Musical Psychoacoustics: The Science of Sound Artifice

Kai Siedenburg

Introduction

Whereas sculptors configure space through form, musicians sculpt time through sound. From an acoustical perspective, however, music is notoriously hard to pin down. Unlike speech, which occupies a relatively stable frequency range around 80 to 6,000 Hz with identifiable temporal modulation and comfortable listening levels around 65 dB sound pressure level (SPL), music spans a wider acoustic space. Musically relevant frequencies range from 16 to 16,000 Hz, featuring a diverse array of sound sources (instruments), with instrument-specific spectro-temporal modulations and ideal listening levels that vary dramatically based on context, from roughly 40 to 120 dB SPL. Accordingly, modern audio codecs (encoding algorithms) process music and speech as distinct signal classes (Valin et al., 2016) and tuning hearing aids for music remains a significant challenge (Lesimple et al., 2024). Nonetheless, understanding the psychoacoustic characteristics of music is crucial for progress in audio engineering, hearing science, and music cognition.

Figure 1. *Chimera of Arezzo, a hybrid creature depicted in an Etruscan bronze statue dating to around 400 BC. Museo Archeologico Nazionale, Florence. Image reproduced from Wikimedia Commons, licensed under a Creative Commons Attribution 3.0 International (CC BY 3.0) license (creativecommons.org/licenses/by/3.0).*

Besides raw acoustic differences between music and speech, the way in which sound sources interact is rather unique in music. Speech communication succeeds when acoustic cues allow segregation of the target speaker from background noise — overcoming the "cocktail party problem" (Cherry, 1953). Music, on the other hand, often creates virtual sound sources, ones that do not exist in reality, by means of exploiting perceptual grouping principles. *Blending* is key in music: A brass section can achieve "one sound" despite multiple players, for instance. The opposite of blending is segregation. *Segregation* can be achieved even with the sounds of single monophonic instruments by splitting melodies in higher and lower voices (so-called virtual polyphony), creating the impression of an instrument accompanying itself. In this sense, music by design aims for perceptual effects or illusions that defy acoustic reality.

Albert Bergman, pioneer of auditory scene analysis research, explains this phenomenon using the notion of a chimera: "The Chimaera was a beast in Greek mythology with the head of a lion, the body of a goat, and the tail of a serpent. We use the word chimera metaphorically to refer to an image derived as a composition of other images. […] Natural hearing tries to avoid chimeric percepts, but music often tries to create them. It may want the listener to accept the simultaneous roll of the drum, clash of the cymbal, and brief pulse of noise from the woodwinds as a single coherent event with its own striking emergent properties. The sound is chimeric in the sense that it does

https://doi.org/10.1121/AT.2025.21.2.59

not belong to any single environmental object." (Bregman 1990, p. 459). A classic chimera of Etruscan origin is depicted in **Figure 1**: a beast that is certainly more than just a lion. Meticulously sculpted music yields similarly ambiguous and chimeric experiences — music is sound artifice that is more than its individual parts.

Musical psychoacoustics has the ambitious goal of reverse engineering music perception by relating the perceptual implications of music to the acoustical properties of music signals. More specifically, the goal is not to make sense of music signals as a whole — which is hardly possible — but to identify some of the core perceptual attributes and processes at the heart of music listening. This feat involves studying a plethora of interesting phenomena, involving elementary auditory attributes (such as pitch, timbre, and loudness), the processes of perceptual organization called musical scene analysis, and even questions around musical consonance and dissonance or inter-individual differences in listening abilities, to name a few. In this article, I will provide a highly selective review of work on musical psychoacoustics, mostly drawing from work that I conducted together with collaborators and PhD students in recent years. The focus will be on auditory frequency perception, auditory salience of voices in musical mixtures, and the role of inter-individual differences in music perception.

The Open Market of Auditory Frequency Perception

Music can be made of melodies, but what are melodies made of? A broad definition would be to state that melodies consist of sounds that change their periodicities or frequency content over time. Perceiving such frequency changes is crucial to following melodies in music or to understanding prosody in speech. This ability allows listeners to hear sounds as high or low and to follow melodic trajectories over time. Such frequency changes have traditionally been described as perceptual changes in pitch or timbre. However, the textbook notion that pitch and timbre are separate auditory attributes is only half true. Here, we will explore some of the connections between the two attributes.

The most important acoustical correlate of pitch is fundamental frequency (f_0). An important correlate of timbre is the spectral centroid (SC), the center of gravity of the spectrum, corresponding to the perceived brightness of sounds (Saitis and Siedenburg, 2020). The sounds of musical instruments feature strong correlations between f_0 and SC, which may, however, vary in strength dependending on the specific instrument class (Siedenburg et al., 2021b). For vocals and saxophones, for instance, there is little correlation between f_0 and SC, whereas for string instruments and oboes there is substantial correlation. The precise relation further depends on the pitch register and the dynamics (playing effort). Yet, even when the use of synthetic sounds rules out inherent correlations between these acoustic features, there are dependencies inscribed in the minds of most listeners. Specifically, changes in pitch and timbre are often confused with each other, as elegantly shown by Allen and Oxenham (2014). Using digitally generated harmonic tone complexes that varied in terms of f_0 and SC, they found that musicians' and non-musicians' discrimination thresholds were similar for SC changes, but musicians were better in discriminating f_0 changes. The authors further measured discrimination with concurrent variation in SC. With pitch as the target dimension, musicians clearly showed the least interference whereas nonmusicians had more difficulties in processing incongruent changes of f_0 and SC. When expressed in units of individual just noticable differences, however, musicians and nonmusicians showed comparable effects of interference. Using brain imaging, Ou et al. (2025) further found shared gradients across the auditory cortex for changes of pitch and brightness, suggesting a shared trajectory of cortical activation along each dimension. Other work consistently suggests that nonmusicians map these two stimulus attributes conjointly to a spatially vertical dimension (Pitteri et al., 2017), but not when both features are presented in isolation. That is, both f_0 and SC appear to be mapped onto the low/high perceptual dimension that is so central to melody perception.

Interestingly, there is one famous stimulus in hearing research that defies the low-to-high dimension of frequency perception (also known as pitch height): the famous "Shepard tone." Stanford psychologist Roger Shepard came up with a complex tone that features octave-spaced partial tones shaped by a global spectral envelope (Shepard, 1964); see **Figure 2A** for an illustration.

Shifting the spectral fine structure below the envelope towards the higher frequencies implies that some partial

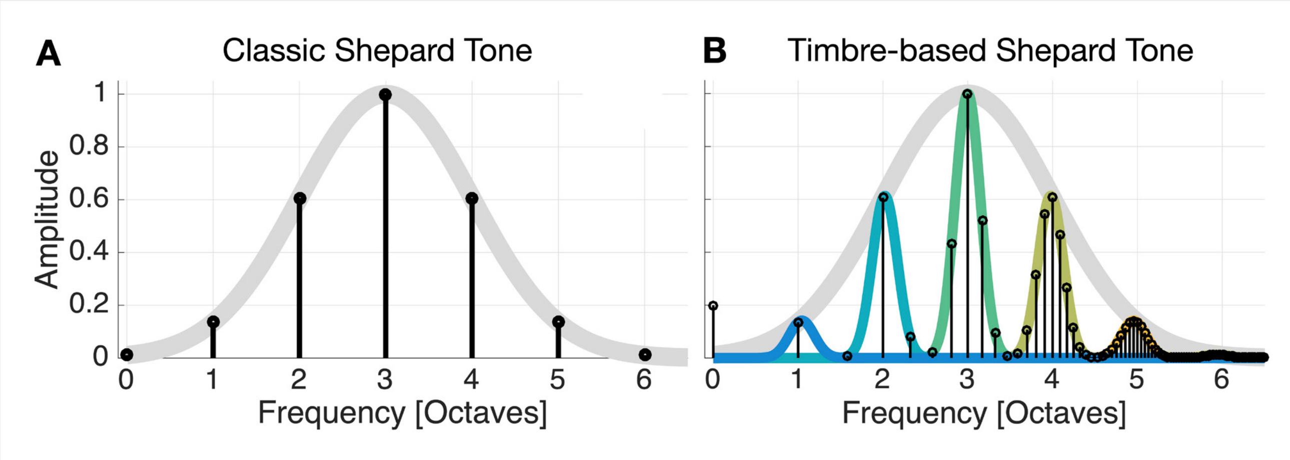


Figure 2. ***A:*** *Classic Shepard tone with octave-spaced partials and global spectral envelope (**gray**).* ***B:*** *Timbre-based Shepard tone with harmonic fine structure and local spectral envelope.*

tones disappear at the higher end whereas partial tones at the lower end are newly introduced. With octave-spaced partials, this somewhat unnatural stimulus yields a host of interesting phenomena in pitch perception: Importantly, the stimulus can be used to construct pitch spirals in which listeners hear pitch increasing or decreasing indefinitely. Furthermore, half-octave shifts of Shepard tones' fine structure tend to yield ambiguous responses with usually half of participants' responses indicating an upward shift and the other half indicating a downward shift. Some authors have interpreted this phenomenon as the auditory equivalent of the Necker cube, which viewers perceive as directing inwards or outwards but never both at the same time. In analogy, half-octave shifts of Shepard tones may be heard ascending or descending, but never both.

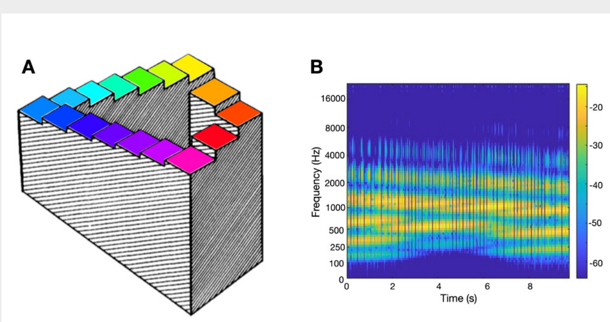
Diana Deutsch considered individual response profiles for such half-octave shifts and reported across several studies that individual listeners show distinct response patterns. One listener may have perceived half-octave shifts starting near G as always moving downwards and shifts starting near C# as always moving upwards, whereas a reversed orientation was observed for other listeners. This is the so-called "tritone paradox" (Deutsch, 2019). Work by Bruno Repp found that this phenomenon is highly dependent on the specific spectral envelope of the stimulus and can be swayed by prior context (Repp, 1997). Moreover, the influence of prior context has been clarified by more recent work, demonstrating that only a few tones suffice to almost completely determine response behavior for ambiguous half-octave shifts (Chambers et al., 2017).

It is straight forward to show that similar perceptual effects can be achieved by manipulations of the spectral envelope. See **Figure 2B**: Instead of using octave spaced partials, one may use partial tones with harmonic spacing, modulated by a local spectral envelope that is octave periodic by itself (Siedenburg, 2018). By shifting the local spectral envelope below the global spectral envelope while leaving the fine structure and hence f_0 intact, one obtains identical patterns of results and similar context dependencies as with classic Shepard tones. I have called this a "timbral Shepard illusion," because the perceptual attribute that is manipulated would, in classic terms, be considered to belong to the realm of timbre.

Daniel Pressnitzer, Jackson Graves, and I went a step further by constructing a stimulus that is ambiguous both in terms of spectral fine structure (pitch) and spectral envelope (timbre) (Siedenburg et al., 2023). Here, we added the trick of even-to-odd attenuation of partial tones' amplitudes to the stimulus described before with harmonic fine structure and octave-periodic local spectral envelope. Starting with a harmonic series and deleting all odd harmonics results in a tone one octave higher than before. We used this trick to construct tones with harmonic fine structure, yet featuring pitch circularity (note that the original Shepard tone in contrast had octave-spaced harmonics). Decreasing f_0 by semitones while attenuating amplitudes of odd harmonics by 2 dB per semitone step yields a sound that has an attenuation of odd harmonics by 22 dB after 11 semitone steps, a sound that may indeed be heard as one octave higher compared to its non-manipulated counterpart (cf., Warren et al., 2003; Deutsch et al., 2008). It turned out that this manipulation of the spectral fine structure resulted in ambiguous half-octave shifts, as for the classic Shepard-tone. Simultaneous shifts of the spectral envelope had similar effects. Overall, these shifts resulted in a type of "two-dimensional (2D) Shepard tones," where both cues, f_0 or envelope, contributed to the perceived direction of the frequency shift. This effect is illustrated in **Figure 3**, combining the impossible staircase with a color circle as a visual analogy to the described stimuli comprising two quasi-circular dimensions.

The described stimuli are not necessarily typical of what we are accustomed to hearing in natural environments, but they can be used to learn about general principles of auditory perception. In experiments using these 2D Shepard tones, we observed that listeners reported antagonistic shifts (cues going in different directions) as falling into one "up/down" dimension. Listeners only rarely reported hearing antagonistic shifts moving in two different directions. Moreover, the strength of the individual cue varied depending on whether stimuli were harmonic or inharmonic. Using a computational model, we found that for harmonic tones, f_0 cues were given much greater weight, whereas for inharmonic tone complexes, the resolved harmonics played a similarly important role. Notably, there were clear indications of strong individual differences in how subjects weighted these auditory cues. Taken together, these results point towards pitch and timbral brightness potentially being part of a compound perceptual dimension that tracks frequency changes over time. As an analogy, auditory frequency perception may be interpreted as an open market, where individual frequency cues such as resolved harmonics, the profile of unresolved harmonics, or temporal cues are traded with one another, depending on the auditory context and the inclinations of individual listeners.

Figure 3. ***A:*** *Analogy of impossible staircase combined with a hue circle, akin to the described 2D Shepard tones.* ***B:*** *Auditory spectrogram for a sequence of 2D Shepard tones with simultaneously increasing* f_0 *and decreasing local spectral envelope.*



The Invisible Hand of Musical Scene Analysis

On the busy marketplace of music perception, melodies and instrumental sounds rarely appear alone but instead tend to intermingle. As sound waves from multiple instruments overlap in time and frequency, music needs to be disentangled by the ear and the brain. Sequences of tones may be grouped to form a melody, depending on their proximity in time and frequency. Depending on whether acoustic components are grouped into one melody (or "stream") or another, different rhythmic relations may emerge. As described by Bregman (1990), auditory scene analysis (ASA) denotes the ways in which the auditory system organizes and represents acoustic signals as auditory events and streams. ASA is an elementary process of auditory perception that, on the one hand, depends strongly on acoustical cues and peripheral physiology, which is also called "primitive" ASA. On the other hand, ASA connects with a variety of cognitive and knowledge-based processes, also called "schema-based" ASA.

For measuring the perceptual implications of ASA, early studies have mainly used synthetic stimuli. A classic stimulus is the so-called "galloping horse," a repeating sequence of two alternating tones that create a repeating ABA_ pattern (the underscore denoting silence; van Noorden, 1975). Importantly, their frequency difference and inter-stimulus time interval determine whether the two tones A and B are heard as forming one or two streams. This elementary stimulus has been used numerous times in the literature, and it demonstrates that both temporal and frequency characteristics directly constrain our perception of auditory scenes. Bey and McAdams (2002) introduced a listening task based on slightly more

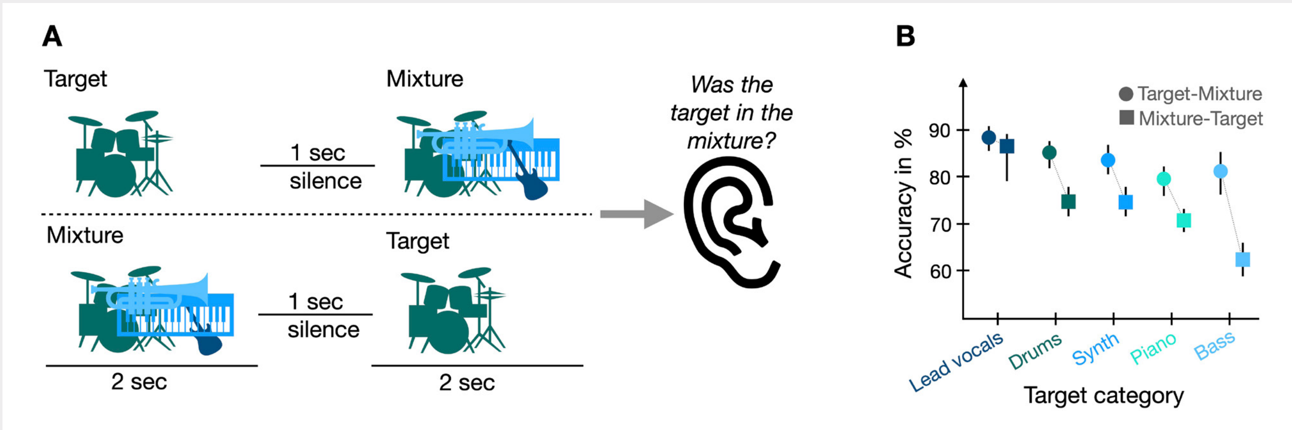


Figure 4. *Testing the salience of musical instrument categories in popular music.* ***A:*** *Task conditions, wherein a target sound is followed by a mixture (**top row**) or a mixture sound is followed by the target (**bottom row**). In both conditions, participants indicate whether they heard the target in the mix (yes/no).* ***B:*** *Effects of musical instrument category (x-axis) and condition (**circles** vs. **squares**). Note the vanishing effect of task condition for the lead vocals compared to all other instruments. Data showing condition means and 95% confidence intervals from Bürgel et al. (2021).*

complex stimuli, so-called interleaved melody recognition, wherein two random melodies, one target and one distractor melody, were interleaved with each other, creating a mixture. The target is presented either before (target-mixture condition) or after the mixture (mixture-target condition). Whereas the target-mixture condition tests selective listening abilities, the mixture-target condition tests a form of undirected listening. Again, mean frequency differences between target and distractor sequences strongly affect the strength of segregation of the two melodies.

In recent work, we have extended the interleaved melody recognition paradigm by using natural mixtures of popular music instead of the sterile sequences of pure tones. In this case, the target-mixture condition would test directed, selective listening in popular music, i.e., the ability to hear out individual sound components from a mix. The mixture-target condition, on the contrary, would test undirected listening. We interpreted the performance gap between the two conditions as a measure of the salience of sounds; the less the deviation from the directed listening, the more likely that this sound was attended to by listeners in the undirected condition. **Figure 4A** shows a sketch of the two conditions, while panel **B** shows the experimental results from Bürgel et al. (2021).

Notably, lead vocals featured almost no difference between the two conditions, suggesting a high degree of salience. To the contrary, all other instrument categories featured notable differences in performance in the two conditions. Two other experiments in the study reproduced these results and confirmed that this specific degree of vocal salience was not due to differences in sound level or dominance in certain spectral regions. In a follow-up study, we tracked down the acoustical features that enable lead vocals to stand out in mixes (Bürgel and Siedenburg, 2023), once again reproducing the unique salience effect with a separate dataset of popular music tunes. Vocal melodies were now modified in their frequency modulations, using the "autotune effect," popular in contemporary music production. The effect removes frequency modulations that are particularly prevalent in note transitions due to natural inaccuracies of singing voices. With pitch trajectories tied to a fixed pitch grid, vocals were not as salient. Importantly, when other instruments, such as the saxophone, played the vocal melody, this melody was particularly salient with the presence of frequency modulations extracted from the original vocal melody. This result suggests that natural inaccuracies in tone production let vocals stand out from the mixture of an instrumental accompaniment.

The most drastic difference between conditions (around 20 percentage points) was observed for bass instruments, as visible in **Figure 4B**, indicating a lack of salience. Could this difference be attributed to a specific absolute bias in human frequency perception, such as preferring to attend to higher frequencies in undirected listening (Pressnitzer et al., 2018)? Additional results from Bürgel et al. (2024) indicated rather a relative bias in frequency perception: the lower and higher outer frequency components in a mixture, that is, so-called "edge frequencies," regardless of which absolute spectral band, possessed drastically more salience compared to middle frequencies. We interpret these results as pointing towards the musical structure, not the acoustics, which often places bass instruments into the background rather than the foreground of musical scenes.

More generally, it seems important to note that music is a highly structured stimulus, deeply attuned to ASA principles. David Huron has demonstrated how central music theoretical rules of voice leading in polyphonic music can be derived from what is known about ASA principles (Huron, 2016). Research on orchestration is pointing out the relation between ASA and the stratification and structuring of highly complex orchestral compositions (McAdams et al., 2022). A similar case can be made beyond classical music where music producers work hard to find the right balance between simplicity and complexity of the musical scenes they produce, which may manifest itself in either clear or muddy listening experiences. Music performers strive to sync their musical phrasing, intonation, tone or sound, and timing with fellow ensemble members to collectively sculpt one musical idea. Sound engineers work hard to create perceptual transparency and giving instruments their space, which allows listeners to discern sound elements and instruments. Room acoustics shape musical clarity, and performers react to the acoustics of a performance space (Schärer Kalkandjiev and Weinzierl, 2013). In all of these domains, musicians face their interactions with musical sound being constrained by the principles of ASA. Once again by means of analogy, I would argue that ASA acts as a type of invisible hand of music generation, which regulates individual actors (composers, performers, sound producers), without a single individual being necessarily aware of the underlying perceptual dynamics at play.

Inter-Individual Differences in Music Perception

Given the many perceptual affordances of music, it is easy to make claims such as "We all hear music differently." But what is the empirical evidence? By considering individual differences in auditory perception, we are addressing important issues. Take sensory hearing loss, which poses a major barrier to hearing for many individuals. Hearing loss is among the most prevalent forms of impairment today, and the World Health Organization estimates that 630 million people will be hard of hearing in 2030. People with hearing loss not only suffer from increased hearing thresholds but also from reduced dynamic range, worsened frequency resolution, distortion, and worsened binaural processing (Moore, 2007). These limitations also imply that elementary perceptual processes such as primitive ASA in music may be impaired for affected individuals, because their sensory representation of sound is washed out. Beyond physiology, it is clear that experience and knowledge shape who we are and how we listen. For example, if a violinist has trained on their instrument for their whole life, this experience will have consequences on how they perceive nuances in pitch and timbre or on what they hear in musical mixtures.

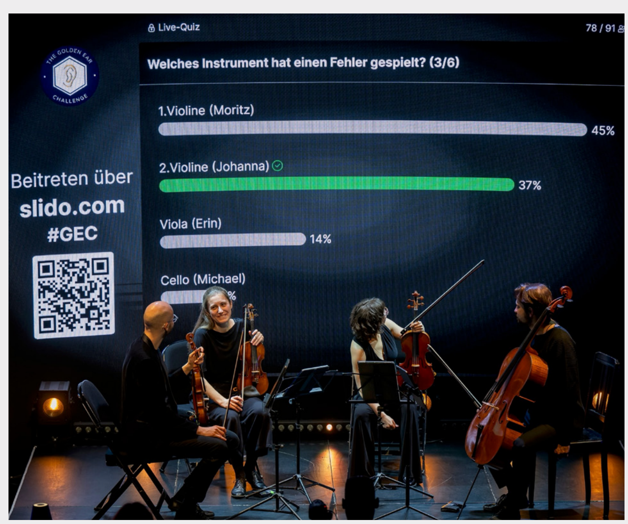
Recently, we have approached the topic of individual differences by developing a new test of musical scene analysis (MSA) ability (Hake et al., 2023). The test uses the target-mixture paradigm from **Figure 4A** and a Bayesian adaptive procedure to efficiently approach individuals' ability levels. Test scores from 96 participants with varying age range and varying degrees of hearing loss were analysed in terms of the importance of four predictors of individual differences (Hake et al. 2025): hearing thresholds (pure tone average, PTA), musical experience or sophistication as measured by the Goldsmiths Musical Sophistication index (Müllensiefen et al., 2014), chronological age, and working memory capacity. We found negative associations of MSA scores with hearing thresholds and positive associations with musical training, whereas the variables age and working memory showed comparatively small associations with MSA scores. To the contrary, speech reception thresholds were dominated by hearing thresholds alone. Consistent with prior work (Siedenburget al., 2020, 2021a; Kirchberger and Russo, 2015), this study reinforces that hearing loss negatively affects ASA not only for speech but also for music, even if other determinants of individual differences may not be negligible.

Searching for real-world manifestations of individual differences in music perception, we recently went beyond the lab to collect data. Specifically, we have started to explore questions around selective listening in music in a live concert event dubbed “The Golden Ear Challenge” (Bürgel and Siedenburg, 2023, goldenears.de). This event was a gamified concert format, wherein the audience was invited to hear out intentional performance errors of musicians, thus testing selective listening abilities. Using a smartphone-based app, listeners were instructed to report on which instrument in the ensemble played a wrong note (see **Figure 5**). Listeners with the most accurate and fastest responses finally received the “Golden Ear Award” as a humorous gimmick. The format serves as a great way to reach out and communicate our research on music perception to musically interested audiences. So far, we have witnessed an extraordinary range of accuracies, from chance to (almost) perfect performance, a range rarely encountered in lab experiments with homogeneous groups of participants.

Beyond quantifying the effects of musical training on ASA ability, the study of qualitative shifts in perception holds promise. As an example, people tend to assume that musicians are capable of perceiving much more detail in music due to an analytical mode of listening compared to nonmusicians. Patrick Susini and colleagues have recently provided a compelling empirical demonstration of this assumption (Susini et al., 2020). They played listeners carefully constructed melodies that were varied in terms of their local or global contour (i.e., the pattern of up and down movements), or both. Listeners were instructed to attend to either local or global characteristics of the melody and report when they heard changes in the melody. Musicians generally achieved better scores. Importantly, however, nonmusicians showed a bias towards global features whereas musicians were biased to attend to local features. That is, musical training over longer time spans may indeed contribute to a perceptual reorganization that diminishes the dominance of global features of melodies in favour of local features. It seems plausible that this situation may extend towards ASA in natural music, where musicians may be biased to analytically segregate chords or textures into their individual components, whereas nonmusicians may be biased to attend to global properties of the music.

Recent years have also seen enormous progress in the characterization of cross-cultural differences in music perception. This issue is particularly complex and thorny due to long-standing controversies about the origin and universality of music. As an example, take the origin of musical consonance and dissonance, which has sparked the interest of scholars since the ancient Greeks. Because western music is globally ubiquitous and musical structures are implicitly learned, it has become almost impossible to find test participants without previous exposure to Western music. Traveling a long way to address this issue, Josh McDermott and a team of researchers have worked with participants from the Tsimane’, a native Amazonian society with minimal exposure to Western culture. Tsimane’ musical culture is very distinct in that it lacks polyphony, harmony, or group performance, arguably with effects on auditory perception. Testing the perception of diads, McDermott et al. (2016) found that native Amazonians were indifferent to musical dissonance, such as the piercing sound of the tritone. Jacoby et al. (2019) further observed that Tsimane’ did not show signs of octave equivalence: when asked to sing back intervals presented beyond their singing range, Westerners would tend to transpose intervals down by

Figure 5. *Musicians of The Golden Ear Challenge event, reviewing the audience’s responses in spotting the instrumentalist who intentionally played a wrong note. Picture courtesy of Daniel Schmidt, University of Oldenburg.*



octaves. Tsimane', however, accurately reproduced intervals but failed to transpose by octaves, suggesting that the octave does not play as a critical role in their perception of musical pitch. Carefully manipulating the strength and spacing of partial tones in complex sounds, Marjieh et al. (2024) showed that musical consonance judgements are affected by timbre and scale structure. Specifically, tones with a stretched or compressed partial series yielded stretched or compressed consonance profiles, respectively. Interestingly, the latter results provide a direct link between scale and partial tone structure. For instance, the five-tone slendro scale of Javanese gamelan coincides with the peaks of the consonance profile obtained with sounds from the bonang, a Javanese gong. Overall, these results highlight how aspects of musical culture are interwoven with the ways in which elementary musical structures are perceived.

Summary and Perspectives

In this article, I provided some selective spotlights of current topics in musical psychoacoustics. Borrowing an idea from Albert Bregman, namely that music often has a chimeric nature, I suggested that music acts as sound artifice, playing tricks with human auditory perception. I then attempted to highlight a few of the tricks that have been used by research on musical psychoacoustics to reverse engineer the workings of the auditory system in response to music. This research involves the construction of ambiguous tones such as 2D Shepard tones that highlight the connectedness of the perceptual dimensions of pitch and timbre. I outlined how vocal sounds bear a particular auditory salience in mixtures and how we studied the acoustical origin of this salience by transferring characteristic vocal frequency micro-modulations to other instruments. Musical scene analysis was portrayed as an "invisible hand" in every stage of music composition, performance, and recording, even before it reaches the ear of the listener. Finally, I considered ways in which individual differences in musical perception can be explored within or even across cultures. It is the rich pallette of topics in musical psychoacoustics that make it a particularly exciting field of study, full of intriguing and unexplored questions.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project ID 352015383 — SFB 1330 A6. The author wishes to thank Keith Wilson, Kat Setzer, and Daniel Pressnitzer for useful help and advice.

References

Allen, E. J., and Oxenham, A. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America* 135(3), 1371 – 1379.

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Perception and Psychophysics* 64(5), 844 – 854.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.

Bürgel, M., Mares, D., and Siedenburg, K. (2024). Enhanced salience of edge frequencies in auditory pattern recognition. *Attention, Perception, and Psychophysics* 86(10), 2811 – 2820.

Bürgel, M., Picinali, L., and Siedenburg, K. (2021). Listening in the mix: Lead vocals robustly attract auditory attention in popular music. *Frontiers in Psychology* 12, 769663.

Bürgel, M., and Siedenburg, K. (2023). Salience of frequency micromodulations in popular music. *Music Perception* 41(1), 1 – 14.

Chambers, C., Akram, S., Adam, V., Pelofi, C., Sahani, M., Shamma, S., et al. (2017). Prior context in audition informs binding and shapes simple features. *Nature Communications* 8, 15027.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America* 25(5), 975 – 979.

Deutsch, D. (2019). *Musical Illusions and Phantom Words: How Music and Speech Unlock Mysteries of the Brain*. Oxford University Press, Oxford.

Deutsch, D., Dooley, K., and Henthorn, T. (2008). Pitch circularity from tones comprising full harmonic series. *The Journal of the Acoustical Society of America* 124(1), 589 – 597.

Hake, R., Bürgel, M., Nguyen, N. K., Greasley, A., Müllensiefen, D., and Siedenburg, K. (2023). Development of an adaptive test of musical scene analysis abilities for normal-hearing and hearing-impaired listeners. *Behavior Research Methods* 55(4), 5456 – 5481.

Hake, R., Müllensiefen, D., and Siedenburg, K. (n.d.). Individual differences in musical scene analysis ability. Under review.

Huron, D. (2016). *Voice Leading: The Science Behind a Musical Art*. MIT Press, Cambridge, MA.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdes, J., Ossandon, T., and McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology* 29(19), 3229 – 3243.

Kirchberger, M. J., and Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing* 36(2), 217 – 228.

Lesimple, C., Kuehnel, V., and Siedenburg, K. (2024). Hearing aid evaluation for music: Accounting for acoustical variability of music stimuli. *JASA Express Letters* 4(9), 1 – 7.

Marjieh, R., Harrison, P. M. C., Lee, H., Deligiannaki, F., and Jacoby, N. (2024). Timbral effects on consonance disentangle psychoacoustic mechanisms and suggest perceptual origins for musical scales. *Nature Communications* 15(1), 1482.

McAdams, S., Goodchild, M., and Soden, K. (2022). A taxonomy of orchestral grouping effects derived from principles of auditory perception. *Music Theory Online* 28(3).

McDermott, J. H., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535(7613), 547 – 550.

Moore, B. C. (2007). *Cochlear Hearing Loss: Physiological, Psychological, and Technical Issues*. John Wiley and Sons, West Sussex, United Kingdom.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE* 9(2), e89642.

Ou, Y., Allen, E. J., Kay, K. N., and Oxenham, A. J. (2025). Cortical substrates of perceptual confusion between pitch and timbre. *bioRxiv*.

Pitteri, M., Marchetti, M., Priftis, K., and Grassi, M. (2017). Naturally together: Pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians. *Psychological Research* 81(1), 243 – 254.

Pressnitzer, D., Graves, J., Chambers, C., De Gardelle, V., and Egré, P. (2018). Auditory perception: Laurel and yanny together at last. *Current Biology* 28(13):R739 – R741.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception* 26(5), 645 – 665.

Schärer Kalkandjiev, Z., and Weinzierl, S. (2013). The influence of room acoustics on solo music performance: An empirical case study. *Acta Acustica united with Acustica* 99(3), 433 – 441.

Saitis, C., and Siedenburg, K. (2020). Brightness perception for musical instrument sounds: Relation to timbre dissimilarity and source-cause categories. *The Journal of the Acoustical Society of America* 148(4), 2256 – 2266.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America* 36(12), 2346 – 2353.

Siedenburg, K. (2018). Timbral Shepard-illusion reveals perceptual ambiguity and context sensitivity of brightness perception. *The Journal of the Acoustical Society of America*, 143(2) EL93 – EL98.

Siedenburg, K., Röttges, S., Wagener, K., and Hohmann, V. (2020). Can you hear out the melody? Testing musical scene perception in young normal-hearing and older hearing-impaired listeners. *Trends in Hearing* 24, 1 – 15.

Siedenburg, K., Goldmann, K., and van de Par, S. (2021a). Tracking musical voices in Bach's The Art of the Fugue: Timbral heterogeneity differentially affects younger normal-hearing listeners and older hearing-aid users. *Frontiers in Psychology* 12, 608684.

Siedenburg, K., Jacobsen, S., and Reuter, C. (2021b). Spectral envelope position and shape in orchestral instrument sounds. *The Journal of the Acoustical Society of America* 149(6), 3715 – 3727.

Siedenburg, K., Graves, J., and Pressnitzer, D. (2023). A unitary model of auditory frequency change perception. *PLOS Computational Biology* 19(1), 1 – 30.

Susini, P., Jiaouan, S. J., Brunet, E., Houix, O., and Ponsot, E. (2020). Auditory local – global temporal processing: Evidence for perceptual reorganization with musical expertise. *Scientific Reports* 10(1), 16390.

Valin, J.-M., Maxwell, G., Terriberry, T. B., and Vos, K. (2016). High-quality, low-delay music coding in the Opus codec. *arXiv preprint*, arXiv:1602.04845.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences* (Vol. 3). Institute for Perceptual Research, Eindhoven, Netherlands.

Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences* 100(17), 10038 – 10042.

Like this article?
Here's more from the author:

LISTEN

Lead Vocal Levels in Pop Music

bit.ly/AA-pop-vocals

Is there an ideal level for lead vocals compared to accompaniment in popular music? Researchers at University of Oldenburg investigated this question by analyzing the Billboard Hot 100 year-end list from 1946 to 2020 as well as Grammy award nominees from 1990 to 2020. In this Across Acoustics interview, author Kai Siedenburg discusses what his group learned about an important aspect of music mixing and the impact the research may have.

About the Author

Kai Siedenburg
kai.siedenburg@uol.de

Institute of Music
Carl von Ossietzky
Universität Oldenburg
Ammerländer Heerstr 114-118
26129 Oldenburg, Germany

Kai Siedenburg is Professor of Systematic Musicology at Carl von Ossietzky Universität Oldenburg (Germany). He served as Professor of Communication Acoustics at Graz University of Technology (Austria) from 2024 to 2025. He holds a PhD in Music Technology from McGill University (2016; Montréal, Canada) and earlier studied mathematics and musicology at Humboldt University Berlin (Germany). His research explores how listeners with diverse hearing profiles make sense of musical sound, covering psychoacoustics, musical acoustics, auditory scene analysis, and hearing devices. He is co-editor of Springer's 2019 volume *Timbre: Acoustics, Perception, and Cognition* and has served as Associate Editor of *The Journal of the Acoustical Society of America* since 2023. In 2025 he received the Early Career Award of the International Commission on Acoustics.