Title	Audio quality moderates localisation accuracy : two distinct perceptual effects?
Author(s)	Lindborg, PerMagnus; Kwan, Nicholas A.
Citation	Lindborg, P., & Kwan, N. A. (2015). Audio quality moderates localisation accuracy : two distinct perceptual effects? 138th Audio Engineering Society Convention 2015, 9313-.
Date	2015
URL	http://hdl.handle.net/10220/25765
Rights	© 2015 Audio Engineering Society. This paper was published in 138 AES Convention and is made available as an electronic reprint (preprint) with permission of Audio Engineering Society. The paper can be found at the following official URL: [http://www.aes.org/e-lib/browse.cfm?elib=17737]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Audio Engineering Society

Convention Paper 9313

Presented at the 138th Convention
2015 May 7–10 Warsaw, Poland

Audio Quality Moderates Localisation Accuracy: Two Distinct Perceptual Effects?

PerMagnus Lindborg[1], Nicholas A Kwan[1]

[1] Nanyang Technological University, Singapore 637458
permagnus@ntu.edu.sg, NAKWAN1@e.ntu.edu.sg

ABSTRACT

Audio quality is known to cross-modally influence reaction speed, sense of presence, and visual quality. We designed an experiment to test the effect of audio quality on source localisation. Stimuli with different MP3 compression rates, as a proxy for audio quality, were generated from drum samples. Participants (n = 18) estimated the position of a snare drum target while compression rate, masker, and target position were systematically manipulated in a full-factorial repeated-measures experiment design. Analysis of variance revealed that location accuracy was better in wide target positions than in narrow, with a medium effect size; and that the effect of target position was moderated by compression rate in different directions for wide and narrow targets. The results suggest that there might be two perceptual effects at play: one, whereby increased audio quality causes a widening of the soundstage, possibly via a SMARC-like mechanism, and two, whereby it enables higher localisation accuracy. In the narrow target positions in this experiment, the two effects acted in opposite directions and largely cancelled each other out. In the wide target presentations, their effects were compounded and led to significant correlations between compression rate and localisation error.

1. INTRODUCTION

People listen to more audio than ever before, and use playback equipment of highly variable quality [14]. Small loudspeaker systems can be improved by pre-filtering to compensate for e.g. lacking bass response [2]. Moreover, people consume music in environments of variable sonic quality, and it has been shown that they spontaneously try to optimise the experience by adjusting equalisation and sound level [13]. This situation presents sound engineers with challenges as to selecting the delivery format, in particular when mastering commercial music (see [24], [25]). Ideally, the benefits of file compression for ease in handling and distribution should not impinge upon the quality of the listening experience. Moreover, research on crossmodal matching has provided evidence that many non-arbitrary correspondences exist between audition and other senses. These have been documented both between simple stimulus dimensions and more complex stimuli. For example, audio quality is known to cross-modally influence reaction speed ([22]), sense of presence ([11]), and perceived visual quality ([28]).

1.1. Localisation and the soundstage

Localisation is the ability to pinpoint a sound source (real or virtual) in space. In localisation, the auditory system makes use of interaural, spectral, and distance cues. Interaural cues (a.k.a. binaural cues) depend on

differences in intensity level and time-of-arrival (phase) between the wavefronts arriving at the ipsilateral (near) and contralateral (far) ear. As a wavefront reflects off the listener's anatomy, in particular the pinnæ, it suffers constructive and destructive interferences, creating spectral cues (a.k.a. monaural cues) that allow for vertical localisation and front-back discrimination (for references, see [15], [17], and [8]).

Among audiophiles, the space within which sources are perceived to be is commonly referred to as the 'soundstage'. The perceived qualities of the soundstage depend on the nature of the sources and their relationships. 'Image' is the term used for describing the localisation of a source in terms of horizontal angle (width) and distance (depth). 'Phantom images' are physically non-existent sources perceived to appear at locations between loudspeakers. Such images might be focused or spread out ([30]). In [18], Moylan describes how a sound may be localised outside the loudspeaker array, resulting in an apparent enlargement of the soundstage: a 'widening illusion'. It might be related to an effect reported in [12], where Liebetrau and collaborators showed that in a situation where a phantom source appears between loudspeakers separated by a large angle, the perceived position tends to move towards one of them. The authors observed that their experiment participants overestimated the elevation of sound sources; however, they could not rule out that this effect might have been an artefact of the apparatus setup.

The accuracy of localisation is affected in multisource environments. Impressive performance of the auditory system was shown in [6] in terms of the ability to pinpoint a source in the presence a large number of concurrent, similar sounds. In an experiment using a large loudspeaker array, a transition between "small" and "large" numbers of interfering sources occurred with about five concurrent sounds in the stimulus. With few sources, exploratory head movements improved localisation accuracy but rapid stimulus onset did not. With many sources, localisation accuracy improved with rapid-onset sounds, but head movements did not contribute ([6], p. 475).

Attention reorientation speed and localisation accuracy depend on how sources are presented within an environment. Sounds can be perceived as internalised (binaurally, e.g. using headphones) or externalised (binaurally with HRTF cues, or using loudspeakers). These might be considered as two levels of spatial rendering quality. In [22], participants were tasked to categorise naturalistic samples, processed as internalized or externalized, and presented in various spatial locations. Sounds presented inside the head received a faster and more accurate response than sounds presented outside the head. For the latter, sounds presented in the frontal hemisphere rather than the back were more accurately localised. The author suggested that externalised sounds might be processed differently by the auditory system.

When increasing numbers of sources fill a soundstage, push/pull- effects (i.e. the source being repelled from or attracted towards a masker) become more important. As pointed out in [31], this might be exploited to create more CPU-efficient algorithms for real-time spatialisation in Audiovisual Virtual Environments, (AVE) In two experiments, Larsson and collaborators manipulated the reverberation in an AVE [11]. They showed that rendering quality significantly affected evaluations of the virtual sonic environment's contribution to overall realism, the feeling of presence, and the ability to localise sources. Here, the participants performed a gamified search task, which was comparatively much slower than the localisation task in [22]. As pointed out in both articles, audio-visual crossmodal associations may lead to audio quality influencing the perceived visual quality in bimodal environments.

1.2. Lossy audio compression

As there are situations where separate encoding of left and right channels may reduce the perceived audio quality reception, joint stereo encoding aims to reduce redundancy and irrelevancy between channels. M/S channel coding utilises matrix operations on the left and right signals to produce a sum channel (mid) and a difference channel (side). This method exploits channel similarity to reduce redundancy without introducing artifacts. Intensity Stereo Coding (ISC) takes advantage of how high frequencies are perceived. It utilizes the energy time envelopes of the channels to approximate the transmitted signal to a target level for each spectral band. ISC is limited to low bitrates and high frequency range. Parametric Stereo Coding reduces time domain artifacts via a dedicated filterbank.

MP3, developed at Fraunhofer Institute [23], is a very common lossy compression format, used by people listening to music through web downloads or streaming services. The different MP3 encoders typically consist of a poly-phase filterbank built on a perceptual model,

quantisation and coding block, and bitstream encoder. Low-salient components are encoded with less detail to simplify the signal. In most situations, the spectral distortion introduced is below the threshold of being noticed. The perceptually most challenging sound types for any psychoacoustic encoder might be "single instruments, which don't give rise to a lot of the masking that arises in a busier track" [25]. To keep noise below the masking threshold, quantisation in MP3 employs a lossy method whereby larger sample values are coded with lower accuracy. The values themselves are coded with a Huffman lossless codebook. For details, see [5] and [27]. Encoding moreover depends on user options including bit depth and sampling rate, which produce varying levels of signal fidelity. Typical bit rates range from 32 to 320 kbit/s, with 128 being the most commonly used. At lower bit rates, quality deficiencies become apparent. For example, the overall listening experience may lose clarity; bass sounds lack power and warmth; and the stereo image lose focus so that point-like sources appear "spread out". As suggested in [7], this might contribute to reduced sense of envelopment. The spatial attribute 'listener envelopment' (LEV) was proposed by [3], who showed that perceived envelopment depends strongly on the relative amount of late lateral reflections e.g. reflections arriving 80 ms or more after the direct sound. This delay is within the masking time interval taken into account in MP3 compression. Therefore decreased bit rate would directly lead to reduced perceived spatial audio quality.

In parallel to lossy encoding effects on quality, it can be noted that the use of dynamic range compression in commercial music has much increased over the past few decades. It is a widespread notion that compression affects the experience of "depth" in music. This was investigated in a recent experimental study [10], which somewhat surprisingly failed to reveal evidence of any perceptual effects. As the authors pointed out, dynamic compression in commercial re-mastering is typically accompanied by additional spectral processing such as equalisation or stereo enhancement (cf. [24], [7]). The study was limited by its use of commercial music samples as stimuli, i.e. comparing differently mixed versions of the same songs. In studying dynamic processing, this did not allow controlling for the confounding influence of spectral processing.

1.3. Aims and hypotheses

While subjective effects of lossy audio encoding on spatial imaging have been described anecdotally (e.g. [7], [16], and [25]), peer-reviewed publications are scarce when it comes to controlled experiments testing perceptual effects caused by reducing audio quality through MP3 compression. After considering the literature reviewed above, we set out to investigate the influence of rendering quality on localisation accuracy. We designed an experiment to test how compression rate[1] and concurrent maskers affect source imaging within a soundstage. Specifically, we formulated three connected questions:

- Audio Quality: is there a break point for localisation accuracy with decreasing compression rate?
- Masking: is there an effect of forward and spectral upward masking on localisation?
- Widening: is the soundstage widening illusion affected by masking and compression rate?

2. PERCEPTUAL EXPERIMENT

2.1. Design

The drum kit is a basic component in modern production music (e.g. studio produced pop and rock). A realistic listening scenario was designed in which the listener is tasked to localise one out of three percussion instrument playing a rhythmic pattern. The experiment was designed as a full-factorial 4-way ANOVA, with one dependent variable, *Accuracy,* and four independently manipulated variables: *Target Position*, *Target Pan, Masker Type*, and *Compression Rate*.

2.2. Stimuli development

Samples of three instruments with sounds typical of a production music drum kit were selected from [29]. The raw samples (44.1 kHz, 24 bits) were filtered to minimize spectral overlap using the default high- and lowpass filters in [1] with the steepest available slope (48 dB/octave). Playback sound levels in the experimental setup were measured using a calibrated SPL meter (Extech 407790). See Table 1 and Section 2.2.2.

[1] In the present text, 'Compression Rate' refers to the bit rate of MP3 compression.

	sample	*filtering*	LZ_{peak}	LA_{peak}
Hihat	"loose 2 – 13in"	Highpass @ 4 kHz	80.3	79.5
Snare	"top 5 – 14in"	Bandpass @ 1…4 kHz	84.5	84.8
Kick	"4 inside – 22in"	Lowpass @ 1 kHz	85.2	69.7

Table 1. Characteristics for the three instrument samples used to create the stimuli. LZ_{peak} = unfiltered sound pressure level (SPL); LA_{peak} = A-filtered SPL; both in dB *re* 20 µPa.

In order to test the masking hypothesis, two rhythmic patterns were created. Participants were asked to focus on the target (Snare drum) which was always the last instrument played in a group of six note onsets (see Figure 1). The idea was that with the target immediately preceded by the low masker (Kick bass drum), as in pattern 1, localisation would be impaired due to spectral upward and forward masking. Conversely, with the target immediately preceded by the high masker (Hihat), as in pattern 2, the masking effect would be negligible. The durations of phrases and the silence between repeated phrases was 1.525 s; between onsets within phrases, 0.254 s; and of the whole stimulus, 10 s.

To test the image spread hypothesis, the target was panned to laterally equally spaced out phantom positions in the stereo image using a standard constant power method, thus engaging intensity level cues only. In a pre-test with two volunteers, five positions were used: fully left and right (i.e. from either loudspeaker only), in the centre (equal levels from the loudspeakers), and midway (i.e. at ±30˚ from the centre axis; see Section 2.2.2). With this arrangement, the localisation task was too easy and ratings were practically 100% accurate. Therefore the number of targets was reduced to four and defined within a smaller azimuth range, in "narrow" and "wide" positions on left and right side of the centre axis, at angles {-20˚, -10˚, +10˚, +20˚}. The masking sounds (Kick and Hihat) were always mixed at the centre, i.e. with equal level from both loudspeakers.

To test the hypothesis related to audio quality, analysis of null-difference spectra of various sounds were first made, considering a range of compression rates larger than what is commonly encountered in music audio on the World Wide Web and elsewhere. Six levels of audio quality were then selected, corresponding to uncompressed PCM and MP3 at 320, 128, 80, 56, and 48 kbps. Lossy compressed stimuli were generated using the LAME compression plugin [4] in [1], with default settings of constant bit rate and joint stereo encoding. The null difference spectral analysis had revealed the greatest frequency response differences between compression rates in the range 1…4 kHz. Since it corresponded to the main energy of a typical snare drum, we designed the experiment with a focus on the perceived localisation of this instrument.

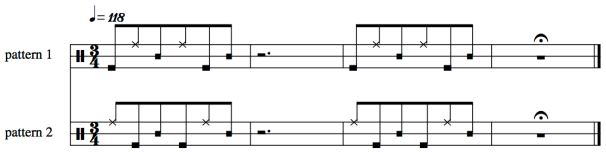

Figure 1 The two rhythm patterns used in the experiment. The target (Snare) is notated on the middle staff line. The Kick drum is on the lower staff line, and the Hihat on the upper staff line.

To sum up, 48 stimuli were prepared from all combinations of 2 rhythm patterns: *MaskType ∈ {Low, High}*, 2 target phantom width positions: *TargPos ∈ {Narrow, Wide},* 2 target pan positions, *TargPan ∈ {Left, Right},* and 6 audio quality levels: *CompRate ∈ {48, 56, 80, 128, 320, PCM}*.

2.2.1. Subjects

20 volunteers responded to an open call. Age was 20…25 years, and 8 were female. They were given a movie voucher as a token of appreciation. Each was individually screened for hearing impairment with an online test [19]. Inspection of the audiogram indicated normal hearing for 18 participants, who proceeded to the perceptual experiment.

2.2.2. Apparatus

The anechoic chamber at the School of Mechanical and Aerospace Engineering at Nanyang Technological University, Singapore was used. The background noise level was measured at LAeq,30s = 45.2 dB (LZeq = 50.1 dB). The relatively high background noise level in the chamber was probably due to slow air flow through an AC duct. A comfortable chair near the centre of the room and two active near-field studio monitor

loudspeakers (JBL Professional LSR305), on stands at 1 m (approximately participant ear height), formed an equilateral triangle 2.5 m wide. The stimuli were played from a computer via an audio interface (Focusrite Saffire Pro 14) and balanced XLR connections. The system was adjusted to produce a comfortable listening level, which produced a peak reading of 77.3 dBA (80.0 dBZ) for PCM (uncompressed) stimuli. When the phrase part of this stimulus was looped continuously (i.e without silences), the level-equivalent sound pressure, measured over 1 minute, was 73.6 dBA (78.2 dBZ). The signal-to-noise ratio, approximately 28 dB, would have been sufficiently large to rule out the possibility that background noise affected the results.

2.2.3. Procedure

Before the experiment, participants were given instructions on how to indicate the perceived direction of the target sound, and presented with training examples. The participant had in front of her a tray-like plank with 4 protruding wood pieces equally spaced 20 cm apart. While blind-folded, she indicated the perceived localisation of a stimulus by placing her hand on one of the wood pieces. The answer (which we here refer to as a 'rating') was registered on the computer keyboard by the experimenter. The 48 stimuli were presented in randomised order for each participant. Presentation was repeated in 3 blocks with a short break to rest in between. The trials were conducted in March 2014.

3. ANALYSIS

3.1. Missing values

Due to a data collection programming bug, 61 ratings were lost, mainly for the stimuli presented in the first and last presentation position in each block for the first 11 participants (for the remaining 7, the problem was corrected). These values were missing at random. In addition, 15 ratings were skipped. This might have been an effect of a systematic task difficulty, but upon inspection no pattern among the corresponding stimuli could be identified, and it was assumed they had been skipped at random due to fatigue. Therefore all missing values were imputed by the mean of ratings from the other two blocks within the same participant. The number of missing values thus substituted was 76 out of a total 2592, or 2.9%.

3.2. Rating consistency and agreement

The consistency of ratings for each participant across blocks was excellent: mean pair-wise correlation between blocks was r = 0.91, and for all except one participant it was above 0.82. The value for the least consistent participant was 0.70 and this was considered high enough to be included. The agreement of ratings between participants (considered as random sampled judges) as measured by Intraclass Correlation ICC2(18) = 1.0, which is considered very high,

3.3. Data coding

The *Error* of each rating was defined as:

$$Error = RatePos - TarPos$$

with *RatePos*, *TargPos* ∈ {-2, -1, 1, 2} corresponding to azimuts of target phantom positions, {-20˚, -10˚, +10˚, +20˚}. The *Accuracy* of each rating was defined as:

$$Accuracy = abs(Error)/4$$

so that *Accuracy* = [0…1].

4. RESULTS

The three hypotheses were tested by conducting an analysis of variance (ANOVA), with *Accuracy* as the dependent variable repeated across blocks, and *TargPos*, *TargPan*, *CompRate*, and *MaskType* as independent variables. Participants were considered as random sampled. The analysis was made in R [20]. The null hypothesis in each test was that neither the factors nor their interactions had an effect on localisation accuracy. The significance was evaluated at the α = 0.05 level. Results are given in Table 2.

4.1. A priori tests: localisation accuracy

The only significant main effect was for *TargPos*, i.e. narrow or wide target position. *Accuracy* was 0.944 for wide positions, and 0.907 for narrow; the effect (Cohen's *d*) of the difference was of medium size, 0.26 SD. In other words, the participants were better at localising the target when presented in wide positions (left or right). See Figure 2.

	F	*p*
CompRate	1.34	0.25
MaskType	0.01	0.93
TargPos	49.3	< 0.00001 ***
TargPan	0.00	0.96
CompRate:MaskType	1.23	0.29
CompRate:TargPos	2.64	0.022 *
MaskType:TargPos	0.08	0.78
CompRate:TargPan	1.44	0.21
MaskType:TargPan	3.67	0.056 .
TargPos:TargPan	3.13	0.077 .

Table 2. ANOVA of *Accuracy*. For clarity, only one- and two-way effects within blocks and participants are included. None of the between-effects or higher within-effects interactions was significant. Codes for p-values: *** p<0.001; ** p<0.01; * p<0.05; . p<0.1.

The only significant interaction effect was for *CompRate:TargPos*, i.e. the interaction between compression rate (proxy for audio quality) and target position (narrow or wide). Because the main effect of *TargPos* was significant and that of *CompRate* was not, the interaction could be interpreted as *CompRate* moderating the effect of *TargPos* onto *Accuracy*. As seen in the interaction plot in Figure 3, *Accuracy* increased with *CompRate* for wide target position, but decreased with *CompRate* for narrow target positions.

The interaction effect was investigated by conducting a multiple linear regression of *Accuracy* onto *CompRate* and *TargPos*, with the dependent variable averaged across blocks, and the independent variables coded as random dummy variables. As in the ANOVA, results showed that the main effect of *TargPos* was significant ($t(1, 863) = 5.2$ ***) while that of *CompRate* was not ($p = 0.62$ ns). However, in the interaction, *CompRate* significantly moderated the effect of *TargPos* ($t(6, 858) = 3.0$ **). The size of the *TargPos* main effect in terms of standardized beta coefficients was ß = 0.17, and that of the interaction effect was ß = 0.10. Note that these results are tentative given the dummy encodings, and also since homoscedasticity was not assessed.

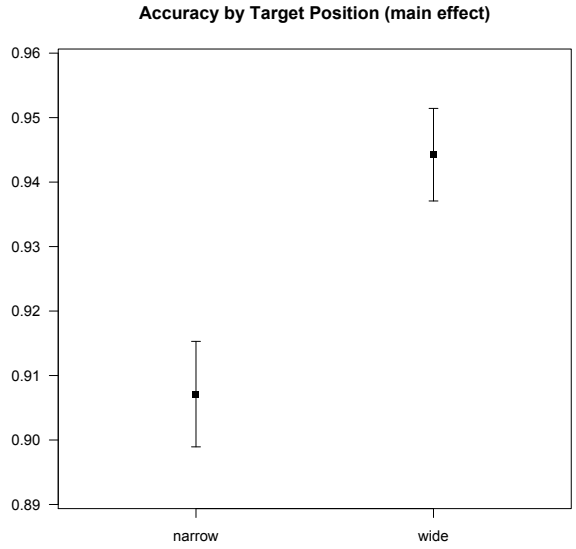

Figure 2 Main effect of *TargPos* (target position) onto *Accuracy*. Means with 95% confidence intervals for 'wide' and 'narrow' positions.

4.2. Further analysis: localisation error

The effect of target position on localisation was investigated by looking separately at the four cases of *TargPos*, i.e. the actual (phantom) position of the target snare drum. Here, we were interested in the localisation *Error,* which, as defined in Section 3.3, can be positive or negative. The correlation between *Error* and *CompRate* was calculated in each case, using the non-parametric Kendall's *tau* statistic (for which the standard error is known and thus the probability can be evaluated). For the wide positions, correlations were significant for both "wide left" and "wide right" (*tau* = -0.15** and 0.13*, respectively), with the opposing signs indicating that rating errors were mainly directed inward, away from the actual target position and towards the centre. For the narrow positions, the correlation was borderline significant for "narrow left" (*tau* = -0.10, p = 0.046*), with rating errors mainly directed outward and away from the target, while the correlation was non-significant for "narrow right".

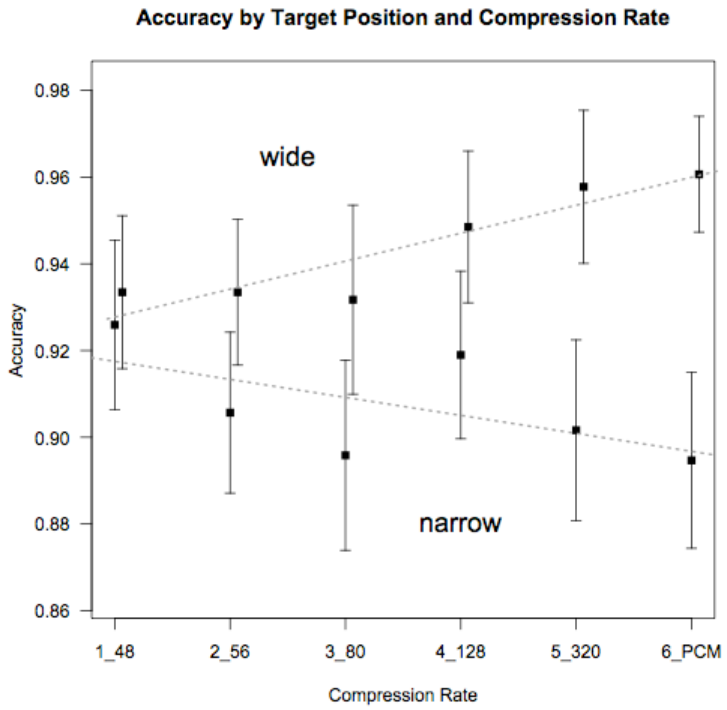

Figure 3. Plot showing the interaction between target position and compression rate onto localisation accuracy; means with 95% confidence intervals. Regression lines have been plotted using the dummy encoding of *CompRate*, but note that in reality it was a fixed ordinal variable.

The results are illustrated in the series of four plots in Figure 4. In each subplot, the x-axis indicates a part of the "absolute" azimuth axis on a scale from -2 to 2, corresponding to the 'wide left' and 'wide right' positions (at -20˚ and +20˚ respective to the centre axis of the setup; recall that the loudspeakers were at ±30˚). The y-axis indicates *Accuracy* of localisation. The vertical dashed line indicates the actual position of the target, and the + symbol indicates the mean rated position. The ellipses indicate rated target positions, grouped in two contrasts: "low' audio quality, i.e. with *CompRate* ∈ {48, 56, 80}, and "high" audio quality, i.e. with *CompRate* ∈ {128, 320, PCM}. The mean ratings for *high* and *low* quality in each subplot are indicated with ellipses where the radii correspond to 95% confidence intervals around the means of *RatePos* and *Accuracy*, respectively.

4.3. Discussion

The plots in Figures 3 and 4 reveal two things. Firstly, targets presented in wide positions have higher *Accuracy* (smaller error) when audio quality is higher. The left-right symmetry is clearly evident for these cases. The wider positioned targets follow an outward trend, and ratings come closer to the 'true target' (dotted line) as bit rate increases. Secondly and by contrast, the narrow positioned targets do not show this pattern: ratings do not come closer to the target as bit rate increases, but rather the opposite (larger error). Moreover, *Accuracy* is lower for higher quality stimuli presented narrowly. Recall that the ANOVA results in Table 2 and the interaction plot in Figure 3 revealed a tendency for rating accuracy to be inversely related to audio quality, for stimuli presented in the narrow position.

The wide plots illustrate the 'widening illusion' on position ratings with increased compression rate. The participants in this experiment tended to perceive higher quality audio as a widened soundstage. As the narrow plots indicate, the widening effect was probably present but less pronounced there; however, the current data do not rule out that the widening effect might be absent at narrow target positions.

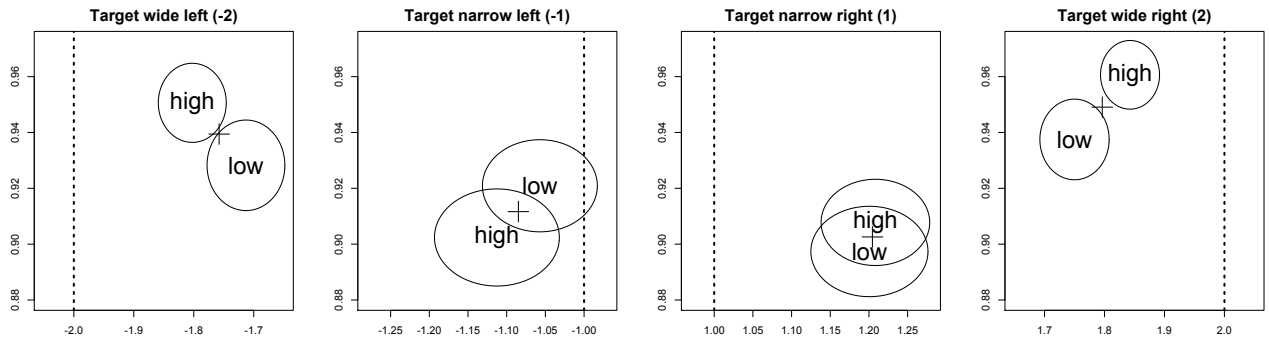

Figure 4. Plots for the four *TargPos* cases. In each, the target's actual pan position is marked by a dotted vertical line. X-axis: rated position (*RatePos*). Y-axis: *Accuracy*. Ellipses marked 'high' indicate stimuli with higher audio quality (*CompRate* = 128 kbit/s and above) and 'low' are for lower audio quality (*CompRate* = 80 kbit/s and below). The radii of ellipses correspond to 95% confidence intervals around the means of *RatePos* and *Accuracy*, respectively. See the main text for detail

At first, these results were somewhat puzzling, but with the concept of soundstage imaging, a plausible explanation emerged. In the present experiment, the rating error in wide positions can only go 'inwards', while in narrow positions the rating error can go in both directions. However, since there was no significant effect of *TargPan* on *Accuracy*, we may assume that the participants made few left-right mistakes. The rating error in narrow positions tended to go 'outwards'. This is consistent with the observations in [12] and [18] reviewed in Section 1.1.

We interpret the findings by suggesting that when listeners are presented with alternating high- and low-quality stimuli, there is a cross-sensory effect happening so that quality becomes associated with spatial position. Perhaps mediated by an emotional response, so that higher quality stimuli are perceived as being "clearer", audio quality becomes associated to a widening of the spatial attribute. The effect seems akin to the SMARC effect (Spatial Musical Association Response Code; see [26]) but is not limited to the modal effect of musical pitch. The results suggest that there might be two perceptual effects of compression rate on sound source imaging: higher quality audio causing a widening of the soundstage; and simultaneously: higher quality audio causing better localisation. In the present experimental results, their combined effects could explain the results. For the narrow target positions, the two perceptual effects acted in opposite directions: the widening effect was directed outward, but the accuracy effect acted inward. The competing forces made the phantom image instable. By contrast, for the wide target positions, the two effects acted in the same outward direction, leading to the clearly defined separation between localisation of high and low quality audio stimuli illustrated in the outer plots in Figure 4.

4.4. Summary

In relation to the hypotheses, the results can be summarised as follows:

- Audio Quality: no 'break point' could be identified. Interestingly, localisation accuracy increased with compression rate for stimuli presented in wide positions, but decreased for stimuli presented in narrow positions. This interaction effect might be caused by two competing perceptual effects of audio quality: a SMARC-like effect and an accuracy effect.
- Masking: there was no significant effect of forward and spectral upward masking on localisation accuracy in this experiment. The onsets of percussive sounds in the stimuli were separated by 254 ms, which is more than the 200 ms cited as as a limit beyond which forward masking is no longer produced ([17], p. 110). The energy level in the later part of the Kick Bass sample (i.e. after 200 ms), less than 13 dB below peak, was insufficient to produce the spectral upwards masking effect.
- A widening of the soundstage was evidenced for targets presented in wide positions, and moderated by compression rate. For targets presented in narrow positions, the SMARC effect was largely cancelled out by the competing accuracy effect, acting in the opposite direction. This finding was a (serendipitous)

result of our experimental apparatus set up to allow strictly four rating positions. Further research might try a modified experimental setup allowing for spatially continuous blindfolded location ratings (including outside the perimeter of the loudspeakers).

5. CONCLUSION

We have presented results from an empirical, controlled experiment where participants estimated the localisation of a target snare drum sound within a rhythmic pattern, while audio quality, masker, and target sound phantom source presentation position were subjected to full-factorial manipulation. ANOVA revealed a significant effect of target position onto location accuracy, so that targets presented in wide positions were more accurately localised than targets presented in narrow positions, with a medium-sized effect. Furthermore, the interaction between target position and compression rate emerged as significant. It could be interpreted as audio quality influencing localisation accuracy differently in wide and narrow target positions.

This finding was explored in a series of regressions and correlations. One possible explanation of the patterns is that they were caused by the existence of two perceptual effects caused by different levels of audio quality: a SMARC-like effect, producing a perceived widening of the soundstage; and an accuracy effect, whereby higher quality audio allowed better source localisation. In the narrow target presentations, these two effects acted in opposite directions, and largely cancelled each other out. In the case of wide target presentations however, these effects were compounded, leading to significant correlations between compression rate and localisation error.

5.1. Limitations of the study and future work

This experiment was limited to effects of MP3 compression on source localisation in the frontal part of the lateral plane. The soundstage spans not only width but also depth and elevation, which have not been covered here. The effects of compression rate on soundstage widening and location accuracy in the dorsal lateral plane might be explored with a 5.1 loudspeaker setup, while effects on vertical plane localisation necessitate a 3D multichannel loudspeaker arrangement.

Furthermore, since only the MP3 decoder is standardised (while the encoder is not) there can be significant quality differences across different implementations. The present results regarding compression rates are conditioned upon the LAME compression scheme that we used. Future work might extend to other ways of manipulating audio quality. For example, going beyond MP3, encoding methods such as high-efficiency AAC [32] are likely to perform better in terms of retaining cues for localisation even under large compression rates. It remains to be tested whether the two distinct perceptual effects (image widening and increased accuracy) that the present results have suggested are also relevant to this and other schemes.

6. REFERENCES

[1] Audacity [software] (2014). *Audacity v2.0.2* http://audacity.sourceforge.net/ (acc. 9 Dec. 2014)

[2] Bahne (2012, Jan./Feb.). "Perceived Sound Quality of Small Original and Optimized Loudspeaker Systems". *J. Audio Eng. Soc. 60:1-2*, 29-37.

[3] Bradley JS & Soulodre GA (1995). "Objective measures of listener envelopment". *J. Acoust. Soc. Am. 98*, 2590-2597.

[4] LAME [software] (2011, Oct.) *LAME v3.99.* http://lame.sourceforge.net/ (acc. 9 Dec. 2014)

[5] Brandenburg K (1999, Sep.). "MP3 and AAC Explained". AES 17th Int. Conf. on High Quality Audio Coding. http://graphics.ethz.ch/teaching/mmcom12/slides/mp3_and_aac_brandenburg.pdf (acc. 9 Dec. 2014).

[6] Brungart D & Simpson B (2005). "Localization in the Presence of Multiple Simultaneous Sounds**".** *Acta Acustica united with Acustica, 91*, 471-479.

[7] Corbett I (2012, Apr.). "What Data Compression Does To Your Music". http://www.soundonsound.com/sos/apr12/articles/lost-in-translation.htm (acc. 9 Dec. 2014).

[8] Gardner WG (1997). *3D Audio Using Loudspeakers*. Kluwer Academic Publisher.

[9] Herre J (2004). "From joint stereo to spatial audio coding - recent progress and standardization". *7th Int. Conf. on Digital Audio Effects.* http://dafx04.na.infn.it/WebProc/Proc/P_157.pdf (acc. 1 Feb. 2015).

[10] Hjortkjær J & Walther-Hansen M (2014 Jan./Feb.). "Perceptual Effects of Dynamic Range Compression in Popular Music Recordings". *J. Audio Eng. Soc. 62-1/2,* 37-41.

[11] Larsson P, Västfjäll D & Kleiner M (2002, Jun.). "Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering". *22nd Audio Engineering Society Conference*, pp. 1-8.

[12] Liebetrau J, Sporer T, et al. (2007, Oct.). "Localization in Spatial Audio - from Wave Field Synthesis to 22.2". *Audio Engineering Society 123rd Convention*, New York, pp. 1-9.

[13] Lindborg PM & Lim MJY (2013, July). "Design of an Interactive Earphone Simulator and Results from a Perceptual Experiment". *Proc Sound and Music Computing Conf. 2013* (SMC 2013), Stockholm, Sweden, pp. 74-79.

[14] Lim MJY & Lindborg PM (2013, June). "How Much does Quality Cost? Listening to Music with Earphones on Buses and Trains". *Proc. 3rd Int. Conf. on Music Emotion (ICME3)*. Jyväskylä, Finland, pp. 1-6.

[15] Loy G (2006). *Musimathics: The Mathematical Foundations of Music, Vol. 1*. MIT Press.

[16] Maher RC (2003). "Lossless Compression of Audio Data". In *Lossless Compression Handbook* (ed. Sayood). http://www.eetimes.com/document.asp?doc_id=1274863 (acc. 9 Dec. 2014).

[17] Moore BCJ (1977, 2012). *An Introduction to the Psychology of Hearing, 6th ed*. Emerald Group Publishing, UK, pp. 1-441.

[18] Moylan W (2006). *Understanding and Crafting the Mix: The Art of Recording, 2nd ed*. Focal Press.

[19] Pigeon S (2014). *Online Audiometric Test and Audiogram Printout.* http://myhearingtest.net/ (acc. 9 Dec. 2014)

[20] R [software] (2014). *R v3.0.3*. http://www.r-project.org/ (acc. 9 Dec. 2014)

[21] Revelle W (in prep). *An introduction to psychometric theory with applications in R*. http://personality-project.org/r/book/ (acc. 10 Feb 2014).

[22] Roginska A (2012, Jul./Aug.). "Effect of Spatial Location and Presentation Rate on the Reaction to Auditory Displays". *J. Audio Eng. Soc. 60:7-8*, 497-504.

[23] Rose M & Frauenhofer IIS (2013). "The mp3 History". http://www.mp3-history.com/ (acc. 9 Dec. 2014).

[24] Rumsey F (2005). *Sound and Recording. 5th Ed.* Focal Press.

[25] Rumsey F (2013). "Mastering for today's media". *J. Audio Eng. Soc. 61:1-2*, 79-83.

[26] Rusconia E, Kwana B, Giordano BL, Umiltà C & Butterworth B (2006). "Spatial representation of pitch height: the SMARC effect". *Cognition* 99:2, 113–129.

[27] Sellars P (2000, May). "Perceptual Coding: How Mp3 Compression Works". *Sound on Sound*. http://www.soundonsound.com/sos/may00/articles/mp3.htm (acc. 9 Dec. 2014).

[28] Spence C (2011). "Crossmodal correspondences: A tutorial review". *Atten. Perceptual Psychophysiology,* 73:971–995. DOI 10.3758/s13414-010-0073-7

[29] SignalToNoise (2012). "12 piece pearl master custom drum kit with zildjian cymbals" http://rhythminmind.net/STN/?page_id=14 (acc. 1 Feb. 2014).

[30] Thiele G (1980). *On the localisation in the superimposed soundfield [Thesis]*. Technische Universität Berlin.

[31] Tsingos N (2007, Sep.). "Perceptually-based auralization". *19th Intl. Congress on Acoustics*, Madrid, pp. 1-6.

[32] Wolters M, Kjörling K, Homm D & Purnhagen H (2003). "A closer look into MPEG-4 High Efficiency AAC". *Proc. AES 115th Convention*, New York, USA.