On the Relationship of Perceived Immersion to Acoustic Properties of Surround Sound Music Reproduction

Roman Kiyan[1], Jakob Bergner[1], Stephan Preihs[1], Daphne Schössow[1], Yves Wycisk[2], Kilian Sander[2], Reinhard Kopiez[2], Jürgen Peissig[1], Friedrich Platz[3]

[1] *Institut für Kommunikationstechnik, Leibniz Universität Hannover*
{roman.kiyan, jakob.bergner, stephan.preihs, daphne.schoessow, peissig}@ikt.uni-hannover.de

[2] *Musikwissenschaftliches Institut, Hochschule für Musik, Theater und Medien Hannover*
{yves.wycisk, kilian.sander, reinhard.kopiez}@hmtm-hannover.de

[3] *Institut für Musikwissenschaft, Hochschule für Musik und Darstellende Kunst Stuttgart*
friedrich.platz@hmdk-stuttgart.de

Introduction

Multichannel loudspeaker setups commonly known as "surround sound" systems promise to deliver an immersive music listening experience by reproducing spatial cues contained in the channel-based audio material. Comparisons of perceptual effects of sound reproduction over different speaker setups have been undertaken in a number of studies [1, 2, 3, 4, 5], evaluating various perceptual attributes for different types of content and loudspeaker configurations. It is clear that in such evaluations, the stimuli themselves play a great role along with the specific conditions in terms of the reproduction setup and listening situation. While listening conditions and speaker setups may be standardized to some extent [6, 7], experimental stimuli for such studies have been produced in various ways. In channel-based spatial audio, these include the creation of different mixes for different reproduction formats [4, 5], downmixes from the format with the greatest number of loudspeakers [1, 3, 2] or muting of subsets of channels [2].

In the main experimental study of the research project Richard Wagner 3.0 the phenomenon of immersion in music listening has been investigated in the psychological, physiological and acoustic domains, with immersion being understood as a high-level psychological construct rather than a basic perceptual attribute. The Immersive Music Experience Inventory (IMEI) [8] has been created within the scope of the project to be used by participants assessing immersion in music.

Based on the results of the study, a modeling approach for the prediction of immersion from acoustic features has been developed [9]. In contrast to this data-driven general approach, this paper is focused on relating some particular observations of the study to the acoustic features of the stimuli.

Experimental study and stimuli

The experimental study with 57 naive subjects (31 female, 26 male) has been carried out in the Immersive Media Laboratory (IML) at IKT [10]. Among other setups, this listening room can represent setup D (five loudspeakers in the horizontal plane and four height loudspeakers) according to ITU-R BS.2051-2 [7]. Musical stimuli used in the study have been mixed by two sound engineers in stereo, 5.1 (2D) and 5.1.4 (3D) versions, aiming to preserve overall aesthetics of the mixes and make them vary only in terms of spatial audio properties. The stimuli incorporate various recording and production techniques such as 3D microphone setups, spatial mixes from individual instrument channels as well as upmixes and downmixes from existing material. An overview of the stimuli used in the study is given in Table 1.

The method for acoustic analysis of the stimuli is adapted from Bergner and Peissig [11] and Bergner et al. [12]. It is based on re-recordings of the stimuli at the listening position using an mh acoustics *Eigenmike*® [13] and the derivation of sound field parameters from a spherical harmonic representation of the signals. Binaural parameters have been computed from a binaural rendering of this Ambisonics representation based on head-related transfer functions (HRTFs) of a Neumann KU 100 dummy head [14, 15]. In contranst to the approach of Bergner et al. [11, 12] which is based on factor analysis of the sound field parameters, mean over time and variance over time have been chosen as summary features derived from sound field parameter time series. All features have been normalized across the stimulus set to the interval $[0, 1]$.

Immersion related to acoustic properties

Based on an analysis of the sound field feature statistics, normalized inter-aural cross-correlation (IACC) as well as diffuseness as defined by Pullki [18] have been identified as particularly relevant features in a model of IMEI immersion based on sound field features [9]. This is consistent with previous findings on perception of sound spaciousness reported in the literature [19, 18, 20]. Hence the temporal means of normalized IACC (`niacc_mean`) and diffuseness (`diff_mean`) will be regarded as sound field features of interest in this paper. For simplicity, the broadband variants of both features will be considered here. Based on these features, some general trends are to be identified, followed by specific observations for

Table 1: Stimuli used in the experimental study. Short forms of titles are highlighted in bold.

Piece	Composer	Genre, Ensemble	Production	Format					
				Mono	Stereo	2D	3D	3D (-5 dB)	multi Mono
Walkürenritt	R. Wagner	Opera (orchestra, fem. voices)	manual upmix from commercial 5.1 [16]		×	×	×		
School's Out	A. Cooper, M. Bruce, G. Buxton, D. Dunaway, N. Smith	Rock (band with male voice)	spot mics + 3D ambience (live)		×	×	×		
In a **Mellow** Tone	D. Ellington, performed by J. Berger	Jazz (band with fem. voice)	3D mic. setup + support mics		×	×	×		
Im **Wunderschönen** Monat Mai	R. Schumann	Art song (piano, male voice)	3D mic. setup + support mics		×	×	×		
Laudate Dominum	J. Vila	Choir (12 singers)	3D mic. setup + support mics		×	×	×		
Bilder einer Ausstellung – Das große Tor von Kiew	M. Mussorgsky	Classical (large orchestra)	3D mic. setup + support mics		×	×	×	×	
Die **Hantel**	F. Thiesen	Electropop (synthesizers, voices)	multitrack studio production	×	×	×	×		
Rokoko variations – Finale	P. Tchaikovsky	Classical (woodwind quintet, cello)	manual upmix from commercial 5.1 [17]	×	×	×	×		×

some of the stimuli.

General observations

Figure 1 shows mean IMEI ratings obtained in the experimental study for each stimulus (represented by a circle) against both IACC and diffuseness computed for that stimulus. Circle colors represent the various reproduction formats. A color-coded surface in this three-dimensional space has been interpolated between the points. It is apparent that there is a general upward trend in mean immersion ratings as IACC decreases and diffuseness increases, which is in line with common perceptual interpretations of these sound field features. It is worth noting, however, that the relationship between the sound field features and the *reproduction formats* is not monotonous, i. e. some stereo stimuli yield *"better"* values of the sound field features than some 2D and 3D stimuli along with higher mean IMEI ratings. This is well-demonstrated by Figure 2, which displays two cross-sections of the three-dimensional space of Figure 1, demonstrating the relationship between IMEI scores and the two sound field features.

With Figure 2a showing IMEI ratings by musical piece and by reproduction format against diffuseness and Figure 2b displaying the same ratings against IACC, the trends observed in Figure 1 become somewhat more apparent. These two views show that several stereo stimuli feature IACC values comparable to the 2D and 3D stimuli, yet are generally rated somewhat lower on the IMEI scale. This can be explained by the stereo and spatial audio stimuli being more clearly distinct in terms of diffuseness, where 2D and 3D reproduction generally leads to a more diffuse sound field as compared to stereo. Based on the two sound field features, it is not readily possible to discern between the 2D and 3D formats, which is reflected in the IMEI ratings as well.

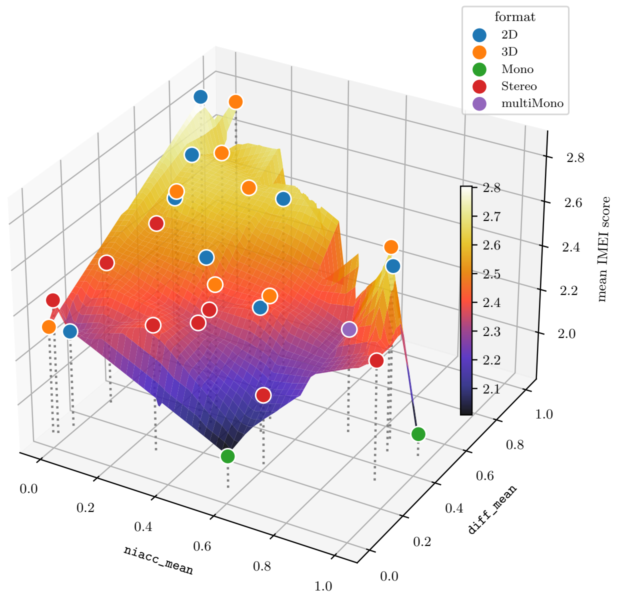


Figure 1: Mean IMEI immersion ratings for the experimental stimuli displayed against normalized IACC (`niacc_mean`) and diffuseness (`diff_mean`). The surface interpolated through the points exhibits a descending trend towards `niacc_mean` $\to 1$ and `diff_mean` $\to 0$.

Observations for specific pieces

Some particular observations shall be discussed for the pieces *Hantel* and *Walküre*. The piece *Hantel* exhibits rather high IACC for all reproduction formats as shown in Figure 2b, yet the IMEI ratings for the stereo, 2D and 3D versions are higher than this would suggest (this corresponds to the anomaly at the right-hand side of the interpolated surface in Figure 1). The IACC values for stereo and spatial reproduction may be skewed in the broadband variant of the feature due to *Hantel* being a bass-heavy electropop piece. If the feature is evaluated in the frequency band above 2.5 kHz, the IACC values for the stereo, 2D and 3D versions of *Hantel* are in line with the other pieces as shown in Figure 2c.

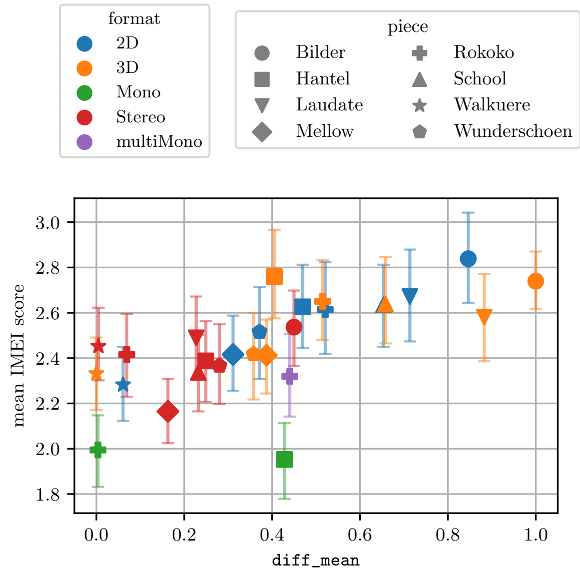
For *Walküre*, all three variants (stereo, 2D, 3D) yield very low diffuseness values, but their IMEI ratings are in line with stimuli of medium diffuseness as highlighted by Figure 2a. At the same time, the three variants of this piece are rated lower than stimuli of comparable IACC as seen in Figure 2b – which may in turn be explained by their low diffuseness. Thus, this particular example serves to demonstrate that a single sound field feature may be insufficient to fully characterize the spatial sound properties of a stimulus and its immersive potential.

Note that the two pieces *Hantel* and *Walküre* exhibiting peculiar behavior in terms of their acoustic features and defying the trends in terms of immersion ratings are a multitrack studio production (*Hantel*) and an existing 5.1 recording with an upmix for the 5.1.4 variant and a downmix for stereo (*Walküre*). On the other hand, the pieces best adhering to the observed trends both in terms of sound field features and immersion ratings (*Laudate*, *Bilder*, *Mellow*, *Wunderschön*) are recordings made with various 3D microphone setups in particular recording spaces. This suggests that the nature of the relationship of perceived immersion to acoustic features may itself be dependent on whether spatial sound field features of a stimulus originate from reverberation or from artificial effects and production techniques. This is plausible considering that the two features used here are most commonly applied for the analysis of sound fields produced by real sources in naturally reverberant environments, such as instruments in concert halls. Nonetheless, the pieces *Rokoko* and *School* do not exhibit any peculiarities in terms of the relationship between sound field features and immersion despite also originating from upmixing and downmixing (*Rokoko*) and individual instrument recordings blended with 3D audience sounds (*School*), respectively.

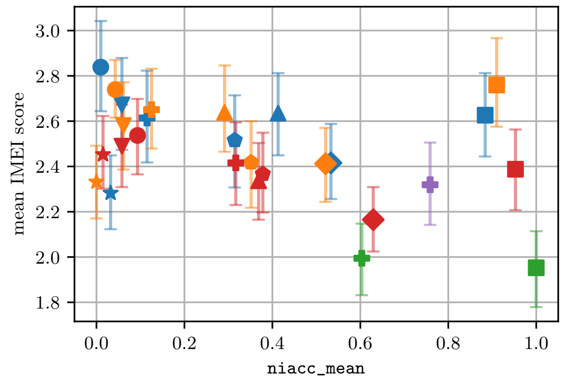
Summary

This paper presents some observations related to sound field properties of channel-based spatial audio music stimuli and their relationship to immersive musical experience as characterized by IMEI scores obtained in an experimental study.

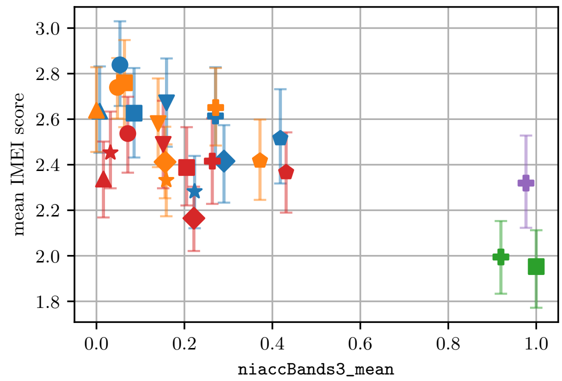
Immersion is observed to generally follow expected trends based on diffuseness and IACC that each stimulus yields in a re-recording at the listening position. Interestingly, the relationship between immersion and sound field properties appears to be independent of the particular reproduction format that rendered a sound field with those properties. For some stimuli produced using artificial spaciousness effects (e. g. upmixes,



(a) Mean IMEI against diffuseness (diff_mean), normalized to [0, 1] across the stimulus set.



(b) Mean IMEI against normalized IACC (niacc_mean), normalized to [0, 1] across the stimulus set.



(c) Mean IMEI against normalized IACC in the frequency band above 2.5 kHz (niaccBands3_mean), normalized to [0, 1] across the stimulus set.

Figure 2: Mean IMEI immersion ratings and bootstrapped 95 % confidence intervals for the experimental stimuli (cross-sections of Figure 1).

multitrack recordings with arbitrarily panned sources), acoustic features and immersion ratings contradict the general trends observed in stimuli including natural reverberation. On the other hand, other "artificially" produced stimuli yield expected sound field features and corresponding immersion ratings. This shows that further research on parameters suitable for acoustic and perceptual characterization of spatial audio musical stimuli is required due to the diversity in production techniques.

Acknowledgement

The project Richard Wagner 3.0 has been funded by "Niedersächsisches Vorab", a joint program by the Volkswagen Foundation in conjunction with the Lower Saxony Ministry for Science and Culture (ZN3497).

References

[1] M. Schoeffler, A. Silzle, and J. Herre. "Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience". In: *IEEE Journal of Selected Topics in Signal Processing* 11.1 (2017), pp. 75–88. DOI: 10.1109/jstsp.2016.2639325.

[2] A. Silzle, S. George, E. A. Habets, and T. Bachmann. "Investigation on the quality of 3D sound reproduction". In: *Proceedings of the International Conference on Spatial Audio.* 2011, pp. 334–341.

[3] C. Eaton and H. Lee. "Subjective evaluations of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings". In: *Acoustical Science and Technology* 43.2 (2022), pp. 149–161. DOI: 10.1250/ast.43.149.

[4] J. Francombe, T. Brookes, R. Mason, and J. Woodcock. "Evaluation of Spatial Audio Reproduction Methods (Part 2): Analysis of Listener Preference". In: *Journal of the Audio Engineering Society* 65.3 (2017), pp. 212–225. DOI: 10.17743/jaes.2016.0071.

[5] E. Hahn. "Musical Emotions Evoked by 3D Audio". In: *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science.* 2018. URL: https://www.aes.org/e-lib/browse.cfm?elib=19640.

[6] *ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems.* International Telecommunications Union, 2015.

[7] *ITU-R BS.2051-2: Advanced sound system for programme production.* International Telecommunications Union, 2018.

[8] Y. Wycisk, K. Sander, R. Kopiez, F. Platz, S. Preihs, and J. Peissig. "Wrapped into sound: Development of the Immersive Music Experience Inventory (IMEI)". In: *Frontiers in Psychology* 13 (2022). DOI: 10.3389/fpsyg.2022.951161.

[9] R. Kiyan, J. Bergner, S. Preihs, Y. Wycisk, D. Schössow, K. Sander, J. Peissig, and R. Kopiez. "Towards predicting immersion in surround sound music reproduction from sound field features". In: *Acta Acustica* (2023). Under review.

[10] R. Hupke, M. Nophut, S. Li, R. Schlieper, S. Preihs, and J. Peissig. "The immersive media laboratory: Installation of a novel multichannel audio laboratory for immersive media applications". In: *Audio Engineering Society Convention 144.* 2018. URL: https://www.aes.org/e-lib/browse.cfm?elib=19522.

[11] J. Bergner and J. Peissig. "On the identification and assessment of underlying acoustic dimensions of soundscapes". In: *Acta Acustica* 6 (2022). DOI: 10.1051/aacus/2022042.

[12] J. Bergner, D. Schössow, S. Preihs, and J. Peissig. "Identification of Discriminative Acoustic Dimensions in Stereo, Surround and 3D Music Reproduction". In: *Journal of the Audio Engineering Society* (2023). Accepted for publication.

[13] J. Meyer and G. Elko. "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield". In: *IEEE International Conference on Acoustics Speech and Signal Processing.* IEEE, 2002. DOI: 10.1109/icassp.2002.5744968.

[14] C. Schörkhuber, M. Zaunschirm, and R. Höldrich. "Binaural rendering of ambisonic signals via magnitude least squares". In: *Fortschritte der Akustik – DAGA 2018.* 2018, pp. 339–342.

[15] B. Bernschütz. "A spherical far field HRIR/HRTF compilation of the Neumann KU 100". In: *Fortschritte der Akustik – AIA-DAGA 2013.* 2013, pp. 592–595.

[16] *Richard Wagner – Die Walküre. The Metropolitan Opera. Deutsche Grammophon 0734855.* 2011.

[17] *Pyotr Tchaikovsky – Rococo Variations for Cello and Wind Quintet. Orchestra Academy of the Bayerisches Staatsorchester. Hänssler Classic.* 2018.

[18] V. Pulkki. "Spatial sound reproduction with directional audio coding". In: *Journal of the Audio Engineering Society* 55 (2007), pp. 503–516. URL: https://www.aes.org/e-lib/browse.cfm?elib=14170.

[19] J. Blauert. *Spatial hearing: the psychophysics of human sound localization.* MIT press, 1997.

[20] T. Hidaka, T. Okano, and L. Beranek. "Interaural cross correlation (IACC) as a measure of spaciousness and envelopment in concert halls". In: *The Journal of the Acoustical Society of America* 92.4 (1992), pp. 2469–2469. DOI: 10.1121/1.404472.