DOI: 10.22032/dbt.39946
DOI (proceedings): 10.22032/dbt.39936

Abstract Reviewed Paper at ICSA 2019

Presented* by VDT.

Capturing 3D Audio: A pilot study on the spatial and timbral auditory perception of 3D recordings using main-array and front-rear separation in diffuse field conditions

J. Grab[1]

[1] *SAE Institute London, United Kingdom, Email: janetmariagrab@gmail.com*

Abstract

This research is a preliminary pilot experiment into the subjectively perceived differences between the recordings resulting from three different 3D microphone arrays: A Bowles-Array (Main-Array) with a vertical coincident height-channel microphone layer, a Fukada-Tree/Hamasaki-Cube configuration (F/R-Array) and a Hybrid-Array containing the signals from the Bowles-Array main layer and the Hamasaki-Cube height layer. It was hypothesised that the arrays in concern will produce recordings that shall lead each to an increased perception of specific attributes for all sources tested (cello, violin, handpan, djembe, guitar). In order to detect possible patterns in spatial and timbral auditory perception subjective listening tests included direct scale magnitude estimations for the attributes Naturalness, Presence, Preference, Width, Localisation Accuracy, Distance/Depth, Envelopment, Spatial Balance, Room Perception, Vertical Image Shift, Vertical Image Spread, and Vertical Frequency Separation and category scaling for the assessment of timbral attributes. Results suggest that none of the arrays in concern conveyed an increased perception of any of the attributes for all sources, which disproves the hypothesis and indicates a source-dependent performance. Simultaneously patterns in the subject responses have been detected which could be explained through psychoacoustic findings focussing on the correlation of perception between the attributes in question. Furthermore, by trying to explain the obtained differences in auditory perception between the different arrays, some assumptions could be made upon what components of which array could have contributed to a specific perception. These findings could serve as a reference for future experiments in the fields of 3D recording techniques or psychoacoustics.

1. Introduction

The current investigation entails a comparison between a main array technique being a Bowles-Array (Main-Array) and a system with front-rear separation being a Hamasaki-Cube/Fukada-Tree configuration (F/R-Array). For the sake of experimentation, a hybrid version containing the signals of the Bowles-Array for the main layer and the signals of the Hamasaki-Cube for the height layer was included. The selection of these arrays was based on previous research [1-15], which at the same time was the foundation of certain assumptions on their behaviour for specific attributes:

The Bowles-Array was expected to convey an enhanced perception of Naturalness, Presence and Vertical Image Spread (VIS), a preference regarding Timbre, an enhanced risk of a Vertical Phantom Image Shift, and a reduced perception of Spatial Balance and Distance/Depth. The Fukada-Tree/Hamasaki-Cube, on the other hand, was thought to favour an enhanced perception of Envelopment, Width, Spatial Balance and Room Perception, a reduced possibility of a VIS, a reduced risk of a Vertical Phantom Image Shift, and a more stable Localisation Accuracy. In addition, the Hybrid-Array was assumed to show a reduced risk of a Vertical Phantom Image Shift, a reduced possibility of a VIS, and a limited perception of Distance/Depth. Besides, there was a possibility of tonal colouration for all three arrays although the nature of these colourations was unknown by the time to the authors' knowledge [10, 12, 14].

2. Methodology

2.1. Preliminary Recording Session

Prior to the recording session, a test session was organised in Bankstock Studios to ensure that the chosen angle between the main and height layer of the Main-Array would achieve an ICLD (interchannel level difference) of at least -9.5dB. This was found to be the localisation threshold for an ICTD (interchannel time difference) of 0ms and thus corresponds to the vertical coincident microphone placement of the setup in question [14]. Although the acoustic properties of the recording space are different, the acoustically controlled environment would at least allow for indications whether such an ICLD could be achieved with the desired microphone angles and directivities. The result of the test session suggested that an ICLD of -9.5dB would be realistic.

2.2. Recording Procedure

The recording took place at All Saints East Finchley, London. For the recording, all microphone signals were routed to three RME Octa Mic II preamps, an Antelope Orion 32 interface, and recorded into Pro Tools in PCM wave format at 96kHz/24bit resolution. To allow for consistency throughout the comparison of the different techniques a modular system was the preferred choice as it allows for different polar patterns while maintaining the same preamps (Schoeps Colette Series, CMC6 preamps, capsules, KA40 diffraction attachments). As having used the same microphone model for all channels, matching the microphone sensitivity was considered achieved apart from slight sensitivity differences caused by the different capsules. The polar patterns, however, were considered an experimental constant. During the recording, the goal was to match the input gain on all channels to minimise the differences of possible colouration amongst the channels and to maintain the natural level relationships between the channels. However, when strictly adhering to this the SNR (signal-to-noise ratio) of the Hamasaki-Cube would exceed acceptable limits. Therefore, the gain applied to the channels of the Hamasaki-Cube was matched within its low and height layer but was higher than the gain of other channels.

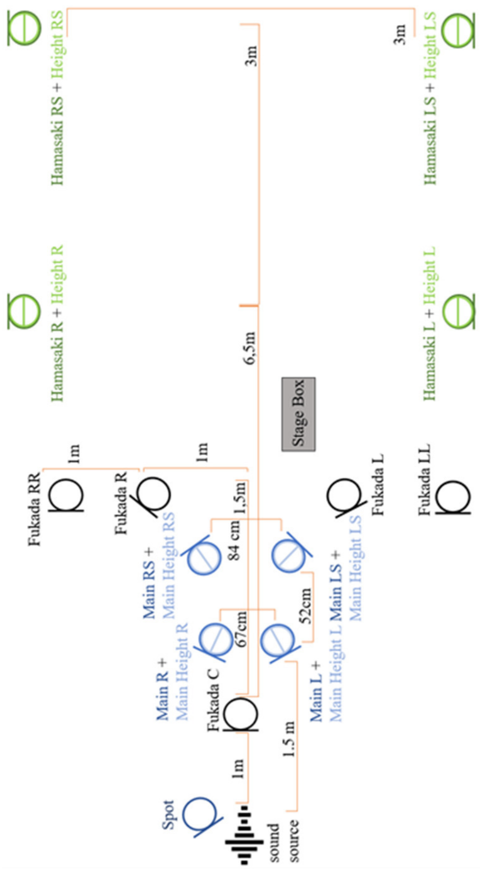


Fig. 1: Floor plan, top view.

Fig. 2: Recording setup, top view.

Position	Capsule on Schoeps CMC 6	Axis	Height
Main L	MK2H	30° outwards, 45° to source	2m
Main R	MK2H	30° outwards, 45° to source	2m
Main LS	MK2H with KA40	30° outwards, 45° to floor	2m
Main RS	MK2H with KA40	30° outwards, 45° to floor	2m
Main Height L	MK4	110° from main upwards	2m
Main Height R	MK4	110° from main upwards	2m
Main Height LS	MK4	110° from main upwards	2m
Main Height RS	MK4	110° from main upwards	2m
Fukada L	MK4	45° outwards	1.8m
Fukada C	MK4	0°	1.8m
Fukada R	MK4	45° outwards	1.8m
Fukada LL	MK2	0°	1.8m
Fukada RR	MK2	0°	1.8m
Hamasaki L	MK8	Positive lobe 90° outwards	3m
Hamasaki R	MK8	Positive lobe 90° outwards	3m
Hamasaki LS	MK8	Positive lobe 90° outwards	3m
Hamasaki RS	MK8	Positive lobe 90° outwards	3m
Hamasaki Height L	MK41	0° to ceiling	5m
Hamasaki Height R	MK41	0° to ceiling	5m
Hamasaki Height LS	MK41	0° to ceiling	5m
Hamasaki Height RS	MK41	0° to ceiling	5m
Spot	MK4	source-dependent	-

Tab. 1: List of microphones.

Fig. 3: Recording setup, side view.

2.3. Listening Test Design

The attributes have been selected and defined based on [8, 12, 16-20]. The chosen response format for the evaluation of the spatial attributes was a direct scale magnitude estimation where the subject assigns a (numerical) value to one stimulus and then judges subsequent stimuli against the first. Timbral attributes, however, have been graded with category scaling, where the subject is asked to assign a category (in this case a timbral or dynamic label) to each stimulus presented.

Double-blind multiple stimuli comparison tests were conducted using a GUI with the Huddersfield Universal Listening Test Interface Generator [21]. The subject could freely turn its head if it stayed in the sweet spot. The task was to grade three stimuli against each other on a continuous rating scale. The scale ranged from 0 (labelled "lesser") to 100 (labelled "greater"), whereas the stimulus with the "greatest" attribute impression was taken as a reference of 100 with the other two stimuli being graded accordingly. This procedure was proposed in [22] to reduce scaling bias. The presentation order of both, the stimuli and the trials was randomised to avoid potential biases. The stimuli were synced in playback, meaning the subject could switch between mixes at any point during the playback. For each test, the subject was to complete a total of five trials, each of which contained the stimuli of the different mixes. When rating Preference, the subject was not given any specific subjective qualities or attributes to consider but was advised to write a comment on its decision on a paper.

2.4. Reproduction Setup

All audio playback including mixing took place in the Auro-3D studio at SAE Brussels, an acoustically treated listening room with a 10.1 Auro-3D setup with Sonodyne SM100AK speakers. The sound pressure level of each loudspeaker was measured and calibrated for 79dB SPL at the listening position using pink noise. The monitor level was calibrated and kept constant to -18dBFS throughout the sessions.

2.5. Stimuli Creation

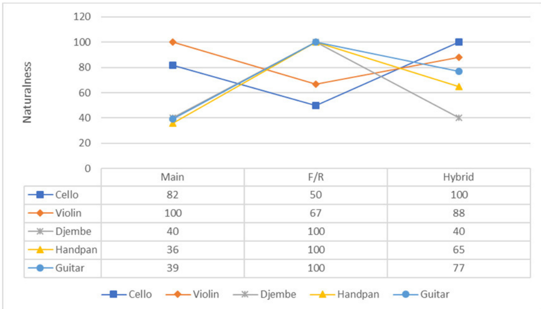
Musical excerpts of 30 seconds were chosen. The selection was mostly aimed towards passages containing pauses of a certain length to ease the perception of room-related attributes. Based on the concept of "ecological validity" [19], a balanced mix was chosen for stimuli creation. Phase relations have been checked aurally and visually, and the polarity was flipped in a few cases where necessary. The stimuli have been mixed in Pro Tools with the Auro-3D Authoring Tools through the Digidesign D-Control (ICON) control surface. Due to the nature of the research questions, no processing was applied. The only exception was a LPF at 250 Hz on the Fukada LL and RR channels, as proposed by Rumsey [23]. All stimuli have been level matched for ± 0.2dB.

2.6. Limitations

Due to limited resources, the process of subject pre- and post-selection could not be accomplished to obtain a listening panel of at least five expert subjects to ensure a sufficient resolution in the test [24]. However, a preliminary listening test was conducted one day before the listening test, and the results of the two tests showed a strong correlation, which indicates intra-subject consistency. Similarly, the validity of the results can be seen reduced as only one balancing engineer was involved, and not several, to minimise the factor of subjectivity in stimuli creation. Also, as the subject was at the same time the experimenter, the subject was familiar with the experimental detail and thus more prone to expectation bias. However, the results throughout the listening tests for the different array-mixes suggest that the magnitude of expectation bias was not dominating the subject's evaluation. Furthermore, the recording was conducted in one space only. Concerning that, the ecological validity of the experiment would be improved when involving ensembles. In addition, based on the recording setup, the results are only applicable to a dry-wet scenario, and not a sound all-around approach.

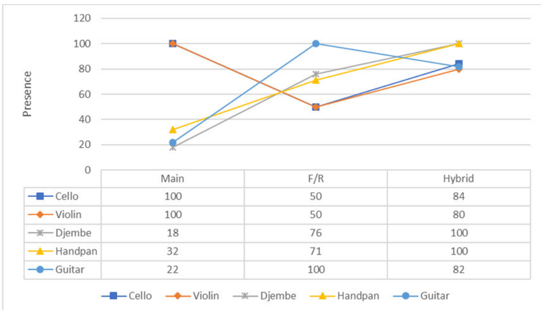
3. Results

3.1. Graphical Representations of the Listening Test Results



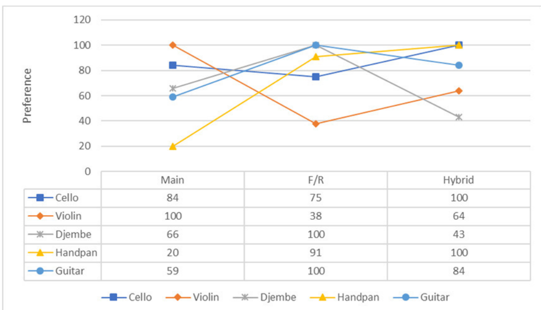
	Main	F/R	Hybrid
Cello	82	50	100
Violin	100	67	88
Djembe	40	100	40
Handpan	36	100	65
Guitar	39	100	77

Tab. 2: Results for the attribute Naturalness.



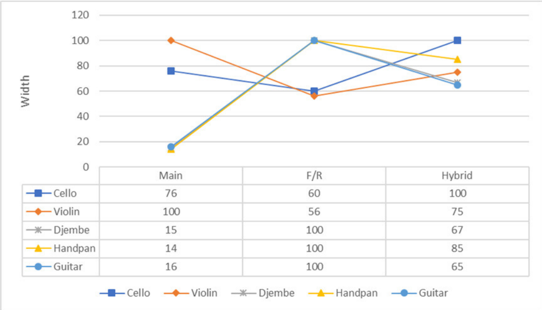
	Main	F/R	Hybrid
Cello	100	50	84
Violin	100	50	80
Djembe	18	76	100
Handpan	32	71	100
Guitar	22	100	82

Tab. 3: Results for the attribute Presence.



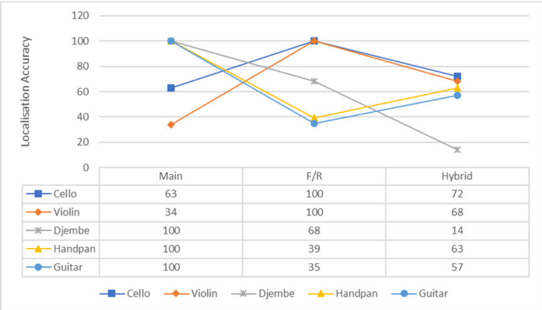
	Main	F/R	Hybrid
Cello	84	75	100
Violin	100	38	64
Djembe	66	100	43
Handpan	20	91	100
Guitar	59	100	84

Tab. 4: Results for the attribute Preference.



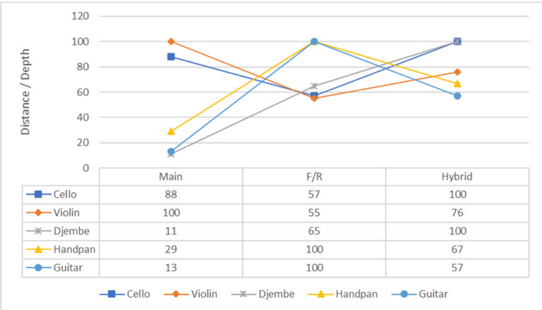
	Main	F/R	Hybrid
Cello	76	60	100
Violin	100	56	75
Djembe	15	100	67
Handpan	14	100	85
Guitar	16	100	65

Tab. 5: Results for the attribute Width.



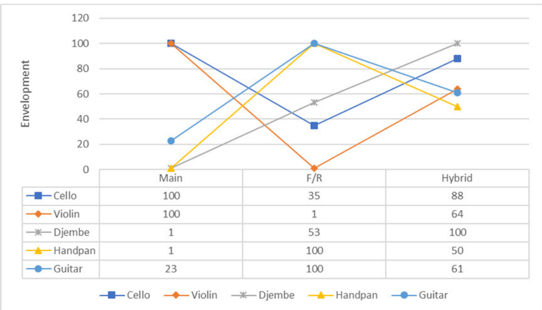
	Main	F/R	Hybrid
Cello	63	100	72
Violin	34	100	68
Djembe	100	68	14
Handpan	100	39	63
Guitar	100	35	57

Tab. 6: Results for the attribute Localisation Accuracy.



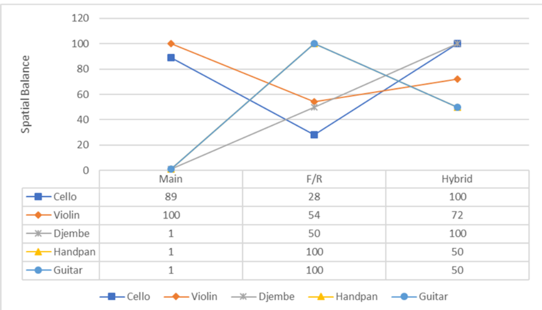
	Main	F/R	Hybrid
Cello	88	57	100
Violin	100	55	76
Djembe	11	65	100
Handpan	29	100	67
Guitar	13	100	57

Tab. 7: Results for the attribute Distance/Depth.



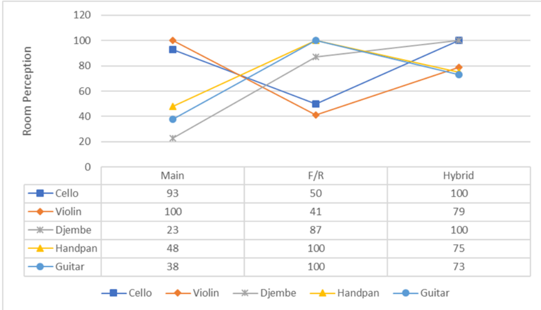
	Main	F/R	Hybrid
Cello	100	35	88
Violin	100	1	64
Djembe	1	53	100
Handpan	1	100	50
Guitar	23	100	61

Tab. 8: Results for the attribute Envelopment.



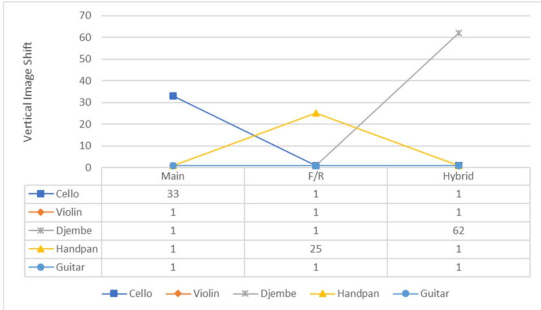
	Main	F/R	Hybrid
Cello	89	28	100
Violin	100	54	72
Djembe	1	50	100
Handpan	1	100	50
Guitar	1	100	50

Tab. 9: Results for the attribute Spatial Balance.



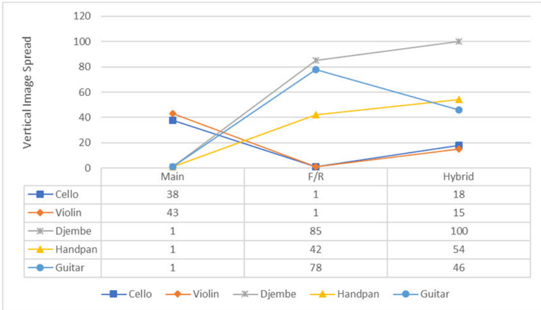
	Main	F/R	Hybrid
Cello	93	50	100
Violin	100	41	79
Djembe	23	87	100
Handpan	48	100	75
Guitar	38	100	73

Tab. 10: Results for the attribute Room Perception.



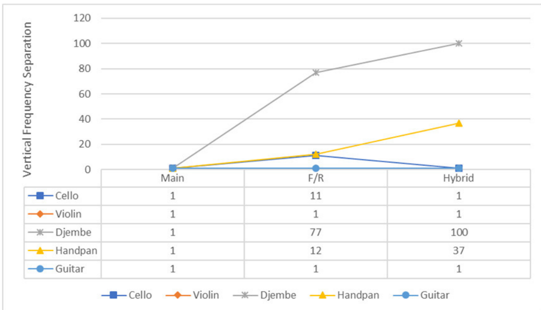
	Main	F/R	Hybrid
Cello	33	1	1
Violin	1	1	1
Djembe	1	1	62
Handpan	1	25	1
Guitar	1	1	1

Tab. 11: Results for the attribute Vertical Image Shift.



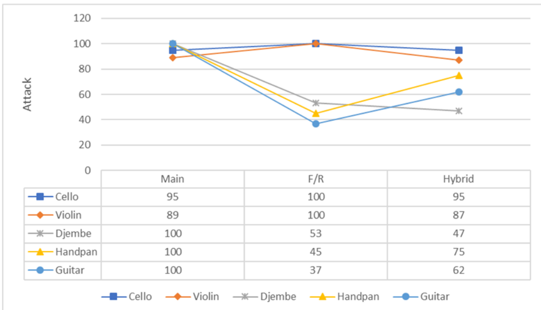
	Main	F/R	Hybrid
Cello	38	1	18
Violin	43	1	15
Djembe	1	85	100
Handpan	1	42	54
Guitar	1	78	46

Tab. 12: Results for the attribute Vertical Image Spread.



	Main	F/R	Hybrid
Cello	1	11	1
Violin	1	1	1
Djembe	1	77	100
Handpan	1	12	37
Guitar	1	1	1

Tab. 13: Results for the attribute Vertical Frequency Separation.



	Main	F/R	Hybrid
Cello	95	100	95
Violin	89	100	87
Djembe	100	53	47
Handpan	100	45	75
Guitar	100	37	62

Tab. 14: Results for the attribute Attack.

Source	Stimulus used for scale A	Stimulus used for scale B	Stimulus used for scale C
Handpan	thin	full, but lacking mids	slightly nasal, less full than B, but more homogeneous, natural, highest treble content in reverb
Cello	full	homogeneous, natural, highest treble content in reverb	(very) thin, nasal
Djembe	homogeneous, natural, highest treble content in reverb	full, but completely lacking mids, excessive bass	canny/nasal, thin
Violin	treble strength neutral	sharp, thin, nasal	brilliance, highest treble content in reverb
Guitar	lacking mids, nasal, thin	homogeneous, brilliant, full	completely lacking midrange, very bright

Tab. 15: Results for the timbral attribute category scaling. The colour blue indicates the mix of the Main-Array, orange the mix of the F/R-Array, and green the mix of the Hybrid-Array.

Source	Stimulus used for scale A	Stimulus used for scale B	Stimulus used for scale C
Guitar	**very open, yet still clear, wide pleasing sound**	very precise but lacking space	compromise between A and B
Violin	**very open sound, good D/R ratio, not sharp**	not as sharp as C but not as wide and opened as A	sharp, edgy, lacking space
Djembe	phase issues not as bad as B, but also not as realistic as C	disturbing frequency, phase issues in the ambient sound with changing frequency	**very realistic**
Cello	dislike timbre	**balanced D/R ratio, further away**	not as wide as B, lows too present, narrowing the picture unnaturally
Handpan	natural but less than C	no space	**natural, real**

Tab. 16: Comments on the grading for the attribute Preference. Comments in bold indicate the comment of the preferred stimulus.

3.2. Correlation between Attributes and Subject Responses

Figure 4 depicts a schematic representation of the correlations found between the assessed attributes based on their response patterns through direct scale magnitude estimation (what array was graded how for what source regarding a specific attribute). Vertically coherent displayed attributes showed an identical response pattern regarding the ranking of the different arrays. Horizontally intended attributes showed an identical response pattern to the attributes above, except for one source. Similarly, the closer the attributes are horizontally to each other, the more similar were their response patterns. The attributes Vertical Frequency Separation and Vertical Image Shift have not been included in this graphic as the obtained values do not allow for a direct comparison with the other attributes.

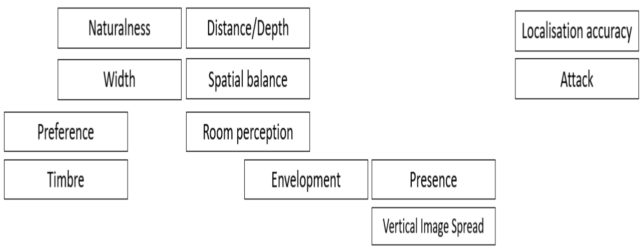


Fig. 4: Correlations between attributes based on their response patterns.

3.3. Correlations between Array Gradings and Sources

First, the Main-Array dominated the high gradings for the violin regarding all attributes. Exceptions are Localisation Accuracy and Attack where the Main- and F/R-Array seemed to have swopped their behaviour. The Main-Array also prevailed the highest ratings for the violin and cello regarding the attributes Envelopment, Presence and Vertical Image Spread. Similarly, the F/R-Array dominated responses of high values for the guitar except for Localisation Accuracy and Attack. Almost to the same extent the F/R-Array led to high scores for the handpan. Analog to these examples, the Hybrid-Array was responsible for most high ratings for the cello and djembe. Considering the sonic nature of the instruments when describing source-array dependencies, the following regularities have been discovered: The Main-Array seemed to dominate primarily the responses of the violin, which exhibits a sustained HF character. Besides, it also featured sustained sources with different frequency content (violin/cello). Analog, the F/R-Array seemed to prevail sources mostly being active in the mid-frequency range and being of both, a sustained and percussive nature. The Hybrid-Array, on the other hand, seemed to have the most influence on high ratings of instruments with enhanced LF content. Also, the Main-Array led to responses of lower scores mostly for instruments entailing a percussive element. An interesting note is also the low values for the Localisation Accuracy of the violin, as for all other attributes the Main-Array resulted in the highest ratings. Secondly, a minor tendency of the F/R-Array towards lower gradings on string instruments can be assumed. Regarding the attributes Vertical Image Shift and Vertical Frequency Separation the only pattern to be discovered was that these phenomena appeared almost exclusively in the F/R- and Hybrid-Array, which both contain the height layer of the Hamasaki-Cube.

4. Discussion

4.1. Correlation between Attributes and Subject Responses

The relationships between attributes as depicted in Figure 4 are confirmed by previous research and will be discussed hereafter. Since the attribute clusters in Figure 4 are based on the rankings of the arrays for the different sources, a close correlation between the attributes implicates an identical or strongly correlated response pattern for the different arrays. Hence, when the attribute correlations can be backed up by previous research, the validity of the array gradings can be seen increased. Naturalness, for example, was proven in experiments "to be by far the most important factor in determining overall preference in sound quality […] and it may have a strong timbral component and be highly context dependent" [23]. Thus, this statement confirms the close relation between Naturalness, Preference and Timbre. It also explains the frequent appearance of the descriptor "natural" in the comments on Preference. As it was shown that timbral fidelity contributes strongly more to the overall quality judgment than spatial fidelity [23, 25], the identical response pattern for Preference and positive timbral descriptors leading to a higher correlation between Preference and Timbre than Preference and any spatial descriptive attributes can be explained this way. The positive timbral descriptors used when describing the auditory perception for stimuli which also had been graded highest in Preference were "homogeneous", "natural", "brilliant" and "highest treble content in reverb." When considering that Naturalness is one of the most critical factors for Preference and has been defined as "how similar to a natural listening experience the sound as a whole sounds," a possible explanation for the negative correlation between negative timbral perception and Preference could be seen in anomalies being specifically related to reproduced sound, such as phase issues, e.g. These are rarely experienced in natural environments and thus lack a reference point against which to compare these experiences [23]. Unlike Timbre, Naturalness appeared mostly as a positive descriptor in the comments for Preference and could be explained through the same concept [19]. Furthermore, the perception of space as a further influential factor for Preference was confirmed by Toole who found that increased quality of spatial ratings can significantly influence the overall sound quality rating [26]. In that regard, the descriptor "open", which frequently appears in the comments on Preference, contributes to "a feeling of space" and leads to higher ratings for Naturalness and Preference [23]. Hence, the rather close relation between the spatial descriptive attributes and the cluster of Naturalness and Preference can be seen confirmed. Having backed up the strong correlation between the array ratings of Preference, Naturalness and Timbre, their correlation to the attribute Width must be included in the discussion. In concert hall acoustics ASW (apparent source width) has been associated with positive listener responses [19]. According to the definitions of the current study ASW could be considered equal to "individual source width", which is a sub-attribute of Width. Therefore, the concept of ASW could be seen to back up the correlation between Width and Preference. In relation to that, the stronger correlation between Width and Preference compared to Depth and Preference was previously confirmed in [19]. Although environmental depth and source distance seem to be dominated by the perception of environmental width, Depth is still crucial to the appreciation of sound quality [23]. This explains why the array gradings for Depth are not as correlated to the array gradings of Preference as are the gradings for Width but can still be found in the same cluster area in Figure 4. Another phenomenon to explain is the correlation between the spatial descriptive attributes, Envelopment and Presence. According to Rumsey, the link of these attributes to Preference, Naturalness and so on can be established in the connection between Width and Room Perception. From there, on one side, he found a correlation between Envelopment, Room Perception and Spatial Balance, and on the other side, he declares that "presence and environmental envelopment are not necessarily the same, although they may be closely related" [19]. Combining these statements, it can be argued that the function of Envelopment as a link between spatial descriptive attributes and Presence in Figure 4 can be seen confirmed. Last but not least the isolated cluster of Localisation Accuracy and Attack requires further explanation. Presumably, a "clear transient response" as by definition for Attack in the listening test would lead to a better Localisation Accuracy. This might explain why these attributes achieved an identical response pattern. The outcome that their response patterns seem uncorrelated to the response patterns of all other attributes was confirmed in the findings of Berg & Rumsey having proved "that localisation in itself is not the attribute closest to naturalness and positive sensations" [27]. Therefore, by having backed up the correlations found between the attributes as depicted in Figure 4 the nature of the response patterns for the different arrays could be considered validated.

4.2. Expected and Unexpected Results

Although the patterns of the listening test were found to be consistent, confirmed by research, and intra-subject consistency is assumed there is a rather distinctive deviation from the expectations of the different arrays outlined in the introduction.

4.2.1. Source-Dependent Array Behaviour

It was assumed that each array would dominate the high ratings for all sources for specific attributes. Therefore, the outcome of the listening test indicating that no array dominates the highest scores for all sources for a specific attribute (or at least four of the five sources) and therefore giving the results a source-dependent character, was not expected. Although some regularities regarding frequency content and acoustic envelopes could be identified amongst the sources, a thorough explanation of the source-dependent results requires further experiments with a more controlled experimental design. When investigating the source-dependence in the psychoacoustic realm, research indicates that the radiation pattern of the different sources could be a factor which could have influenced the current results. Martin *et al.* proved that the radiation pattern impacts the instrument's perceived audio image whereas the non-coincident arrays featured the most irregular source image perceptions [28]. This is worth mentioning as in the current study only non-coincident arrays have been applied. Although the research of Martin *et al.* was only concerned about imaging, the diverse perception of the source images within the same arrays indicates that also other attributes could be affected by radiation patterns. Even if the approach of radiation patterns won't lead to an explanation of the current results, the insight gained therein could provide a better understanding of 3D recording, according to Bowles [1].

4.2.2. Vertical Image Spread, Vertical Image Shift and Vertical Frequency Separation

The unexpected result of the F/R- and Hybrid-Array dominating the perception of Vertical Image Spread, Vertical Image Shift and Vertical Frequency Separation, whereas all these attributes have been initially assigned exclusively to the Main-Array, could be explained when having a look at psychoacoustics. Spectral graphs obtained of the Hamasaki-Cube signals indicate a slightly enhanced HF-content compared to the ambient signals of the Main-Array. As the F/R- and Hybrid-Array both contained the signals of the Hamasaki-Cube for their height layer, this could have caused a pitch-height effect [10], which, amongst other possible factors, may have led to the perception of VIS, Vertical Image Shift or Frequency Separation. The scenario of an exceeded localisation threshold during mixing leading to vertical ICCT and thus to these effects [10, 14] can be considered unlikely, as the Hamasaki-Cube is optimised to capture mainly ambient sound [3]. A remaining question here would be why some of these effects have been observed in the Main-Array for string instruments, but no other sources. As the ICCT leading to these effects depends on the ICLDs between the main and height layer signals [14], this would indicate the option that the ICLDs of the string instruments were smaller than the ICLDs of the other sources. However, the reason for this would be currently unknown to the author.

4.2.3. Naturalness and Width

Although the Main-Array was claimed to convey an enhanced perception of Naturalness, in fact, only one source scored highest for this attribute for the Main-Array. This implies that the F/R- and Hybrid-Array dominated the attribute Naturalness. A possible reason could be that the highly decorrelated signals of the Hamasaki-Cube, being a part of both, the F/R- and Hybrid-Array, lead to a decreased IAC (interaural cross correlation) and thus to an increased ASW [2]. ASW can be considered being part of the attribute Width as per definition used, which resulted in the same response pattern as Naturalness. Having said this, the expected result of the F/R-Array achieving the highest scores for Width may be explained the same way.

4.2.4. Room Perception, Spatial Balance and Distance/Depth

Although it was the Hybrid-Array scoring highest for Room Perception, Spatial Balance and Distance/Depth, and not the F/R-Array, as expected, it can be said that the Main-Array scored considerably lower than any of the other two arrays for these attributes, even if this was expected. It could be assumed, that since both, the F/R- and the Hybrid-Array share the common ground of the Hamasaki-Cube height layer, that a possible explanation could be the way early reflections have been captured, as stated in [3] and [13]. Therefore, one could think that the increased height, distance and directivity of the Hamasaki-Cube microphones compared to the directivity and placement of the Main-Array height layer microphones would have led to a more distinctive capture of early reflection key parameters, such as mentioned in [13]. As a consequence, this would ease Room Perception and the perception of Spatial Balance and Distance/Depth. This, however, does not explain why the Hybrid-Array entailing omnidirectional microphones in the main layer and thus having no directivity in its capture could score higher than the F/R-Array containing the Hamasaki-Cube main layer which is optimised for capturing lateral early reflections. This is an unexpected outcome as the increased importance of lateral early reflections compared to ceiling reflections for Room Perception was originally found by Barron [29], although his experimental design slightly differs from the current study.

4.2.5. Envelopment

Based on the same arguments no explanation could be found why the Hybrid-Array scored higher on average than the F/R-Array for Envelopment. Although it is not early lateral reflections influencing the perception of Envelopment, but late lateral reflected energy [13, 19], the parameter of directional lateral capture remains the same. What could be explained instead is the positive correlation between the average scores of Envelopment and Spatial Balance, based on Hanyu *et al.* [4]. Consequently, as the Hybrid-Array was perceived to have the highest degree of Spatial Balance, it thus might also have been perceived as most enveloping.

4.2.6. Presence

Similarly, based on [19], the increased perception of Spatial Balance of the Hybrid-Array might have led to an increased perception of Presence compared to the other arrays. Based on the rather high ratings for Presence of both, the Hybrid- and F/R-Array compared to the Main-Array, it can only be hypothesised that the increased treble content in the Hamasaki-Cube height layer might have contributed to the perception of “realism” and thus to Presence [18]. If this would be the case, however, the question would arise why the Hybrid-Array was scored higher than the F/R-Array as the F/R-Array also contains the Hamasaki-Cube main layer signals, which have been shown to exhibit a slight increase of HF-content compared to the Hybrid-Array main layer signals. In any case, the results regarding this attribute stand in contradiction with the claim that one of the main advantages of the Main-Array would be its ability to convey Presence [13].

4.2.7. Timbral Colouration

Dummy head recordings have been conducted to gather an objective reference for comparison to the timbral qualities perceived in the auditory evaluation. However, no direct indications could be derived from the spectral graphs derived thereof about the nature of the possible timbral colourations of the different arrays. It is proposed that an experimental design with fewer uncontrolled variables should be applied to approach this complex matter, similar as in [12].

5. Conclusions

This work has been useful in gaining an understanding of the spatial and timbral perception of a Bowles-Array, a Fukada-Tree/Hamasaki-Cube configuration and their hybrid version. The report of the different experimental techniques and the discussion of their outcomes gave further indications on how individual array parts might have contributed to the perception of specific attributes. This insight could be seen as a basis for recording engineers experimenting with 3D recording techniques for informing some of their decisions.

6. References

[1] Bowles, D.: A microphone array for recording music in surround-sound with height channels, presented at the 139th International AES Convention (2015 October), preprint 9430.

[2] Gribben, C. and Lee, H.: The Frequency and Loudspeaker-Azimuth Dependencies of Vertical Interchannel Decorrelation on the Vertical Spread of an Auditory Image. Journal of the Audio Engineering Society (2018), 66(7/8), 537-555.

[3] Hamasaki, K. and Van Baelen, W.: Natural Sound Recording of an Orchestra with Three-dimensional

Sound, presented at the 138th International AES Convention (2015 May), preprint 9348.

[4] Hanyu, T., Kimura, S. and Chiba, S.: A New Objective Measure For Evaluation of Listener Envelopment Focusing on the Spatial Balance of Reflections. Journal of Architecture and Planning (1999), 64(520), 9-16.

[5] Holman, T.: Surround Sound: Up and Running, 2nd ed., Focal Press, Oxon, 2008.

[6] Howie, W. and King, R.: Exploratory microphone techniques for three-dimensional classical music recording, presented at the 138th International AES Convention (2015 May), e-Brief 196.

[7] Howie, W., King, R. and Martin, D.: A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound, presented at the 141st International AES Convention (2016 September), paper 9612.

[8] Howie, W., Martin, D., Benson, D., Kelly, J. and King, R.: Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques, presented at the AES International Conference on Spatial Reproduction - Aesthetics and Science (2018 July), convention paper P10-1.

[9] Kassier, R., Lee, H., Brookes, T. and Rumsey, F.: An Informal Comparison Between Surround-Sound Microphone Techniques, presented at the 118th International AES Convention (2005 May), preprint 6429.

[10] Lee, H., Gribben, C. and Wallis, R.: Psychoacoustic Considerations in Surround Sound with Height, presented at the 28th Tonmeistertagung VDT International Convention (2014 November).

[11] Riaz, H., Stiles, M., Armstrong, C., Chadwick, A., Lee, H. and Kearney, G.: Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality, presented at the 143rd International AES Convention (2017 October), e-Brief 384.

[12] Robotham, T., Stephenson, M. and Lee, H.: The Effect of a Vertical Reflection on the Relationship between Preference and Perceived Change in Timbral and Spatial Attributes, presented at the 140th International AES Convention (2016 May), preprint 9547.

[13] Theile, G. and Wittek, H.: 3D Audio Natural Recording, presented at the 27th Tonmeistertagung VDT International Convention (2012 November).

[14] Wallis, R. and Lee, H.: The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localisation Thresholds for Natural Sound Sources. Applied Sciences (2017), 7(3), 278-297.

[15] Zielinsky, G., Lemmens, P., Dabringhaus, W., Wittek, H. and Nettingsmeier, J., Proceedings of ICSA 2011. DVD. VDT, Detmold, 2011.

[16] Berg, J. and Rumsey, F.: Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique. Journal of the Audio Engineering Society (2006), (5), 365-379.

[17] Berg, J. and Rumsey, F.: Systematic Evaluation of Perceived Spatial Quality, presented at the 24th AES International Conference (2003 June), paper 43.

[18] Gerzon, M.: Whither Four Channels?. Link House Publications, Croydon, 1971.

[19] Rumsey, F.: Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm. Journal of the Audio Engineering Society (2002), 50(9), 651-666.

[20] Pedersen, T. and Zacharov, N.: The development of a Sound Wheel for Reproduced Sound, presented at the International AES Convention (2015 May), preprint 9310.

[21] Gribben, C. and Lee, H.: Towards the Development of a Universal Listening Test Interface Generator in Max, presented at the 138th International AES Convention (2015 May), e-Brief 187.

[22] Howie, W., King, R., Martin, D. and Grond, F.: Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio, presented at the 142nd International AES Convention (2017 May), preprint 9797.

[23] Rumsey, F.: Spatial Audio. Focal Press, Oxon, 2013.

[24] Bech, S. and Zacharov, N.: Perceptual Audio Evaluation - Theory, Method and Application. Wiley, Chichester, 2006.

[25] Rumsey, F., Zieliński, S., Kassier, R. and Bech, S.: On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. The Journal of the Acoustical Society of America (2005), 118(2), 968-976.

[26] Toole, F.: Subjective Measurements of Loudspeaker Sound Quality and Listener Performance. Journal of the Audio Engineering Society (1985), 33(1/2), 2-32.

[27] Berg, J. and Rumsey, F.: Correlation between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction, presented at the 109th International AES Convention (2000 September), preprint 5206.

[28] Martin, B., King, R. and Woszcyk, W: Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three Dimensional Capture of Acoustic Instruments, Part I, presented at the 141st International AES Convention (2016 September), preprint 9613.

[29] Barron, M.: The subjective effects of first reflections in concert halls - The need for lateral reflections. Journal of Sound and Vibration (1971), 15(4), 475-494.