.org/10.1037/0096-1523.30.4.643>
- Dibben, N., Coutinho, E., & Vilar, J. A. (2018). Do individual differences influence moment-by-moment reports of emotion perceived in music and speech prosody? *Frontiers in Behavioral Neuroscience*, 12, 1–11, Article 184. <https://doi.org/10.3389/fnbeh.2018.00184>
- Fischer, M., & McAdams, S. (2024). Instrument Timbre Combinations Influence the Relative Prominence of Perceptual Layers in Orchestral Music. *Music Perception* – preprint. <https://osf.io/preprints/osf/pkcmd> [accessed on 29 September 2024].
- Fischer, M., Soden, K., Thoret, E., Montegry, M., & McAdams, S. (2021). Instrument Timbre Enhances Perceptual Segregation in Orchestral Music. *Music Perception*, 38(5), 473–498. <https://doi.org/10.1525/M.P.2021.38.5.473>
- Foyle, D. C., & Watson, C. S. (1984). Stimulus-based versus performance-based measurement of auditory backward recognition masking. *Perception & Psychophysics*, 36(6), 515–521. <https://doi.org/10.3758/bf03207511>
- Goebel, W. (2003). *The Role of Timing and Intensity in the Production and Perception of Melody in Expressive Piano Performance*. PhD Thesis. Dissertation for the attainment of the Doctorate in Philosophy at the Faculty of Humanities of the Karl-Franzens-University Graz, <https://unipub.uni-graz.at/obvugroa/content/titleinfo/5368419/full.pdf> [accessed on 27 September 2024].
- Griesinger, D. (2015). Acoustic quality, proximity, and localization in concert halls: The role of harmonic phase alignment. *Psychomusicology: Music, Mind, and Brain*, 25(3), 339–344. <https://doi.org/10.1037/pmu0000116>
- Griesinger, D. (2011). Clarity, Cocktails, and Concerts: Listening in Concert Halls. *Acoustics Today*, 7(2), 15–22.
- Haapaniemi, A. (2018). *Perceptual studies in concert hall acoustics*. Helsinki: Unigrafia Oy.
- Hausfeld, L., Disbergen, N. R., & Valente, G. (2021). Modulating cortical instrument representations during auditory stream segregation and integration with polyphonic music. *Frontiers in Neuroscience*, 15, 1–13, Article 635937. <https://doi.org/10.3389/fnins.2021.635937>
- Heng, L., & McAdams, S. (2024). The function of timbre in the perception of affective intentions: Effect of enculturation in different musical traditions. *Musicae Scientiae*, 28(4), 1–25. <https://doi.org/10.1177/10298649241237775>
- Hermes, D. J. (2023). Auditory-Stream Formation. In *The Perceptual Structure of Sound. Current Research in Systematic Musicology*, Vol. 11. Cham: Springer. https://doi.org/10.1007/978-3-031-25566-3_10
- Hochgraf, K. A. (2019). The Art of Concert Hall Acoustics: Current Trends and Questions in Research and Design. *Acoustics Today*, 15(1), 26–34. <https://doi.org/10.1121/at.2019.15.1.31>
- Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception*, 19(1), 1–56. <https://doi.org/10.1525/mp.2001.19.1.1>
- Jain, S., Nataraja, N. P., & Narne, V. K. (2022). The Effect of Subjective Fatigue on Auditory Processing in Musicians and Non-musicians. *Music Perception*, 39(3), 309–319. <https://doi.org/10.1525/mp.2022.39.3.309>
- Jain, S., & Nataraja, N. P. (2022). The Effect of Fatigue on Working Memory and Auditory Perceptual Abilities in Trained Musicians. *American Journal of Audiology*, 28(2S), 484–492. https://doi.org/10.1044/2019_AJA-IND50-18-0102
- Jensen, M. S., & Mathewson, K. E. (2011). Simultaneous perception of both interpretations of ambiguous figures. *Perception*, 40(8), 1009–1011. <https://doi.org/10.1068/p6880>
- Kallman, H. J., & Massaro, D. W. (1979). Similarity effects in backward recognition masking. *Journal of Experimental Psychology: Human Perception & Performance*, 5(1), 110–128. <https://doi.org/10.1037//0096-1523.5.1.110>
- Kallman, H. J., Hirtle, S. C., & Davidson, D. (1986). Recognition masking of auditory duration. *Perception & Psychophysics*, 40(1), 45–51. <https://doi.org/10.3758/bf03207593>
- Kaya, E. M., & Elhilali, M. (2017). Modelling auditory attention. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1714), 1–6, Article 20160101. <https://doi.org/10.1098/rstb.2016.0101>
- Lokki, T. (2016). Why is it so hard to design a concert hall with excellent acoustics? *Proceedings of the Australian Acoustical Society*. Brisbane, Australia. https://www.acoustics.asn.au/conference_proceedings/AA_SNZ2016/papers/p43.pdf
- Lerdahl, F., & Jackendoff, R. S. (1985). *A Generative Theory of Tonal Music*. The MIT Press.
- Lynch, M. P., Eilers, R. E., & Oller, D. K. (1990). Innateness, Experience, and Music Perception. *Psychological Science*, 1(4), 272–276. <https://doi.org/10.1111/j.1467-9280.1990.tb00213.x>
- Marozeau, J., Innes-Brown, H., & Blamey, P. J. (2013). The effect of timbre and loudness on melody segregation.

- Music Perception*, 30(3), 259–274. <https://doi.org/10.1525/mp.2012.30.3.259>
- Massaro, D. W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America*, 58(5), 1059–1065. <https://doi.org/10.1121/1.380765>
- Massaro, D. W. & Burke, D. (1991). Perceptual development and auditory backward recognition masking. *Developmental Psychology*, 27(1), 85–96. <https://doi.org/10.1037/0012-1649.27.1.85>
- Massaro, D. W., & Idson, W. L. (1977). Backward recognition masking in relative pitch judgments. *Perceptual and Motor Skills*, 45(1), 87–97. <https://doi.org/10.2466/pms.1977.45.1.87>
- Massaro, D. W., & Idson, W. L. (1978). Target-mask similarity in backward recognition masking of perceived tone duration. *Perception & Psychophysics*, 24(3), 225–234. <https://doi.org/10.3758/BF03206093>
- McAdams, S. (1987). Music: A Science of the Mind? *Contemporary Music Review*, 2(1), 51–56.
- McAdams, S., Depalle, P., & Clarke, E. (2004). Analyzing Musical Sound. In E. Clarke & N. Cook (Eds), *Empirical Musicology: Aims, Methods, Prospects*. *Empirical Musicology: Aims, Methods, Prospects*. <https://doi.org/10.1093/acprof:oso/9780195167498.003.0008>
- McAdams, S., Goodchild, M., & Soden, K. (2022). A Taxonomy of Orchestral Grouping Effects Derived from Principles of Auditory Perception. *Music Theory Online*, 28(3), 1–17. <https://doi.org/10.30535/mto.28.3.6>
- Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 919–931. <https://doi.org/10.1098/rstb.2011.0355>
- Nishihara, N., & Hidaka, T. (2012). Loudness perception of low tones undergoing partial masking by higher tones in orchestral music in concert halls. *Journal of the Acoustical Society of America*, 132(2), 799–803. <https://doi.org/10.1121/1.4729547>
- Overgaard, M. (2021). Introspection in cognitive science: A defence. PsyArXiv. <https://osf.io/preprints/psyarxiv/zv6w8> [accessed on 5 April 2025].
- Pressnitzer, D., & McAdams, S. (2000). Acoustics, Psychoacoustics and Spectral Music. *Contemporary Music Review*, 19(2), 33–57. <https://doi.org/10.1080/07494460000640251>
- Rakowski, A. (1983). Pitch discrimination and musical interval recognition in backward masking. In R. Klinke & R. Hartmann (Eds.), *Hearing – physiological bases and psychophysics*. Berlin–Heidelberg: Springer. https://doi.org/10.1007/978-3-642-69257-4_46
- Rosiński, A. (2021). Wpływ wykształcenia muzycznego na grupowanie dźwięków sekwencji ABA-ABA w rytm galopujący. In A. Rosiński (Ed.), *Przestrzenie akustyki. Professional Acoustics*. Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.
- Rosiński, A. (2023). Influence of Music Education and Pitch Scales on the Grouping of the AB-AB Sequence Sounds. In A. Rosiński (Ed.), *Przestrzenie akustyki. Professional Acoustics 2*. Olsztyn: Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.
- Rosiński, A. (2024a). *Wpływ melodyki i rytmiki na tworzenie wielowariantowości odbioru wrażeń słuchowych. Przetwarzanie dźwięków w strumienie percepcyjne na przykładzie wybranych dzieł muzycznych*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego. <https://doi.org/10.31338/uw.9788323562757>
- Rosiński, A. (2024b). Multivariantism of Auditory Perceptions as a Significant Element of the Auditory Scene Analysis Concept. *Arts*, 13(6), 1–18. <https://doi.org/10.3390/arts13060180>
- Rosiński, A. (2024c). Pitch-based Stream Segregation of Tone Sequences: A Comparative Study of Musicians and Non-musicians. In A. Miśkiewicz & T. Rogala (Eds), *Sonic Art, Sound Perception and Audio Engineering*. Warsaw: Chopin University Press.
- Rosiński, A. (2024d). Influence of Music Education and Interval Size on Grouping of the AB-AB Sequence Sounds. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46, 1531–1537. https://escholarship.org/content/qt6db683ch/qt6db683ch_noSplash_0389048271078601fd62c14bcddae9c8.pdf?t=sgrg6x [accessed on 5 April 2025].
- Rosiński, A. (2025). Variable Properties of Auditory Scene Analysis in Music. *Arts*, 14(1), 1–21. <https://doi.org/10.3390/arts14010019>
- Shamma, S.A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–123. <https://doi.org/10.1016/j.tins.2010.11.002>
- Shin, M. (2012). *Metric Analyses of Selected Modern "Dances" for Organ*. PhD Thesis, Indiana University, Bloomington. <https://scholarworks.iu.edu/iuwrrest/api/core/bitstreams/f3990010-f0f7-4776-8bcd-6c7924ac714a/content> [accessed on 26 September 2024].
- Snyder, J. S., Lee, S.-K., Carter, O. L., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, 34(4), 1007–1016. <https://doi.org/10.1037/0096-1523.34.4.1007>
- Song, W., Ellermeier, W., & Hald, J. (2011). Psychoacoustic Evaluation of Multichannel Reproduced Sounds Using Binaural Synthesis. *Journal of Acoustical Society of America*, 129(6), 2063–2074. <https://doi.org/10.1121/1.3628323>
- Stevens, C. J. (2012). Music Perception and Cognition: A Review of Recent Cross-Cultural Research. *Topics in Cognitive Science*, 4(4), 653–665. <https://doi.org/10.1111/j.1756-8765.2012.01215.x>
- Szalárdy, O., Bendixen, A., & Böhm, T. M., Davies, L. A., Denham, S. L., & Winkler, I. (2014). The effects of rhythm and melody on auditory stream segregation. *Journal of the*

- Acoustical Society of America*, 135(3), 1392–1405.
<https://doi.org/10.1121/1.4865196>
- Thurman, D. L. (2010). *Insight on Selected Standard Audition Repertoire for the Trombone*. PhD Thesis, University of Alabama, Tuscaloosa. <https://ir-api.ua.edu/api/core/bitstreams/76deb6cf-4391-444e-bb2f-a3be582b3df9/content> [accessed on 26 September 2024].
- Timmers, R., Arthurs, Y., & Crook, H. (2020). Stream segregation revisited: Dynamic listening and influences of emotional context on stream perception and attention. *Consciousness and Cognition*, 85, 1–12, Article 103027. <https://doi.org/10.1016/j.concog.2020.103027>
- Wang, C. C., & Sogin, D. W. (1990). The Recognition of Melodic Fragments as Components of Tonal Patterns. *Psychology of Music*, 18(2), 140–147. <https://doi.org/10.1177/0305735690182003>
- Wang, X., Wei, Y., Heng, L., & McAdams, S. (2021). A Cross-Cultural Analysis of the Influence of Timbre on Affect Perception in Western Classical Music and Chinese Music Traditions. *Frontiers in Psychology*, 12, 1–20. Article 732865. <https://doi.org/10.3389/fpsyg.2021.732865>
- Wilson, A. (2012). Multistable Perception of Art-Science Imagery. *Leonardo*, 45(2), 156–161. https://doi.org/10.1162/LEON_a_00382