Aural assessment by means of binaural algorithms −The AABBA project−

Jens Blauert[1], Jonas Braasch[2], Jörg Buchholz[3], H. Steven Colburn[4], Ute Jekosch[5], Armin Kohlrausch[6], John Mourjopoulos[7], Ville Pulkki[8], and Alexander Raake[9]

[1] *Inst. of Communication Acoust., Ruhr-Univ. Bochum, Bochum, Germany*

[2] *Ca³rl Lab., Rensselaer Polytechn. Inst, Troy NY, USA*

[3] *Ctr. for Appl. Hearing Research, Techn. Univ., Lyngby, Denmark*

[4] *Hearing-Research Ctr., Boston Univ., Boston MA, USA*

[5] *Chair of Communication Acoust., Techn. Univ. Dresden, Dresden, Germany*

[6] *Techn. Univ. Eindhoven and Philips Research Europe, Eindhoven, The Netherland*

[7] *Wire Communication Lab., Patras Univ., Patras, Greece*

[8] *Deptm. of Signal Proc., and Acoust., Helsinki Univ. Techn., Helsinki, Finland*

[9] *Techn. Univ. Berlin and Deutsche Telekom Labs., Berlin, Germany*

AABBA is an intellectual grouping with the goal of collaborating for 4−5 years on the application of computational models of human binaural hearing. The grouping will compile a battery of common software components for setting-up dedicated full-scale binaural models for various technological applications − particularly in areas related to communication acoustics. Examples are: quality assessment of audio channels and loudspeakers as well as rooms for acoustical performances, assessment of disorders of binaural hearing, assessment of speech-understanding capabilities in acoustically adverse surroundings, auditory-scene mapping, assessment of spatial properties of product sounds as well as of the sense of envelopment and immersion, analysis of human spatial hearing in a multimodal world with, for example, the listeners moving in space and/or receiving additional visual and/or tactile cues.

INTRODUCTION

Computational models of binaural hearing have been developed and discussed since the advent of computers in acoustics, that is, since the late sixties of last century. For current reviews of the state of the art see, e.g., Colburn (1996), Blauert (1997), Braasch (2005). Although the scientific curiosity to model the auditory system is still unbroken, and many scientist are actually engaged in developing these models further, it has not yet been commonly realized that these models have a promising potency for technological application.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

To exploit on this fact, the *AABBA* consortium has decided to join forces to explore a number of dedicated applications of binaural models. Although this endeavour requires the further development of existing binaural models, for the time being it is agreed within this project that the main focus will not be on model building but on actual applications.

All founding member-groups of *AABBA* have a background in auditory modelling – see the list of references below. They will exchange their existing model components and data collections between each other to arrive at a common model-components library. This library will later be made available to the public.

AABBA-project planning, consequently, comprises two main task sections, namely,

- Provide a collection of components of auditory models to be included in a common consistent architecture, that is, a "tool box", and
- Explore and develop specific technological applications, with at least one application per member being planned so far.

By the way, *AABBA* is not a closed circle – new members can be accepted by decision of the steering committee, which is formed by the current member groups.

THE "TOOLBOX"

A library of sufficiently documented computer programs in MATLAB, which enables users with a basic knowledge in binaural hearing to compose models of binaural signal processing – tailored to the topics of aural assessment in their given application scenarios is envisioned. The library will include signal-driven as well as hypothesis-driven algorithms on both the signal and the symbolic level, with the aim of including expert systems that draw upon multimodal information and cognition. It will further include a collection of data as needed to run the models.

A provisional listing of readily available components is given in the following,

- **External and middle ear:** artificial heads, HRTF catalogues, HRTF-measurement capabilities, middle-ear models
- **Inner ear**: basilar-membrane models, ear-filter banks, hair-cell models
- **Monaural preprocessing:** echo canceller, masking models, models for signal-dependent-compression
- **Binaural interaction:** Jeffress-type models, extended Jeffress models (e.g., with contra-lateral inhibition), Durlach-type EI models, combined EE/EI models, count-comparison models, Grothe-type models, and
- **Feature extraction from binaural-activity maps:** horizontal position, ILDs, ITDs, running coherence, reverberation effects, auditory spaciousness, parameters for spatial coding

Input	acoustic signals
EXTERNAL EAR	HRTFs
Interface 1	binaural ear signals
MIDDLE EAR	program module(s)
Interface 2	modified binaural signals
COCHLEA	program module(s)
Interface 3	multi-channel signals
MONAURAL PREPROCESSING	program module(s)
Interface 4	multi-channel signals
BINAURAL PROCESSING	program module(s)
Interface 5	ILDs, ITDs, k(t), $\underline{S}(\omega)$, ect.
BINAURAL ACTIVITY MAPS	modules(s)
Interface 6	3D data representation
PRESEGMENTATION	program module(s)
Output	enhanced 3-D representation

Fig. 1: Modules and interfaces defined within the bottom-up section of the *AABBA* model k(t)… interaural coherence, S(ω)…binaural and interaural spectral cues.

In a first step, a common architecture is set up to host the different model components which already exist at the partners sites. Where components with similar functionally are available, it will be arranged that users can switch between different components within the same functional module. To this end, data contents and data structures are to be specified at the interfaces between functional modules. Figure 1 illustrates the idea as far as existing components are concerned.

APPLICATION ACTIVITIES

Application projects have been defined individually by each of the nine member groups. It is envisaged that each group reports on their relevant R&D results in the form of a book chapter – besides other appropriate reporting. In this way, the proceedings of *AABBA* will be available to the public right after the project ends.

Before each of the member groups decided on their specific field of application, the following generic application areas for binaural models had been identified and discussed among partners.

- **Audio technology:** binaural-cue selector, quality assessment of audio channels, quality assessment of loudspeakers, automatic surveillance of transmission quality.
- **Audiology:** assessment of disorders of binaural hearing, assessment of binaural dereverberation and binaural decolouration, assessment of speech-understanding capabilities in acoustically adverse surroundings, binaural-loudness meter.
- **Aural virtual environments:** auditory-scene mapping, identification of virtual sources, assessment of the perceived room size.

- **Hearing aids:** fitting of binaural hearing aids, diagnosis of dysfunctions of hearing aids.
- **Product-sound quality:** assessment of spatial properties of product sounds.
- **Room acoustics:** echo detector, spaciousness meter, detectors of image shifts, assessment of the sense of envelopment and immersion, assessment of the *precedence effect,* assessment of a global "quality of the acoustics".
- **Speech technology:** speaker-position mapping, binaural speech intelligibility, assessment of speech recognition in adverse acoustical conditions, assessment of the *cocktail-party effect.*
- **Binaural models as a research tool:** to be employed for the evaluation and assessment and analysis of human spatial hearing in a multimodal world, e.g., with the listener moving in space and/or receiving additional visual and/or tactile cues.

At the end of this selection process it turned out that the consortium, in the application section of *AABBA*, will speficifically focus on the following main areas of activity,

- **Spatial scanning and mapping of auditory scenes:** Estimation of the position and the spatial extents of auditory events which form an auditory scene – be it a natural scene as in room acoustics or virtual scene as in virtual-reality applications, or at the play-back end of audio systems – including spatially diffuse auditory events, often perceived as components of reverberance.
- **Analysis of auditory scenes with the aim of deriving parametric representations at the signal level.** Estimated these parameters may be intended to be used, e.g.,
  - For coding and/or re-synthesis of auditory scenes,
  - For speech-enhancement in complex acoustic environments (incl. hearing aids),
  - For systems to enhance the spatial perception in sound fields, (such as better localization and/or a better sense of envelopment; further, decoloration and dereverberation), or
  - For the identification of perceptual invariances of auditory scenes.
- **Evaluation of auditory scenes in terms of "quality",** where "quality" will strictly be judged from the user's point of view, e.g.,
  - Quality of "the acoustics" of spaces for musical performances,
  - Quality of systems for holophonic representation of auditory scenes, such as auditory displays and virtual-reality generators,
  - Spatial quality of audio-systems (for recording, transmission and play-back) incl. systems that employ perceptual coding, or

  - Performance of speech-enhancement systems (incl. hearing aids).

- **Analysis of auditory scenes with the aim of deriving parametric representations at the symbolic level :** e.g.,
  - Identification of determinants of *meaning* contained in binaural-activity maps, or
  - Assignment of meaningful symbols to the output of binaural models.

QUALITY OF EXPERIENCE

In communication technology the term "Quality of Experience" (QoE) has recently become a kind of a "buzz word". In contrast to the more traditional concept of "Quality of Service" (QoS), QoE comprises those aspects of the quality of systems that cannot be assessed with traditional instrumental methods. In other words, the term QoE denotes the perceptual quality of systems. A lot of effort is currently put into the task of simulating the perception process, with the aim of arriving at predictions of perceptual quantities, which might then be used to characterize aural percepts and, in a further step, help predict the related perceptual quality.

Following this way of thinking, algorithms to estimate perceptual attributes like loudness, pitch, sharpness and speech intelligibility (e.g., STI) have been derived, and further, more complex ones such as estimates for the perceptual quality of speech (e.g., PESQ) and music (e.g., PEAQ) are in use. Additional algorithms have been proposed and/or are available for other sensory domains, e.g., for video and/or for multimodal systems.

Most of the *AABBA* members plan to include "QoE" evaluation and assessment in their application studies, more specifically, by making use of binaural models as a part of quality-prediction algorithms.

THE *AABBA* APPROACH TO QUALITY EVALUATION

In *AABBA* we make a distinction between the *character* of a sound, which denotes a profile of all its nameable and measurable attributes – such as acoustic, auditory, emotional, Gestalt and semantic ones − and the *quality* of the sound. Quality judgements evolve from a comparison which listeners perform in the actual use situations, namely a comparison of the sound character with a set of references that the listeners have in mind. A possible rough model of the quality-assignment process suitable for engineering purposes is sketched in Fig. 2 (Blauert and Jekosch 2003). An interesting feature of this modelling approach is that the individuality and the task-specificity of the quality-assignment process are taken care of in the reference module of the model. The character-evaluation module can thus be left fairly general.

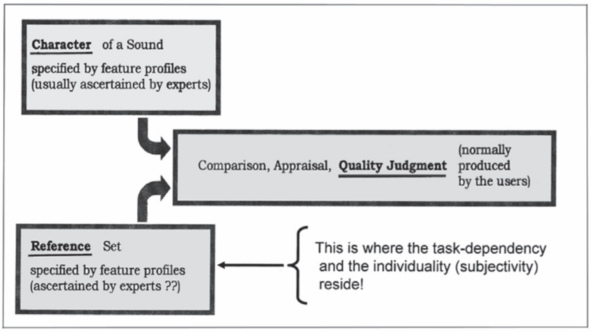

Fig. 2: An "engineering" approach to the quality-assignment process.

The reference sets used for quality judgments may have different levels of complexity. In an attempt to put some order into this issue, Blauert and Jekosch (2007) have proposed to arrange the references according to the amount of intellectual abstraction involved. Figure 3 depicts this idea. Starting from simple perceptual attributes, the amount of abstraction increases to instrumentally measurable attributes, then to Gestalt attributes and finally to content-related attributes. When trying to define the abstraction levels by the measuring methods applied to evaluate the attributes at each particular level, it becomes obvious that the higher the level of abstraction, the more knowledge is involved in the assessment procedures.

To illustrate how the different reference attributes relate to the different levels of abstraction, we take the aural quality of musical performances as an example, be it in actual rooms like concert halls or be it mediated via audio systems like surround-sound systems.

- The lowest level of abstraction is given when basic psycho-acoustic attributes are judged, such as loudness, pitch and sharpness. This is the case when the *sound quality* concerned is the quality of the auditory event *as such*, i.e. in its pure, non-interpreted form. The assessment methods in this case are basic psychoacoustical ones. In other words, they require listening tests which employ basic psychometric methods, where the evocation of higher abstraction layers of perception is intentionally excluded.
- A higher level of abstraction occurs in connection with instrumental (… physical) measurements. Instrumental measurement methods have been developed for providing ultimate objectivity, i.e. independence of the results from a particular laboratory or experimenter. This is the kind of data engineers would generally prefer – if this were feasible. These kinds of measurements are used, for example, when the *quality of transmission* is the issue of interest. Physical (…acoustical) reference data may then be available. Please note that spaces like concert halls may be treated as transmission systems for acoustic signals.

- The next higher level of abstraction is denoted *auditory-scene quality* here − *aural-gestalt quality* could be an alternative label. What is actually meant, is the degree as to which the *presentation* or *Gestalt* of the auditory scene meets the reference. Hereby, it is not necessarily aimed at authentic reproduction, rather terms like plausibility, perceptual coherence, presence and immersion become issues of interest. The measurement methods to capture these issues are perceptual ones, though more complex than those applied under the sound-quality level – yet, surely also cognitive ones.

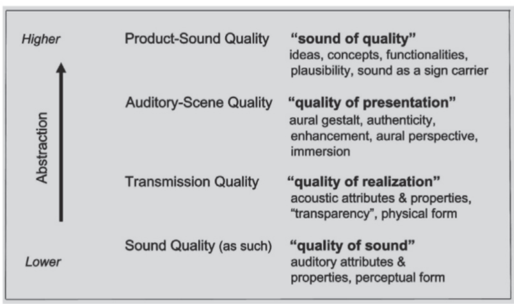

Fig. 3: Levels of abstraction regarding references employed in sound-quality assessment.

- An even higher level of abstraction is reached when the role of sounds as signs comes to the fore. This is, for example, the case for the overall quality of musical pieces in the performance situation but also for *product-sound quality.* Product sounds are considered to have a high quality when they are able to convey to the potential or actual product user that the product itself has a high quality. In these cases, it is not the sound quality as such that is the issue of interest. Rather, the focus lies on the fact that the product sound supports the quality of the product by acting as a sign for the actual product quality, that is, the content. Note that this holds for the sound of industrial products as well as for performed musical pieces. The methods to measure these effects and to transform the results into profiles to be included into references for quality assessment draw heavily upon cognition. It certainly takes expertise in cognitive psychology to manage these tasks.

When applying these ideas to binaural models, it appears that the current architecture of binaural models must be expanded to be able include knowledge and to set up hypotheses based on this knowledge. To this end, it is envisaged to complement the binaural model with a "brain", that is, expert components which "interpret" the output of the lower, signal-driven sections of the model. Figure 4 shows how the architecture of an augmented system must be structured, however, the final architecture has not yet been decided. Note that software architectures of this or similar types have already been discussed in artificial intelligence and/or speech technology, e.g., in the context of instrumental speech-recognition systems.

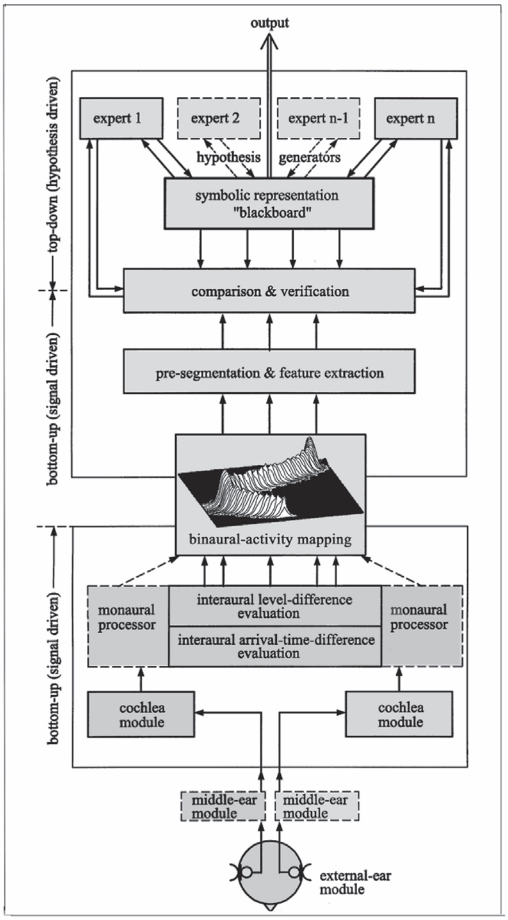

Fig. 4: Enlarged architecture of a binaural model including knowledge-based components.

CONCLUSION

The *AABBA* consortium aims at putting current intellectual knowledge as is available in the field of models of the binaural system to work in eight selected, real technological application scenarios. As far as possible, a common set of tools will be employed for this purpose.

However, for each individual application, domain-specific knowledge and domain-specific interpretation algorithms have to be incorporated in the respective model algorithms. To this end the general architecture of the binaural model will be expanded by a knowledge-based overall structure.

The model as a whole will attempt to simulate bottom-up signal processing in the subcortical monaural and binaural pathways of the auditory system as well as hypothesis-driven processing as attributed to the cognitive parts of the central nervous system – the latter at least as far as needed in the specific application areas of *AABBA* (Fig. 5).

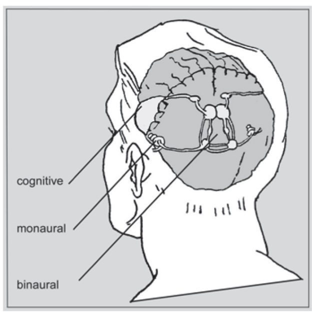

Fig. 5: Sections of the auditory pathway that are touched upon by our modelling efforts.

The *AABBA* members are well aware of the fact that the sub-cortical auditory system obviously determines *where* and *how* sounds appear in the perceptual world. Yet, the reaction of listeners to sounds surely depends primarily on what these sounds *mean* to them in their actually life situation. Consequently, for actual application of binaural models to real-life tasks, assignment of *meaning* must be taken into account.

REFERENCES

These references represent a selection of literature that *AABBA* members consider relevant as evidence for their specific experience and way of thinking of binaural modelling and Quality of Experience.

Blauert, J., and Braasch, J. (**2007**). "Räumliches Hören". In *Handbuch der Audiotechnik*, edited by S. Weinzierl (Springer, Berlin–Heidelberg–New York), pp. 87-122.

Blauert, J., and Guski, R. (**2009**). "Critique of 'pure' psychoacoustics", Fortschr. Akust. NAG/DAGA'09, 1518−1519, Dtsch. Ges. Akust., D-Berlin.

Blauert, J., and Jekosch, U. (**2003**). "Concepts behind sound quality: some basic consideration", Proc. InterNoise`03, 72-76, KR-Jeju Isld.

Blauert, J., and Jekosch, U. (**2007**) "Auditory quality of concert halls – the problem of references". Proc. 19th Int. Congr. Acoust., ICA 2007 Madrid, paper RBA 06−004, Revista de Acústica **38**, ES–Madrid.

Blauert, J. (**1997**). *Spatial Hearing − The psychophysics of human sound localization*, 2nd enhanced edition (MIT, Cambridge, MA).

Blauert, J. (**1999**). "Models of binaural hearing: architectural considerations" , In: *Proc. 18th DANAVOX Symposium 1999*, edited by A. N. Rasmussen *et al.* (DANAVOX Jubilee Foundation, DK−Copenhagen), 189-206.

Braasch, J. (**2005**). "Modelling of binaural hearing", in *Communication Acoustics*, edited by J. Blauert (Springer, Berlin–Heidelberg–NewYork), pp. 75-108.

Buchholz, J. M., and Mourjopoulos, J. (**2004a**). "A computational auditory masking model based on signal-dependent compression. I. Model descriptions and performance analysis," Acta Acustica united with Acustica **90**, 873-886.

Buchholz, J. M., and Mourjopoulos, J. (**2004b**). "A computational auditory masking model based on signal-dependent compression. II. Model simulations and analytical approximations," Acta Acustica united with Acustica **90**, 887-900.

Colburn, S. H. (**1996**). "Computational models of binaural processing. In:, *Handbook of auditory research: Auditory computation,* edited by. Fay, R. R., and Popper, A. N. (Springer, New York), pp. 332-400.

Hess, W., Braasch, J., and Blauert, J. (**2003**). "Acoustical evaluation of virtual rooms by means of binaural-activitity patterns", 115th Conv. Audio Engr. Soc., preprint (Audio Engr. Soc., New York NY, USA)

Jekosch, U. (**2004**). "Basic concepts and terms of quality − reconsidered in the context of product-sound quality," in acta acustica united with Acustica **90**, 999-1006.

Jekosch, U. (**2005a**). "Assigning meaning to sounds − Semiotics in the context of product-sound design," in *Communication Acoustics,* edited by J. Blauert (Springer, Berlin–Heidelberg–NewYork), pp. 193-221.

Jekosch, U. (**2005b**). *Voice and speech quality perception.− assessment and evaluation* (Springer, Berlin−Heidelberg−New York).

Jepsen, M. L., Ewert, S. D., and Dau. T. (**2008**). "A computational model of human auditory signal processing and perception," J. Acoust. Soc. Am. **124**, 422-438.

Pulkki, V., and Hirvonen, T. (**2009**). "Functional count-comparison model for binaural decoding," Acta Acustica united with Acustica **95**, 883-900.

Pulkki, V. (**2007**). "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc. **55**, 503-516.

Raake, A. (**2002**). "Does the content of speech influence its perceived sound quality?". Proc. 3rd Int. Conf. on Language Resources and Evaluation (LREC 2002), **4**, 1170-1176, ES−Las Palmas.

Raake, A. (**2006**). *Speech quality of VoIP – assessment and prediction* (John Wiley and Sons Ltd, UK).

Spors, S., and Ahrens, J. (**2008**). "A comparison of wave-field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," 125th AES Conv. USA−San Francisco CA.

Spors, S., Wierstorf, H., Geier, M., and Ahrens, J. (**2009**). “Physical and perceptual properties of focused sources in wave-field synthesis,” 127th AES Conv. USA−New York NY.

Stern, R. M., and Colburn, H. S. (**1978**). “The theory of binaural interaction based on auditory-nerve data, IY: A model for subjective lateral position,” J. Acoust. Soc. Am. **64**, 127-140.

Wan, R., Durlach, N.I., and Colburn, H.S. (**2009**). “Modified EC model applied to speech intelligibility with spatially distributed maskers,” *submitted* to J. Acoust. Soc. Am.

Zarouchas, T., and J. Mourjopoulos (**2009**) “Modeling perceptual effects of reverberation on stereophonic sound reproduction in rooms,” J. Acoust. Soc. Am. **126**, 229-242.