Review of MM-004756

I. SUMMARY

The paper "High Quality Low Bitrate Spatial Audio Coding" is concerned with optimizing spatial encoding for audio compression. The authors point out that existing techniques for spatial audio coding do not seek to minimize the decoded audio reconstruction error, and they suggest that this should be done.

The idea seems both reasonable and appealing, and the results seem convincing, but I have some concerns with repect to the presentation.

II. LIST OF CONCERNS

1) The paper has a lot of overlap with the authors' previous work in [59], something that I personally feel should be acknowledged already at the beginning of the paper. Some of the extensions from [59] are not quite clear. Specifically, Secs. III and IV are very confusing to me. I do not follow the authors' explanation of AbS in the context of spatial encoding, and it is especially unclear to me how the results in Sec. III are the basis for the computations described in Sec. IV. I will try to describe this confusion further in the following points:

2) The general model described by Fig. 3 (and the accompanying description) is confusing in that it has no definite final output and that it is not clear what is communicated back to the "Spatial analysis" block. Also, it seems like the "Error minimisation" block depends on the "Spatial Parameters", even though they are not fed through. Do not get me wrong, I understand what the authors are trying to express here. But it is my opinion that if you choose a schematic diagram to convey your idea, it should be complete and consistent.

3) It seems like the conclusion after Eqs. (22) and (23) is simply that if the downmix signal is calculated from the quantized (or estimated) energy constants instead of the actual energy constants, then the reconstruction is error free. However, I do not see what is gained by the (to me) rather inaccessible derivation in III-B; it does not seem to open for any other solutions, extensions, or conclusions. I think that the authors need to look over their presentation of the method and evaluate whether this is the most effective method of getting their points across. Especially, since this derivation is already given in [59], I think a lot of the more repetitive

4) I realize that no background section is ever complete, but I would still encourage the authors to include a section on Ambisonics and Higher-Order Ambisonics (HOA). This is especially relevant because spatially targeted compression techniques have been applied here (see e.g. E. Hellerud, A. Solvang, U. P. Svensson, "Spatial redundancy in Higher Order Ambisonics and its use for lowdelay lossless compression", ICASSP 2009).

III. CONCLUSION

My main problem with this paper is that a lot of space is used to cover material that has been thoroughly covered by an earlier paper of the authors. The connections to the new additions to the theory are not quite clear to me, and I think the authors should perhaps make Secs. III and IV a bit more accessible.