PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022, in Gyeongju, Korea

Extended abstract

ABS-0626

The Auditory Modeling Toolbox 1.x

Clara HOLLOMEY[1]; Piotr MAJDAK; Robert BAUMGARTNER

Acoustics Research Institute, Academy of Sciences, Austria

ABSTRACT

The Auditory Modeling Toolbox (AMT) is an open source Matlab/Octave toolbox dedicated to promoting reproducible research in the hearing sciences. Through its unified interface, it provides access to implementations of auditory models written in various programming languages, to experimental data, and to code focusing on the reproduction of published results. An extensive in-code generated documentation and software demonstrations of the relevant aspects of the models assist in getting quickly familiar with their functioning. Moreover, the AMT 1.x comprises tools to facilitate the extension and modification of existing auditory models, such as the caching of results, on-the-fly download of experimental data and head-related transfer functions from online repositories, as well as general purpose auditory functions such as filters, signal generators, and plotting functionality. For the contributors, the AMT offers the option of multi-licensing of their implementations, a clear display of authorship, and citations to their authors' publications. With the AMT 1.2, over 60 auditory models and experiments as well as 50 sets of data contributed by researchers from a wide range of scientific fields are provided. The AMT's code, documentation, and resources are provided at http://amtoolbox.org.

Keywords: Computational modeling, psychoacoustics, open source software, reproducible research

1. INTRODUCTION

The Auditory Modeling Toolbox (AMT) 1.x [1] is a framework for making auditory models and their associated experimental data available to the general public. Rooted in the idea of promoting reproducible research, the AMT 1.x is distributed under the open source GNU Public License Version 3, and is freely available from amtoolbox.org. Its basic structure is implemented in Matlab/GNU Octave, but models in any programming language are supported. Its online documentation is generated directly from within the code, providing transparency to the user via a direct link between the source code and its results.

Besides the software package, AMT 1.x comprises experimental and auxiliary data that can be downloaded on the fly and directly from within the toolbox. Additionally, a caching mechanism allows for the online retrieval of pre-calculated data, thus shortening the computation time for many auditory models. Third-party toolboxes complement AMT 1.x with functionality for conducting advanced signal analysis [2], processing head-related transfer functions (HRTFs) in the spatially oriented format for acoustics (SOFA) [3], carrying out statistical analysis [4], and synthesizing 3D sound fields [5]. Figure 1 shows the structure of the AMT 1.x, outlining the connections between the modules of the code, downloadable resources, and their origins.

Models and model stages are central to the code structure of AMT 1.x. They are linked to a specific publication, named after its first author and year, and are ideally supported by experiment functions reproducing that publication's results. The achievable degree of reproducibility as well as the quality of the code and documentation are ranked on the AMT's homepage to provide further transparency. Additional demonstrations showcase the model's most important aspects. Common functions offer auditory modeling-related functionality that can be used in conjunction with several models. Data functions, finally, give access to experimental data from various publications, usually via download from the online repository of the AMT 1.x.

[1] clara.hollomey@oeaw.ac.at

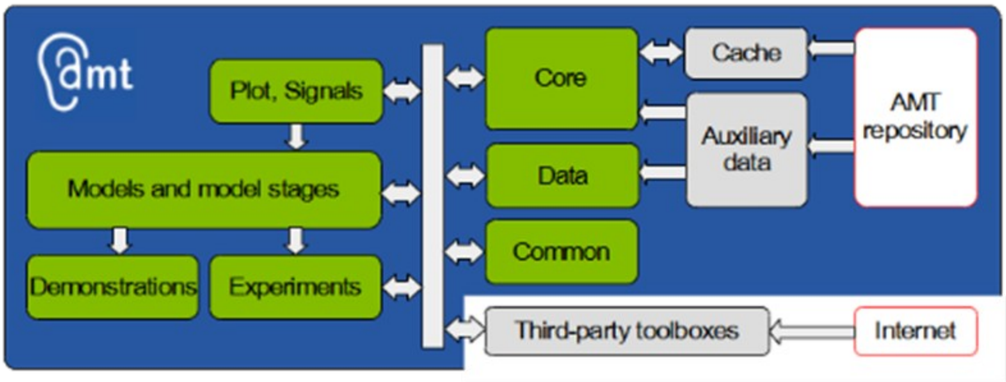

Figure 1: Structure of the AMT 1.x. Green: Code consisting of functions and algorithms. Grey: Data and third-party toolboxes. White: Online repositories. Figure reproduced from [1].

2. AUDITORY MODELS

With the release of AMT 1.2, the AMT comprises more than 50 models functionally covering the auditory pathway from the outer ear up to the brainstem. Moreover, there are models for higher level behavioral and perceptual features, such as loudness, speech, and spatial perception. In the following, we summarize the available model implementations in AMT 1.x and outline their correspondence to the auditory processing stages depicted in Figure 2.

2.1 The auditory pathway

The outer and middle ear processing is largely done via common functions. HRTFs need to be in SOFA format and can be checked for their geometrical consistency via the model `ziegelwanger2014` [6]. `pausch2022` [7], another acoustic-geometric model, estimates the interaural time differences based on hearing-aid related transfer functions.

The models `lopezpoveda2001` [8], `hohmann2002` [9], `lyon2011` [10], and `verhulst2012` [11] yield the basilar membrane velocity as a function of frequency. `zilany2007` [12], `zilany2014` [13], and `bruce2018` [14] implement the complete chain from sound pressure to spike rates of AN fibers. `ewert2000` [15] and `carney2015` [16] account for the sensitivity to temporal modulations of the auditory system, which is commonly attributed to the neural auditory pathway from the cochlear nucleus to the inferior colliculus. Models integrating the more peripheral stages with higher-level neural stages comprise `dau1996` [17], `dau1997` [18], `roenne2012` [19], `verhulst2015` [20], `verhulst2018` [21], `relanoiborra2019` [22], and `king2019` [23]. Binaural processing is supported by the models `lindemann1986` [24], `breebaart2001` [25], `dietz2011` [26], and `takanen2013` [27].

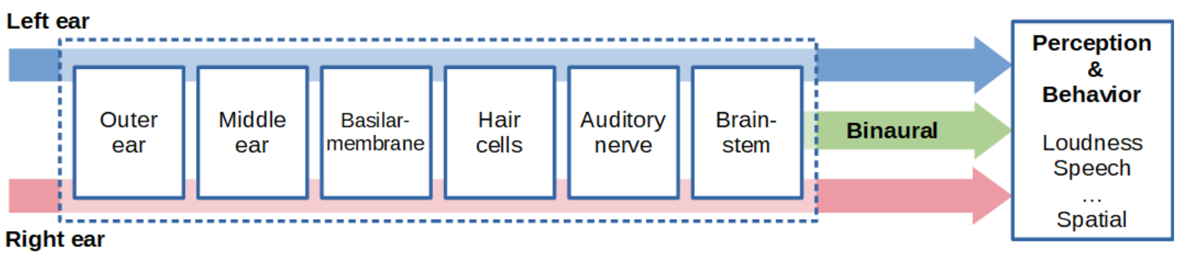

Figure 2: Typical structure of auditory models reflecting the monaural processing stages of the auditory periphery (Left ear, Right ear), followed by an optional stage of binaural interaction (Binaural) and stages modeling perceptual or behavioral outcomes (Perception & Behavior). Figure reproduced from [1].

2.2 Perception and behavior

In the AMT 1.2, we have monaural loudness models represented by `moore1997` [28], `glasberg2002` [29], and `chen2011` [30] as well as `moore2016` [31], a loudness model considering binaural inhibition. We have monaural speech perception models such as `joergensen2011` [32], `taal2011` [33], and `joergensen2013` [34], complemented by models considering binaural speech processing, such as `culling2004` [35], `jelfs2011` [36], `leclere2015` [37], `hauth2020` [38], `prudhomme2020` [39], `vicente2020nh` [40], `vicente2020` [41], and `lavandier2022` [42]. For perceptual similarity, we have `osses2021` [43] and `mckenzie2022` [44]. Models for spatial perception comprise `zakarauskas1993` [45], `langendijk2002` [46], `may2011` [47], `baumgartner2013` [48], `georganti2013` [49], `wierstorf2013` [50], `baumgartner2014` [51], `reijniers2014` [52], `kelvasa2015` [53], `baumgartner2016` [54], `hassager2016` [55], `li2020` [56], `baumgartner2021` [57], `barumerli2022` [58], `llado2022` [59] and `mclachlan2021` [60].

3. OTHER IMPORTANT INFORMATION

The AMT is not limited to models and their original published experiments, but also comprises contributions from researchers applying AMT models in their publications. At the moment, the AMT comprises five such experiments: `exp_baumgartner2015` [61], `exp_baumgartner2015 binweight` [62], `exp_steidle2019` [63], `exp_engel2021` [64], `exp_osses2022` [65], and `exp_roettges2022` [66].

The AMT is under active maintenance and we welcome new contributions, be it models, data, or experiments, from all fields of hearing science and in all programming languages. All contributors are listed on the AMT's website, are declared in the source code, and are cited in AMT-related publications.

All information and resources related to the AMT can be found at https://amtoolbox.org. The AMT version 1.2 can be downloaded from https://sourceforge.net/p/amtoolbox/files. For the development, the source code is available at https://git.code.sf.net/p/amtoolbox/code.

ACKNOWLEDGEMENTS

This work was supported by the European Union (EU, project 'SONICOM' 101017743, RIA action of Horizon 2020).

REFERENCES

[1] P. Majdak, C. Hollomey, and R. Baumgartner. "AMT 1.x: A toolbox for reproducible research in auditory modeling," Acta Acust, vol. 6, 19, 2022.

[2] Z. Pruša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr and P. Balazs. "The large time-frequency analysis toolbox 2.0," in Sound, Music, and Motion, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds. Cham, Switzerland: Springer, 2014, pp. 419–442

[3] P. Majdak,, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig. "Spatially Oriented Format for Acoustics 21: Introduction and Recent Advances," J Audio Eng Soc 70, 2022, pp. 565–584.

[4] P. Berens, CircStat: A Matlab Toolbox for Circular Statistics, Journal of Statistical Software, Volume 31, Issue 10, 2009

[5] I. Engel, BinauralSH, https://zenodo.org/record/5012460#.Yt3CO-xBxpQ, assessed 21/07/2022, 19:30

[6] H. Ziegelwanger and P. Majdak, 'Modeling the direction-continuous time-of-arrival in head-related transfer functions', J. Acoust. Soc. Am., vol. 135, no. 3, pp. 1278–93, Mar. 2014, doi: 10.1121/1.4863196.

[7] F. Pausch, S. Doma, and J. Fels, Hybrid multi-harmonic model for the prediction of interaural time differences in individual behind-the-ear hearing-aid-related transfer functions. Acta Acust, vol. 6, 34, 2022

[8] E. A. Lopez-Poveda and R. Meddis, 'A human nonlinear cochlear filterbank', J Acoust Soc Am, vol. 110, no. 6, pp. 3107–18, Dec. 2001, [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/11785812

[9] V. Hohmann, ‘Frequency analysis and synthesis using a Gammatone filterbank’, Acta Acust. United Acust., vol. 88, no. 3, pp. 433–442, May 2002.
[10] R. Lyon, ‘Cascades of two-pole–two-zero asymmetric resonators are good models of peripheral auditory function’, J Acoust Soc Am, vol. 130, pp. 3893–3904, 2011, doi: 10.1121/1.3658470.
[11] S. Verhulst, T. Dau, and C. Shera, ‘Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission’, J. Acoust. Soc. Am., vol. 132, pp. 3842–3848, 2012.
[12] M. S. A. Zilany and I. C. Bruce, ‘Representation of the vowel $/\epsilon/$ in normal and impaired auditory nerve fibers: Model predictions of responses in cats’, J. Acoust. Soc. Am, vol. 122, pp. 402–248, 2007.
[13] M. S. A. Zilany, I. C. Bruce, and L. H. Carney, ‘Updated parameters and expanded simulation options for a model of the auditory periphery’, J. Acoust. Soc. Am., vol. 135, no. 1, pp. 283–286, Jan. 2014, doi: 10.1121/1.4837815.
[14] I. C. Bruce, Y. Erfani, and M. S. A. Zilany, ‘A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites’, Hear. Res., vol. 360, pp. 40–54, Mar. 2018, doi: 10.1016/j.heares.2017.12.016.
[15] S. Ewert and T. Dau, ‘Characterizing frequency selectivity for envelope fluctuations’, J Acoust Soc Am, vol. 108, pp. 1181–1196, 2000, doi: 10.1121/1.1288665.
[16] L. H. Carney, T. Li, and J. M. McDonough, ‘Speech Coding in the Brain: Representation of Vowel Formants by Midbrain Neurons Tuned to Sound Fluctuations’, eNeuro, vol. 2, no. 4, Jul. 2015, doi: 10.1523/ENEURO.0004-15.2015.
[17] T. Dau, D. Püschel, A. Kohlrausch: A quantitative model of the “effective” signal processing in the auditory system. I. Model structure. J. Acoust. Soc. Am vol. 99, no. 6, pp. 3615–3622, 1996.
[18] T. Dau, B. Kollmeier, and A. Kohlrausch, ‘Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers’, J. Acoust. Soc. Am., vol. 102, no. 5, pp. 2892–2905, 1997.
[19] F. M. Rønne, T. Dau, J. Harte, and C. Elberling, ‘Modeling auditory evoked brainstem responses to transient stimuli’, J. Acoust. Soc. Am., vol. 131, no. 5, pp. 3903–3913, 2012, doi: 10.1121/1.3699171.
[20] S. Verhulst, H. M. Bharadwaj, G. Mehraei, C. A. Shera, and B. G. Shinn-Cunningham, ‘Functional modeling of the human auditory brainstem response to broadband stimulation.’, J. Acoust. Soc. Am., vol. 138, no. 3, pp. 1637–1659, Sep. 2015, doi: 10.1121/1.4928305.
[21] S. Verhulst, A. Altoè, and V. Vasilkov, ‘Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss’, Hear. Res., vol. 360, pp. 55–75, Mar. 2018, doi: 10.1016/j.heares.2017.12.018.
[22] H. Relaño-Iborra, J. Zaar, and T. Dau, ‘A speech-based computational auditory signal processing and perception model’, J. Acoust. Soc. Am., vol. 146, no. 5, pp. 3306–3317, Nov. 2019, doi: 10.1121/1.5129114.
[23] A. King, L. Varnet, and C. Lorenzi, ‘Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept’, J. Acoust. Soc. Am., vol. 145, no. 4, pp. 2277–2293, 2019.
[24] W. Lindemann, ‘Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals’, J. Acoust. Soc. Am., vol. 80, no. 6, pp. 1608–1622, 1986, doi: 10.1121/1.394325.
[25] J. Breebaart, S. van de Par, and A. Kohlrausch, ‘Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters’, J Acoust Soc Am, vol. 110, no. 2, pp. 1105–17, Aug. 2001, [Online].
[26] M. Dietz, S. D. Ewert, and V. Hohmann, ‘Auditory model based direction estimation of concurrent speakers from binaural signals’, Speech Commun., vol. 53, pp. 592–605, 2011.
[27] M. Takanen, O. Santala, and V. Pulkki, ‘Binaural assessment of parametrically coded spatial audio signals’, in The technology of binaural listening, J. Blauert, Ed. Berlin, Germany: Springer, 2013, pp. 333–358.
[28] B. C. J. Moore, B. R. Glasberg, and T. Baer, ‘A Model for the Prediction of Thresholds, Loudness, and Partial Loudness’, J Audio Eng Soc, vol. 45, no. 4, pp. 224–240, 1997.
[29] B. R. Glasberg and B. C. J. Moore, ‘A Model of Loudness Applicable to Time-Varying Sounds’, J Audio Eng Soc, vol. 50, no. 5, pp. 331–342, 2002.
[30] Z. Chen, G. Hu, B. R. Glasberg, and B. C. J. Moore, ‘A new model for calculating auditory excitation patterns and loudness for cases of cochlear hearing loss’, Hear. Res., vol. 282, no. 1, pp. 69–80, Dec. 2011, doi: 10.1016/j.heares.2011.09.007.

[31] B. C. J. Moore, B. R. Glasberg, A. Varathanathan, and J. Schlittenlacher, 'A Loudness Model for Time-Varying Sounds Incorporating Binaural Inhibition', Trends Hear., vol. 20, Dec. 2016, doi: 10.1177/2331216516682698.
[32] S. Jørgensen and T. Dau, 'Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing', J. Acoust. Soc. Am., vol. 130, no. 3, pp. 1475–1487, 2011, doi: 10.1121/1.3621502.
[33] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, 'An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech', IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 7, pp. 2125–2136, Sep. 2011, doi: 10.1109/TASL.2011.2114881.
[34] S. Jørgensen, S. D. Ewert, and T. Dau, 'A multi-resolution envelope-power based model for speech intelligibility', J. Acoust. Soc. Am., vol. 134, no. 1, pp. 436–446, Jul. 2013, doi: 10.1121/1.4807563.
[35] J. F. Culling, M. L. Hawley, and R. Y. Litovsky, 'The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources', J Acoust Soc Am, vol. 116, no. 2, pp. 1057–65, Aug. 2004.
[36] S. Jelfs, J. F. Culling, and M. Lavandier, 'Revision and validation of a binaural model for speech intelligibility in noise.', Hear. Res., vol. 275, no. 1–2, May 2011, doi: 10.1016/j.heares.2010.12.005.
[37] T. Leclere, M. Lavandier, and J. F. Culling, 'Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation.', J. Acoust. Soc. Am., vol. 137, no. 6, pp. 3335–3345, Jun. 2015, doi: 10.1121/1.4921028.
[38] C. F. Hauth, S. C. Berning, B. Kollmeier, and T. Brand, 'Modeling Binaural Unmasking of Speech Using a Blind Binaural Processing Stage', Trends Hear., vol. 24, p. 2331216520975630, Jan. 2020, doi: 10.1177/2331216520975630.
[39] L. Prud'homme, M. Lavandier, and V. Best, 'A harmonic-cancellation-based model to predict speech intelligibility against a harmonic masker', J. Acoust. Soc. Am., vol. 148, no. 5, pp. 3246–3254, Nov. 2020, doi: 10.1121/10.0002492.
[40] T. Vicente and M. Lavandier, 'Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises', Hear. Res., vol. 390, p. 107937, May 2020, doi: 10.1016/j.heares.2020.107937.
[41] T. Vicente, M. Lavandier, and J. M. Buchholz, 'A binaural model implementing an internal noise to predict the effect of hearing impairment on speech intelligibility in non-stationary noises', J. Acoust. Soc. Am., vol. 148, no. 5, pp. 3305–3317, Nov. 2020, doi: 10.1121/10.0002660.
[42] M. Lavandier, T. Vicente, and L. Prud'homme, 'A series of SNR-based speech intelligibility models in the Auditory Modeling Toolbox'. Acta Acust, 6, 20, 2022. doi: https://doi.org/10.1051/aacus/2022017.
[43] A. Osses Vecchi and A. Kohlrausch, 'Perceptual similarity between piano notes: Simulations with a template-based perception model', J. Acoust. Soc. Am., vol. 149, no. 5, pp. 3534–3552, May 2021, doi: 10.1121/10.0004818.
[44] T. McKenzie, C. Armstrong, L. Ward, D. Murphy, and G. Kearney. Predicting the colouration between binaural signals. Appl. Sci., 12(2441), 2022.
[45] P. Zakarauskas and M. S. Cynader, 'A computational theory of spectral cue localization', J Acoust Soc Am, vol. 94, pp. 1323–1331, 1993.
[46] E. H. A. Langendijk and A. W. Bronkhorst, 'Contribution of spectral cues to human sound localization', J Acoust Soc Am, vol. 112, no. 4, pp. 1583–96, Oct. 2002, doi: doi:10.1121/1.1501901.
[47] T. May, S. van de Par, and A. Kohlrausch, 'A probabilistic model for robust localization based on a binaural auditory front-end', IEEE Trans Audio Speech Lang Proc, vol. 19, pp. 1–13, 2011.
[48] R. Baumgartner, P. Majdak, and L. Bernhard, 'Assessment of sagittal-plane sound localization performance in spatial-audio applications', in The Technology of Binaural Listening, J. Blauert, Ed. Berlin, Heidelberg: Springer, 2013, pp. 93–119.
[49] E. Georganti, T. May, S. van de Par, and J. Mourjopoulos, 'Sound Source Distance Estimation in Rooms based on Statistical Properties of Binaural Signals', Audio Speech Lang. Process. IEEE Trans. On, vol. 21, no. 8, pp. 1727–1741, Aug. 2013, doi: 10.1109/TASL.2013.2260155.
[50] H. Wierstorf, A. Raake, and S. Spors, 'Binaural Assessment of Multichannel Reproduction', in The Technology of Binaural Listening, J. Blauert, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 255–278.
[51] R. Baumgartner, P. Majdak, and B. Laback, 'Modeling sound-source localization in sagittal planes for human listeners', J. Acoust. Soc. Am., vol. 136, pp. 791–802, 2014, doi: 10.1121/1.4887447.
[52] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans, 'An ideal-observer model of human sound localization', Biol. Cybern., vol. 108, no. 2, pp. 169–181, 2014, doi: 10.1007/s00422-014-0588-4.

[53] D. Kelvasa and M. Dietz, 'Auditory Model-Based Sound Direction Estimation With Bilateral Cochlear Implants.', Trends Hear., vol. 19, 2015, doi: 10.1177/2331216515616378.

[54] R. Baumgartner, P. Majdak, and B. Laback, 'Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane', Trends Hear., vol. 20, p. 2331216516662003, 2016, doi: 10.1177/2331216516662003.

[55] H. G. Hassager, F. Gran, and T. Dau, 'The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment', J. Acoust. Soc. Am., vol. 139, no. 5, pp. 2992–3000, 2016, doi: 10.1121/1.4950847.

[56] S. Li, R. Baumgartner, and J. Peissig, 'Modeling perceived externalization of a static, lateral sound image', Acta Acust., vol. 4, no. 5, Art. no. 5, 2020, doi: 10.1051/aacus/2020020.

[57] R. Baumgartner and P. Majdak, 'Decision making in auditory externalization perception: model predictions for static conditions', Acta Acust, vol. 5, p. 59, 2021, doi: 10.1051/aacus/2021053.

[58] R. Barumerli, P. Majdak, R. Baumgartner, M. Geronazzo, and F. Avenzini, 'Predicting human spherical sound-source localization based on Bayesian inference', Acta Acust., vol. (submitted to), 2022.

[59] P. Lladó, P. Hyvärinen, and V. Pulkki, 'Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation', Acta Acust, vol. 6, p. 1, 2022, doi: 10.1051/aacus/2021056.

[60] G. McLachlan, P. Majdak, J. Reijniers, and H. Peremans, 'Towards modelling active sound localisation based on Bayesian inference in a static environment', Acta Acust, vol. 5, p. 45, 2021, doi: 10.1051/aacus/2021039.

[61] R. Baumgartner and P. Majdak, 'Modeling Localization of Amplitude-Panned Virtual Sources in Sagittal Planes', J Audio Eng Soc, vol. 63, no. 7/8, pp. 562–569, 2015, doi: 10.17743/jaes.2015.0063.

[62] R. Baumgartner, P. Majdak, and B. Laback, 'The Reliability of Contralateral Spectral Cues for Sound Localization in Sagittal Planes', presented at the Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, USA, 2015.

[63] L. Steidle and R. Baumgartner, 'Geometrical Evaluation of Methods to approximate Interaural Time Differences by Broadband Delays', in Fortschritte der Akustik, Rostock, 2019, pp. 368–370.

[64] I. Engel, D. F. M. Goodman, and L. Picinali, 'Assessing HRTF preprocessing methods for Ambisonics rendering through perceptual models', Acta Acust, vol. 6, p. 4, 2022, doi: 10.1051/aacus/2021055.

[65] A. Osses et al., 'A comparative study of eight human auditory models of monaural processing', Acta Acust 2022, 6, 17, doi: https://doi.org/10.1051/aacus/2022008.

[66] S. Röttges, C. F. Hauth, T. Brand, and J. Rennies-Hochmuth, 'Challenging a non-intrusive EC-mechanism: Modelling the Interaction between binaural and temporal speech processing', Acta Acust., vol. (submitted to), 2022.