Audio Engineering Society

Conference Paper 7

Presented at the Conference on
Immersive and Interactive Audio
2019 March 27 – 29, York, UK

Auditory Scenography in Music Production: Case Study Mixing Classical Turkish Music in Higher Order Ambisonics

Can Karadoğan[1] and Thomas Görne[2]

[1]*Istanbul Technical University, Department of Music Technology*
[2]*Hamburg University of Applied Sciences, Department of Media Technology*

Correspondence should be addressed to Can Karadoğan (karadoganc@itu.edu.tr)

ABSTRACT

Music production has always been influenced by and evolved alongside the newest technological standards and listener demands. As the general apparatus for music consumption moves towards headphones, and with the advent of 3D cinema sound and virtual reality (VR) technologies, questions on how to translate musical standards into a three-dimensional listening environment are in debate. This paper discusses the 3D mix aesthetics of Ambisonics beyond 6th order taking a classical Turkish music production as a musical case. An ensemble recording was made in the recording studio of Istanbul Technical University (İTÜ) MİAM. The channels of that session were mixed in the Immersive Audio Lab of Hamburg University of Applied Sciences (HAW), exploring generic ways of spatial music production. The results were rated by means of a survey grading immersive audio parameters.

1 Introduction

For so-called immersive audio formats like Auro-3D, Dolby Atmos, Ambisonics and other channel, object and soundfield based 3D audio standards, music production aesthetics is an open issue, especially when it comes to formats without a predominant direction of perception where the audience can freely orientate in a virtual auditory scene. Examples would include playback via horizontally uniform loudspeaker arrays or head-tracked binaural audio / VR Audio. In these listening situations, a new aesthetics of music production is obviously necessary.

Within a system with fixed spatial orientation like stereo or surround, both the position and the direction of movement of an auditory object are perceived in relation to the fixed direction of sight. One can unambiguously distinguish between front and rear, left and right. Consequently, in the typical production aesthetics of stereo and surround and even "surround with height" 3D Audio formats, the position in front of the listener is regarded as the "main" direction and thus is typically reserved for important musical elements. In contrast, when there is no predominant direction, all horizontal directions become equally meaningful, and production concepts based on a hierarchy of directions no longer apply.

The mixing process in such "immersive" formats might be considered as a kind of auditory scenography, where auditory objects / instruments / voices can occupy virtually any position in three-dimensional space, either static, or dynamic by moving freely, and where the se-

mantics of position and movement are rather dependent on the objects' spatial relations than on their absolute positions.

To investigate the creative and aesthetic options of such a novel scenographic production approach in a rather traditional musical genre, Classical Turkish Music (CTM) was chosen. Apart from some occasional sound engineering publications [1], CTM is a genre with almost no established production guidelines and no strict mix expectations. It therefore fits the explorational purposes of this research topic very well. Another advantage of having CTM as our chosen genre is its heterophonic musical structure that forces us to break the typical arrangement hierarchy for a mix project. In order to emphasize this point, a pop song was also mixed in 3D in parallel to the CTM recordings.

A stereo production for a Classical Turkish Music ensemble was made in Istanbul in the İTÜ Dr. Erol Üçer MİAM recording studio. The recording approach was quite traditional, where the group was performing live together, with accent microphones on each instrument and a main pair in A-B stereo configuration capturing the room (Fig. 1).

Fig. 1: CTM Ensemble in MİAM live room

The genre in focus has many descriptions: The most commonly used term used to label this particular style in the Turkish music market is "Türk Sanat Müziği" which translates as Turkish Art Music. In this paper it is referred to as Classical Turkish Music mostly because of the way it was performed and recorded. The traditional performance lives on and music production today uses mostly the same methods to capture it.

2 Spatial Audio as an Artistic Tool

2.1 Physical, Perceptual and Metaphoric Space

The perceived space can be regarded as as a cognitive construct based on two perceptual aspects: (1) the sonic space spanned by sound sources, perceived as auditory objects, and (2) the architectural space, perceived through reflections of boundaries and rigid objects.

Beyond this explicit spatial perception caused by well-known physical spatial cues, auditory perception has a metaphoric spatial dimension as well, expressed in linguistic metaphors like...

- pitch height,
- melodic movement,
- auditory object size or volume and
- auditory object shape (sharp / round / smooth etc.).

These metaphoric spatial properties of sound are manifestations of the *crossmodal correspondences* of perception[1], partly described in the late 19th century by philosopher Carl Stumpf as "spatial symbolism of tones" [3], in-depth investigated mainly since the mid 1980's [4, 5, 6, 7, 2]: A high pitched sound corresponds to a small, bright, sharp-edged object with a position high in space, while a low pitched sound corresponds to a large, dark, round object low in space.

The correspondence of pitch and perceived position in space, reflected in the metaphor of the "height" of a sound, has first been investigated by Pratt in 1930: *"The results are clear-cut and unequivocal. High tones are phenomenologically higher in space than low ones"* [8]. This "pitch-height" effect, also referred to as "SMARC" effect (spatial–musical association of response codes) [9] is clearly the most prominent and important crossmodal effect in audio production and therefore crucial for an aesthetic approach to spatial audio.

The impact of the pitch-height effect on phantom image perception and its relevance and applicability for spatial audio production has recently been investigated by Lee [10, 11]. One remarkable finding is the listener's preference to a congruent representation of physical and metaphoric height in a 2D to 3D audio upmix in comparison to a "proper" 3D recording [10].

[1] Terminology according to Spence [2]; by different authors also referred to as synaesthetic correspondences / associations or crossmodal equivalences / similarities / mappings.

In summary, both the vertical position (elevation) and the size of the auditory object have a physical as well as a metaphoric dimension, referring to directional cues and spectral content respectively. These dimensions may interfere, as physical and metaphorical position of an auditory object in space can either be congruent or incongruent.

2.2 Spatial Audio in Electro-Acoustic Music and Sonic Art

Sonic art and electro-acoustic (acousmatic, electronic) music were the original genres where spatial audio was utilized from the first days the technology was available, though rarely as a distribution format, but rather for live performance. Composers like Schaeffer, Xenakis and Stockhausen introduced live multichannel reproduction in the 1950's, typically with quadraphonic, and later often octaphonic, setups. *Moving* auditory objects have been introduced in the 1950's likewise, one example being a 1958 composition by Pierre Boulez, where sounds from wall-mounted loudspeakers, complementing the orchestra, were rotating in a spiral movement upwards during the performance [12].

From the mid 1970's, the staging of highly complex spatial scenes has been facilitated with "Loudspeaker Orchestras" – best known being the Acousmonium of the Groupe de Recherches Musicales, set up with some 80 loudspeakers of fairly individual acoustic specifics, distributed throughout the performance space.

A rather modern approach to performing spatially complex electro-acoustic music and sonic art is the utilization of High Density Loudspeaker Arrays (HDLA), used either as live instruments for multichannel diffusion in the Loudspeaker Orchestra tradition, or as tools for "sonic imaging" in combination with object or soundfield based coding like Ambisonics [13]. Examples of prominent installations are the Birmingham ElectroAcoustic Sound Theatre (BEAST) at the University of Birmingham with 30+ loudspeakers in various setups, the 43.4 "Klangdom" at ZKM Karlsruhe, or the 58.2 ASPIRe and 134.6 "Cube" at Virgina Tech.

At HAW Hamburg's Immersive Audio Lab (IAL) a HDLA in 5 height layers in a fairly dry studio (RT ≈ 0.29 s) is available (Fig. 2). The IAL is designed for Higher Order Ambisonics (HOA) but capable of rendering other formats [14].

Fig. 2: Immersive Audio Lab (IAL) at HAW Hamburg

2.3 Auditory Scenography

In this project we follow an approach to producing music in spatial audio inspired by the aesthetics of sonic art and electro-acoustic music, where space is often regarded as an inseparable element of the composition. Naturally, this leads to a different aesthetics than the common practice in stereo, surround and even 3D audio production, where a typical aesthetic goal is creating a virtual stage, populated with the instruments and voices and surrounded by a proper architectural space.

In contrast, an *auditory scenograpy* regards space literally as performance-space. And advanced spatial audio tools like HOA facilitate realism beyond sheer listener envelopment, *"breaking the paradox between reality and fiction"*, as Barrett points out [15].

Movement as an important element of auditory scenography might be motivated with different concepts, i.e.:

- the "dance concept", following the idea of music as a main impulse of bodily movement, described e.g. by Ihde: *"To listen is to be dramatically engaged in a bodily listening that 'participates' in the movement of the music."* [16],
- the "metaphor concept" referring to the crossmodal correspondences of auditory perception and accordingly the crossmodal metaphors of the music, mainly pitch height and size.

The former concept aims to translate the *"'internal' dance of rhythms and movements felt bodily while quietly listening"* (Ihde [16]) into the physical movement

of the auditory objects. A similar "choreographic" approach to spatial audio production has recently been investigated, employing terms and concepts of dance in a spatial panning tool [17].

The latter concept aims to translate the metaphors of space and movement into physical movement. Both concepts might overlap within the lively movement of a melodic line.

3 Aesthetic Challenges and Constraints

The artistic aim of conventional music production may be similar to the avant-garde electro-acoustic music works; however, the complexity of the audience's expectations in any given genre makes these artistic goals sometimes difficult to achieve.

Nevertheless, the exploration of spatial audio as a tool for production is advisable even in rather conservative musical genres, as more and more listeners have access to immersive technology. At least one can expect an intensified listener's experience [18].

Since *Fantasia* (1940) of Walt Disney, the idea of moving instruments over the panorama has been appealing for many artists, given that there is a visual image to relate to.

In a music production where the job is to complete a stereo mix, the movement of certain sources might span the whole "stereo stage" but this technique is sometimes – e.g. in case of a classical production – not customary practice at all. We might perceive a three-dimensional scene but our vantage point "window" is limited. Thus the movement of an auditory object in a stereo mix might be rather small in absolute angular displacement, but quite large with regard to the stereophonic scene, resulting in an obtrusive and distracting sound effect.

On the contrary, within a fully immersive spatial audio environment, we expect the movement to be an unobtrusive yet effective tool for taking advantage of the above-mentioned concepts of auditory scenography.

Another aesthetic issue of a spatial mix might arise with the perceived complexity of the auditory scene: When we analyse the arrangement elements of a generic pop production, instruments can be grouped by their functions [19], and will most likely be perceived in a hierarchical structure. But when we compare a string quartet to a pop band, the perception of musical elements changes in a strange way: We are stuck between perceiving a single group of strings and four string instruments individually. The upper limit of simultaneously perceivable objects becomes relevant here: Following Miller's famous investigation of conscious perception, one can expect blending / merging effects when more than 7 ± 2 sources are presented in the scene [20], rather less when the objects are similar, as experience in film sound design shows [21]. We expect this effect to be highly dependent on the space occupied with the auditory objects, as the closeness of auditory objects facilitates their coalescence.

4 Technical Realisation

The production format was chosen to be Higher Order Ambisonics, as this provides in practice a quite large "sweet area" in the production studio if one can dispense with the advantages of accurate soundfield reproduction [22, 14], and moreover allows to transfer the mix to different technical environments including binaural playback. The loudspeaker array at the IAL supports HOA beyond 6th order in extended half-space (elevation $-15°...90°$, see Fig. 3).

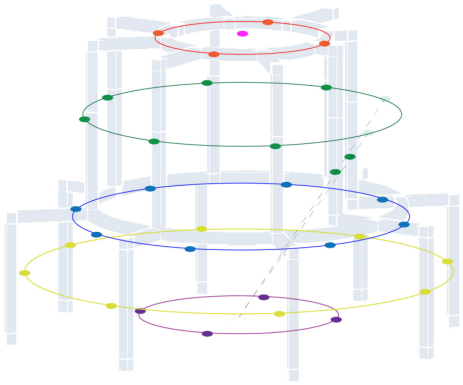

Fig. 3: Loudspeaker dome / HDLA layout in the Immersive Audio Lab; dots indicate loudspeaker positions. Second loudspeaker ring at ear level.

As a design environment, we used Cockos Reaper DAW and Ambisonics plug-ins developed by Zotter et al. and Rudrich at IEM (Institute of Electronic Music and Acoustics) Graz [23, 24]. The AllRADecoder was configured with the IAL speaker positions, and the IEM StereoEncoder was our primary plug-in to pan channels and write automation lines for spatial trajectories.

The transition from a rather conventional recording towards a fully immersive 3D mix had to deal with the question: Which channels of the original live studio session can be used in a HOA project in order to obtain a decent localization and allow for spatial panning?

After first attempts of using the main A-B stereo pair of the ensemble, we began generating mix models using only the accent microphones of the recording, for the simple reason that they were mono tracks resembling the instrument in a better way.

Consequently, "room" or perceived architectural space was created in the Ambisonics domain, using a HOA algorithmic reverb plug-in.

5 Results

A classical Turkish music session was made in the typical music production ensemble recording way: The main A-B stereo microphone was in the middle of the hemisphere of 5 instruments and each instrument had its close microphone. The ensemble consisted of following instruments: tanbur, kemençe, kanun, oud and cello. The players did not use headphone monitoring.

The repertoire of the ensemble for this session consisted of 18th and 19th Century compositions. Besides these compositions, the performers played free improvisations (taksim) in certain musical microtonal scales (makams) of related pieces. Most of these taksim's were played solo, but there were also one with two instruments and another one with three. Five instruments performed the remaining tracks together. There were no overdubs.

As a comparison, another piece from the same ensemble but from a previous session with percussion instruments were also brought in to the mix, in order to compare the complexity of the panning problematic and spatial rhythm perception. This recording included the following instruments: tanbur, kemençe, kanun, oud, cello, ney, bendir and daire.

In order to have a reference for arrangement hierarchy, a pop rock project with the following instrumentation was also mixed in 3D: drums, bass, piano, accordion, acoustic guitar, electric rhythm guitar, electric lead guitar, reverse effects and lead vocals.

In order to understand the complexity, all mixes were prepared in parallel and the impressions and observations were noted in the lab journal. The following is a list of every project and a description of its mix approach.

5.1 Solo Instrumental Mix:

The oud plays a *taksim*, an improvisation in the *makam* (modal scale) of *acemaşiran*. This mix was approached as if a single actor appeared on stage and performed his lines. We tried to set the tempo of panoramic moves quickly at first and but then slowed them down. The height changed in relationship with the pitches of the scale. Our hypothesis was that because there is only a single source present, the listener wants to hear the performer from the front, hence, no overly-exaggerated rotational gestures. A reverb plug-in was used for spatialization but no other effect was inserted.

5.2 Duo Instrumental Mix:

The tanbur and kanun play an improvisation in the *makam* of *hüzzam*. The interaction between the two instruments immediately initialized the idea of dance choreography in a room. The musical motion created by the ascending notes were underlined with the idea of the instruments taking off and climbing up the dome in spirals to the top speaker (VoG) and then similarly, as the melodic line descends down to the cadence (karar perdesi) the spatial motion would be quasi downwards to ear level. With this mix, it was clear that mono spot microphones work better than the main pair for rotational gestures [15]. The idea of "meeting-points" was applied as well, as both instruments finish their sentences in the same point in space. Our expectation was that the survey subjects would immediately feel the room created by the dance of these two instruments.

5.3 Trio Instrumental Mix:

Tanbur, kemençe and cello play an improvisation in the *makam* of *nihavend*. The cello sets on a drone tone, keeping the fundamental pitch (*karar*) and the pause pitch (*durak*) of the *makam* to give the tanbur and the kemençe a reference for their free improvisation. The cello track is rotating at a very slow pace all around the space. Being a low register instrument the rotation is not easily traceable by the listener. This results in the effect that the cello creates volume based fluctuations in the room. The interaction between tanbur and the kemençe is again resembling a conversation, in other words, a dialogue with short musical sentences in a question-answer fashion. Similar to the duo description, they "dance around each other". Long spatial "jumps" are used for the single overtones of the tanbur as the kemençe plays its sentence. Long sustained

notes of the kemençe are underlined by small angular panoramic moves. Because there are only 3 instruments, one drone accompaniment and two soloists, it does not feel "uncanny" when they sometimes play behind our heads, hence, in the rear position. When the improvisation ends, both instruments finish their musical motions at the VOG speaker position.

5.4 Ensemble of Five Mix:

Tanbur, kemençe, kanun, oud and cello play an instrumental composition in the *makam* of *acemaşiran*. The heterophonic character of classical Turkish music becomes very obvious in the performance: Most of the time all five instruments play the same melodic line together and when they complete musical sections a verse or a chorus (*hane* or *teslim*), some of the instruments stop playing the line and fill in with an ornamentation. This creates a strong potential for a background-foreground placement.

In order to have some organized panning constellations, the mix needed to be planned scene for scene, similar to any time-based visual format such as dance choreography or film production. The metaphorical choreography has the overall impression of being something between a 2D mix and a 2D enhanced with a third rotational spatial dimension. The first reason for this type of design is the need to avoid spatial chaos which would make the perception of the instruments nearly impossible. The second reason is that our head needs to be pointed to the front in order to get used to the interplay.

We started the mix by fixing certain instruments in static points within the front hemisphere. This presents a first picture of the ensemble as they sit in the studio recording session. The cello, oud and the tanbur were fixed to ear level left, front and right, respectively, where the cello, being a low register instrument, was kept mostly beneath ear level. The remaining two high register instruments kemençe and kanun were rotating mostly at about 45° elevation both for their ascending lines and their ornamentations.

After the fermata at 3:28, the legato passage begins where the kanun and cello are left alone. They play a staccato line where the cello employs the pizzicato technique. The kanun is panned step by step on every single note which is a perceivable move probably because there are only two instruments present. After this section the so far static tanbur and the oud begin their rotational moves, and at 3:44 all instruments start moving around the back hemisphere. The mix concludes with a unison phrase where they don't come back to their origin points.

5.5 Ensemble of Eight Mix:

Tanbur, oud, ney, kanun, kemençe, cello, bendir and daire play an instrumental composition in the *makam* of *buselik*. Comparing with the ensemble of five, three more instrument channels are added to the mix: two percussion instruments bendir and daire and also the wind instrument ney. As the number of instruments goes up, so too does the complexity of spatial placement, panning and perception.

The first spatial version of this piece did not have any frontal perspective at all: All instruments were arriving at the ears from all directions, which gave an uncanny effect. Therefore, a second mix was designed to have a quasi "stage placement" at the beginning. Just like the previous mix, most of the instruments were located in the frontal hemisphere: The percussion instruments fixed to the left, cello to the front-left bottom, plucked instruments tanbur and oud to the front-right. The high register instruments ney, kanun and kemençe were freely moving, mostly underlining their ornamentations (long trills of the kanun moving over the VoG speaker back and forth; long sustained blown notes of the ney, rotating very fast, creating a pad-like texture and getting lost into the background).

Because of the instrumental *semai* form (ABAB) the sections mostly repeat themselves. In order to break that routine, fixed instruments were also relocated at the change of each section. For example, percussion instruments appear rotated about 45 degrees to the right, remaining in the frontal hemisphere. Similarly, the plucked group and the cello is rotated so that the ensemble has varying static perspectives to follow.

5.6 Comparative Pop-Rock Mix:

This comparative pop-rock project uses drums, bass, piano, accordion, acoustic guitar, electric rhythm guitar, electric lead guitar, reverse effects and lead vocals in its arrangement. The main difference to a CTM project is that the arrangement introduces the instruments step by step and they all have clear musical roles as expected in a typical music production: Lead, Fills, Pad, Rhythm and Foundation [19]. Every instrument was mixed

down to mono tracks, none of them was used as a stereo source in the 3D space. Again, this was done to improve the localization so that designing scenes would be easier.

As the song had typical arrangement hierarchy, it made the decision process for static and dynamic elements easier: The foundation instruments drums and bass had their fixed positions throughout the song, again spread out in the frontal hemisphere, ear level left and bottom right, respectively. The song begins with the accompanying electric guitar that first appears as a fixed source at 30° front-right but then begins to slowly move within an angular span of about 45° when the lead vocal begins from an symmetrically juxtaposed point, namely ear level front-left. At that point the listener would realize the priority that is given to the lead vocal. The lead vocal completes the verse at the origin point and is then very slowly panned towards the front, where it finishes the chorus. Then it reappears in the ear-level front right position when the second verse begins. This technique might be regarded as a 'sectional placement' where the panning steps only happen when there is a section change. The singer is singing while moving across the stage, always remaining in the foreground.

Finally, in the last chorus the voice appears in its original position, this time left alone with the acoustic guitar panned exactly to the same point. This is to give the impression that the singer is accompanying himself with his guitar. The accordion, on the other hand is rotating freely: It plays sustained notes in a fill function answering to the lead vocals lines. The instrument appears first in the back position, then moves around the listener, mostly at ear-level.

As soon as instruments are panned to the VoG speaker they tend to fall into the background. A similar effect happens when they appear in the back, as sources behind the listener are not easy to localize and recognize at the first. Similarly, auditory objects spinning around at a fast angular speed cannot be located at all and tend to disappear as well. This technique was applied to the reversed electrical guitar effect signals in the bridge part of the song: Rotating quickly, they created the ambiguous effect of a pad like texture in 3D space.

5.7 The Stereo Mix

All above mentioned examples were originally produced for 2.0 stereo. So in order to understand the differences, the stereo mixes were prepared and kept ready for comparison.

5.8 The Static Mix

To evaluate the panning effects, all above-mentioned examples were mixed with static sources in 3D as well. However, the static mixes created a different enveloping impression than the stereo versions.

5.9 Some Findings of the Production Process

In accordance with some professional producers[2], we agree that mixing in spatial audio doesn't have much in common with the approach that is used for a stereo production; traditional aesthetic rules often don't apply.

When developing aesthetics of music production for spatial / immersive audio, we suggest considering the following points:

- The static mix in spatial audio does not create the hierarchic effect of a stereo or even surround mix. Music that completely immerses one's body needs to be further spatialized in some way, i.e. using a good reverb to enhance the spatial feeling.
- The limits of cognition in perceiving complex scenes lead to a basically different perception of a fully immersive auditory scene compared with a stereo or surround mix, or even "surround with height" when used in the conservative way (i.e. virtual stage plus room).
- Likewise, due to loss of the predominant direction and due to the listener's rather active role in perceiving an immersive complex scene by shifting her or his focus of attention, the typical mixing approach of guiding one's attention through leveling and panning must be rethought.
- Depending on the musical style, the metaphoric content of different musical elements can be converted into position and movement, much easier and more convincing than in stereo or surround or "surround with height".
- Consequently, the panning styles may be categorized as static or dynamic, and dynamic panning might result in an enhanced spatial effect and enhanced impact through the "physical rendering" of metaphoric musical content, much different from the effect resulting from dynamic panning e.g. in stereo.

This is what we suggest to call auditory scenography.

[2] e.g. Martyn Ware [British Electric Foundation / Heaven 17 / The Human League / Illustrious Co. with Vince Clarke], personal communication (2017)

6 Listening test

In order to evaluate the hypothesis that a musical genre like CTM, usually produced in a classical style, can benefit from a fully immersive scenographic approach, we performed a short survey.

34 subjects participated in this test. The subjects were mostly undergraduate and graduate students experienced in audio engineering. The 20-minute long test was designed as an assessment of 6 different musical excerpts (Solo, Duo, Ens. of 3, Ens. of 5, Ens. of 8, Pop) in 3 versions each, i.e. the original spatial HOA mix with moving sources, a static spatial HOA mix, and the stereo version. Thus in total 18 excerpts from the productions were assessed in randomized order, with a length of some 1 to 2 minutes each. The venue was the production studio, i.e. HAW Immersive Audio Lab. Not more than 6 subjects were simultaneously present in the studio. The subjects were situated in the perceptual sweet area of the studio [22, 14], but outside the rather small accurate-reproduction area.

The questionnaire evaluated the following perceptual parameters, performed as direct rating on a metric scale ranging from 1 (lowest value) to 9 (highest):

1. spaciousness (Räumlichkeit)
2. envelopment (Umhüllung)
3. definition of the instruments (Klarheit)
4. genre congruence or appropriateness (Angemessenheit)
5. preference (Vorzug)

Parameters 1 and 2 refer to the technical immersive effect or "3D-ishness", parameter 3 to the perceived complexity of the auditory scenes. Parameters 4 and 5 are aimed at the production aesthetics, as an impressive spatial mix nevertheless might be considered as flashy and inappropriate. All parameters were clarified to the subjects before the test. Below are some findings.

6.1 The Number of Instruments and Perception

Results of the survey show that there is a noticeable change in the audience perception of immersive audio when listening to mixes with varying numbers of instruments. When we look at the scores for spaciousness and envelopment-which are both descriptive for immersive audio, the audience is not that satisfied with a moving single instrument in space (Table 6). On the contrary, for the CTM mixes with 3, 5 and 8 instruments these values become higher and with a smaller standard deviation (Table 3).

Additionally, the preference average of the dynamic mix for the trio is higher than both of the stereo and the static 3D mix. As Table 4 shows, the stereo mix is rated as having a high definition of instruments but unexpectedly not as genre congruent as the 3D versions. The immersive factors are low as expected.

6.2 Movement and Space

As Table 1 and Table 4 indicate, stereo mixes do not create enough spaciousness and envelopment compared to 3D. Furthermore, thinking about room perception, moving sources, hence dynamic 3D mixes get higher spaciousness ratings in comparison to static 3D mixes (Tables 2 and 3). This can also be seen when we comparatively analyze the results of Tables 5 and 6: The dynamic versions show higher averages than the static versions for spaciousness and envelopment.

6.3 Definition in Immersive Audio

As the scores in results show that the definition value of mixes with 1 or 2 instruments are higher than the ones with 3 to 8 instruments. However, definition is not really a quality that the audience is looking for when listening to 3D audio. This can be supported by the results as Table 6 shows: When there are less instruments the definition increases but it does not directly mean that the preference increases as well. Even for the stereo versions of the 1 or 2 instrument mixes the definition average is high but all other criteria deliver lower ratings.

	N	Mean	Std Dev
spaciousness	102	4.47	2.008
envelopment	102	3.54	1.988
definition of instruments	102	5.25	1.992
genre congruency	101	5.38	1.666
preference	102	4.49	1.893
valid cases	101		

Table 1: CTM 3...8 instruments, stereo mixes

	N	Mean	Std Dev
spaciousness	102	6.62	1.635
envelopment	102	6.60	1.780
definition of instruments	102	6.92	1.264
genre congruency	102	6.72	1.181
preference	100	6.59	1.545
valid cases	100		

Table 2: CTM 3...8 instruments, static spatial mixes

	N	Mean	Std Dev
spaciousness	102	7.05	1.492
envelopment	102	7.57	1.182
definition of instruments	102	6.67	1.444
genre congruency	102	6.46	1.507
preference	102	6.65	1.663
valid cases	102		

Table 3: CTM 3...8 instruments, dynamic spatial mixes

	N	Mean	Std Dev
spaciousness	68	4.56	2.083
envelopment	68	3.74	2.176
definition of instruments	68	7.07	1.696
genre congruency	68	5.79	1.626
preference	65	5.09	1.902
valid cases	65		

Table 4: CTM 1...2 instruments, stereo mixes

	N	Mean	Std Dev
spaciousness	68	6.01	1.749
envelopment	67	5.19	2.141
definition of instruments	68	7.12	1.662
genre congruency	68	6.28	1.610
preference	65	6.03	1.854
valid cases	64		

Table 5: CTM 1...2 instruments, static spatial mixes

	N	Mean	Std Dev
spaciousness	68	6.78	1.303
envelopment	67	6.58	1.662
definition of instruments	67	7.03	1.537
genre congruency	68	5.87	1.495
preference	68	6.06	1.884
valid cases	66		

Table 6: CTM 1...2 instruments, dynamic spatial mixes

7 Summary

For a genre like CTM in fully immersive coding and playback, we propose dynamic panning, or moving objects in the scenes, as this mixing approach provides the listener not just with an intensified spatial experience, but also is considered highly appropriate for the music. This kind of dynamic spatialization, quite common with electro-acoustic / acousmatic music or sonic art, seems to be a promising approach also for traditional musical genres where one might expect rather conservative production aesthetics.

With a higher number of instruments the spatial effects of such a "scenographic" approach are more unobtrusive yet seemingly still effective, as movement appears to be less obvious in complex scenes.

In creating trajectories one might follow either the crossmodal metaphor of melodic movement, translating this into physical movement, or the "dance" metaphor, inspired by the inherent movement of rhythms and melodies. Particularly promising is the combination of both: letting the auditory objects of instruments or voices freely move in space, dancing to the rhythms, but according to and congruent with the crossmodal metaphor of height.

Acknowledgement

A special thanks goes to Refik Hakan Talu and his ensemble for supporting this project with their beautiful performances and recording session tracks, and to Oğuz Öz for helping with the recording. Thanks likewise to Anette Stobert for reliable and unfailing technical assistance in the IAL.

This work was performed during a research exchange supported by Istanbul Technical University (İTÜ) and Hamburg University of Applied Sciences (HAW).

References

[1] Karadogan, C. and Sarier, O., "Recording of 18th Century Classical Turkish Music Works," Proceedings VDT International Convention / 29. Tonmeistertagung, Cologne, 2016.

[2] Spence, C., "Crossmodal correspondences: a tutorial review," *Attention Perception & Psychophysics*, 73(4), 2011.

[3] Volke, S., "Carl Stumpf und die 'Raumsymbolik der Töne'," in G. Rötter and M. Ebeling, editors, *Hören und Fühlen*, pp. 55–69, Carl Stumpf Gesellschaft, Peter Lang Verlag, 2012.

[4] Marks, L. E., "On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination," *Journ. Exp. Psychol.: Human Perception and Performance*, 13(3), 1987.

[5] Marks, L. E., "On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness," *Journ. Exp. Psychol.: Human Perception and Performance*, 15(3), 1989.

[6] Gallace, A. and Spence, C., "Multisensory synesthetic interactions in the speeded classification of visual size," *Perception & Psychophys.*, 68, 2006.

[7] Evans, K. K. and Treisman, A., "Natural cross-modal mappings between visual and auditory features," *Journ. Vision*, 10(1), 2010.

[8] Pratt, C. C., "The spatial character of high and low tones," *Journ. Experimental Psychology*, 13, pp. 278–285, 1930.

[9] Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., and Butterworth, B., "Spatial representation of pitch height: the SMARC effect," *Cognition*, 99(2), 2005.

[10] Lee, H., "2D-to-3D Ambience Upmixing Based on Perceptual Band Allocation," *Journ. Audio Eng. Soc.*, 63(10), 2015.

[11] Lee, H., "Perceptual Band Allocation (PBA) for the Rendering of Vertical Image Spread with a Vertical 2D Loudspeaker Array," *Journ. Audio Eng. Soc.*, 64(12), 2016.

[12] Voit, J., *Klingende Raumkunst. Imaginäre, reale und virtuelle Räumlichkeit in der Neuen Musik nach 1950*, chapter 2.2.2, p. 202, Tectum, 2014.

[13] Reese, K., "Mediales Komponieren im Raum. Konstellationen und Wirkungen," *Auditive Perspektiven: Artist's Perspectives on Sound in Space,* http://kunsttexte.de, 2/2017-1, 2017.

[14] Kessling, P. and Görne, T., "Studio for immersive media research and production: Immersive Audio Lab at HAW Hamburg," e-Brief EB453, AES 145th Convention, New York, 2018.

[15] Barrett, N., "Ambisonics and acousmatic space: a composer's framework for investigating spatial ontology," Proc. 6th Electroacoustic Music Studies Network Conference, Shanghai, 2010.

[16] Ihde, D., *Listening and Voice. Phenomenologies of Sound*, chapter 13, pp. 155–156, State University of New York Press, 2nd edition, 2014.

[17] Erruz, G. and Schmele, T., "Sound-Space Choreography," Conference Paper, AES Conference on Spatial Reproduction – Aesthetics and Science, Tokyo, 2018.

[18] Hahn, E., "Musical Emotions Evoked by 3D Audio," Conference Paper, AES Conference on Spatial Reproduction – Aesthetics and Science, Tokyo, 2018.

[19] Owsinski, B., *The Mixing Engineer's Handbook*, Alfred Music, 2nd edition, 2006.

[20] Miller, G. A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review*, 101(2), 1955.

[21] Görne, T., *Sounddesign: Klang, Wahrnehmung, Emotion*, Hanser, 2017.

[22] Frank, M. and Zotter, F., "Exploring the perceptual sweet area in Ambisonics," AES 142nd Convention, Berlin, 2017.

[23] Frank, M., Zotter, F., and Sontacchi, A., "Producing 3D Audio in Ambisonics," AES 57th Int. Conference, Hollywood, 2015.

[24] Rudrich, D., "Introducing the IEM Plug-in Suite," 'klingt gut' Intern. Symposium on Sound, Hamburg, 2018.