ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/243770079

Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony

Article · January 1995

CITATIONS
115

READS
2,290

3 authors, including:

Jean-Marc Jot
Virtuel Works LLC
57 PUBLICATIONS **1,459** CITATIONS

SEE PROFILE

Veronique Larcher
Sennheiser GmbH
4 PUBLICATIONS **146** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project BiLi Project View project

Project Reverb View project

All content following this page was uploaded by Jean-Marc Jot on 08 October 2017.

The user has requested enhancement of the downloaded file.

Meta Platforms, Inc.
Exhibit 1076
Page 001

Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony.

Jean-Marc Jot, Véronique Larcher** and Olivier Warusfel*,*
**IRCAM, Paris, France, **Télécom Paris, France.*

Presented at
the 98th Convention
1995 February 25 - 28
Paris

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Meta Platforms, Inc.
Exhibit 1076
Page 002

Digital signal processing issues in the context of binaural and transaural stereophony

Jean-Marc Jot [†#], *Véronique Larcher* [‡], *Olivier Warusfel* [#]

[#] IRCAM, 1 pl. Igor Stravinsky, 75004 Paris, France
[†] Espaces Nouveaux, 56 bd. Davout, 75020, Paris, France
[‡] Télécom Paris, dépt. Signal, 46 rue Barrault, 75634 Paris, France

e-mail: jmjot@ircam.fr

ABSTRACT

Signal processing aspects of the modeling of head-related transfer functions (HRTFs) are examined, with application to the real-time mixing and synthesis of two-channel signals for headphone or loudspeaker reproduction of three-dimensional sound images. The merits of IIR and FIR implementations of the synthesis filters are discussed and efficient solutions are proposed, including time-varying implementations (allowing head-tracking or the simulation of moving sound sources). Further modeling of early reflections and the use of feedback delay networks for reproducing the later diffuse reverberation make it possible to include accurate binaural room effect synthesis in the simulation, without exceeding the capacity of recent programmable digital signal processors.

0 INTRODUCTION

The purpose of the present paper is to describe some principles allowing to implement a binaural or transaural spatial processor for a minimal computational cost, while attempting to preserve maximal fidelity of the reproduction. This study has been carried out within the Spatialisateur project, conducted by Espaces Nouveaux and IRCAM, in collaboration with the Centre National d'Etudes des Télécommunications, and prolonging earlier work carried out at Télécom Paris [1-4]. The focus is not made here on the experimental aspects of HRTF measurements or the validation of the simulation. The psycho-experimental validation of the techniques presented herein is currently in progress, using the real-time binaural Spatialisateur described in [5], and preliminary results are presented in [6]. The Spatialisateur was developped in the Max object-oriented graphical signal processing software environment, and is available as a Max object named *Spat~* running in real time on the IRCAM Musical Workstation [7].

0.1 Principles of binaural and transaural synthesis

Binaural synthesis is a process by which, from a primary monophonic recording of a source signal, a three-dimensional sound image can be reproduced on headphones. The monophonic signal is transformed by two linear filters in order to reconstruct, at the listener's ears, the pressure signals that would be captured by the eardrums in a natural listening experience, in the presence of a real sound source (Fig. 1). This process is equivalent to synthesizing a dummy-head recording (appropriate for headphone reproduction) from a conventional microphone signal. Early attemps to mix

Meta Platforms, Inc.
Exhibit 1076
Page 003

conventional microphone recordings with dummy-head recordings were made in the 1970's and led to the concept of "binaural mixing console": a mixing console where each sound channel is processed trough a directional filter ("binaural panpot"), allowing to give each source signal a chosen apparent direction on a sphere around the listener (see e.g. [8, 9]).

The two "head-related transfer functions" (HRTFs), which realize a two-channel directional filter associated to a particular direction of incidence, account for the transformation undergone in free field by a sound along its path from a source to the listener's ear canals (this transformation is caused by the diffraction of the sound on the listener's pinnae, head, shoulders and torso). A database of HRTFs can be measured using a dummy head and a loudspeaker in an anechoic room, by sampling a number of directions on a sphere around the dummy head (the dummy head can be replaced by a human with small microphones inserted in his/her ears canals). To simulate a free-field listening situation, the binaural synthesis can be performed by convolving a source signal with the pair of HRTFs measured for a given direction of incidence, and including an additionnal equalization filter for compensating the coupling of the earphones to the ears (Fig. 1).

The same process can be performed for any natural environment (e.g. a concert hall) instead of the anechoic room. In this case, the pair of measured transfer functions (or impulse responses) will contain the effect of room reflections and reverberation and this room effect will be faithfully reproduced over headphones, by convolving a pre-recorded monophonic source signal with the measured binaural impulse response. However, one must be aware that the reproduced listening experience will be that of the loudspeaker emitting the monophonic signal in the room (instead of the original natural sound source radiating in the room). The differences between the directivity patterns of the natural source and the loudspeaker can have a dramatic effect on the result of the simulation, particularly on the subjective source distance [10]. This is a serious limitation of binaural simulation using measured binaural room responses.

A more powerful approach to the binaural simulation of complex sound fields is offerred by combining the free-field simulation process (using the free-field HRTF database) with the principle of superposition. An approximation of the binaural impulse response can be computed from a physical description of the room and obstacles. Geometrical models of sound propagation in rooms (the image source or ray-tracing methods, see e.g. [11]) can be extended to incorporate, along with the modeling of wall reflections and air absorption, a model of the source directivity (which can be described in the same way as the HRTF database describes the directivity of the receiver). An additionnal advantage of this synthesis method over convolution by measured binaural room responses is that it allows to construct virtual spaces containing one or several virtual sound sources with possible variations in time (movements or any transformations of the sources, the room or obstacles). However, the validity of this synthesis method depends on the accuracy of the models used to predict the propagation of sounds in rooms. The validation of these models is still an open field for research (see e.g. [12]) and its discussion is beyond the scope of this paper.

Dummy-head signals or binaural signals obtained by the above methods can be reproduced over headphones or further converted into loudspeaker-compatible signals, by means of a "cross-talk cancelling" process assuming that the listener will be placed at a particular position with respect to the loudspeaker pair. This loudspeaker reproduction technique, initially invented by Schroeder and

Meta Platforms, Inc.
Exhibit 1076
Page 004

Atal [13, 14] was further optimized by Cooper and Bauck [15], who coined the term "transaural" for this reproduction mode.

0.2 Applications of binaural and transaural synthesis

Due to the fact that spatial cues are conveyed by reproducing the signals at the listener's ears rather than recreating a sound field in an extended listening space, binaural techniques are, in essence, individual reproduction techniques: they are not intended for adressing a large audience with a loudspeaker system (see e.g. [16, 17] for details on these limitations and some extensions or alternative approaches). Despite this limitation, the potential applications of binaural and transaural synthesis are found in many different contexts where the reproduction of spatial sound information is of importance.

According to the application, it may or may not be necessary to process sounds in real time or to realize a time-varying implementation allowing the simulation of moving sources. Some applications will require the synthesis of room reflections and reverberation. An artificial room effect should be included if the apparent positions of virtual sounds sources are to cover a genuinely *three-dimensional* space, because the perception of auditory distance is not reliable in anechoic listening conditions (see e.g. [8]). When an artificial room effect is necessary, it may be sufficient to reproduce a small number of early reflections in some cases, while other applications may require the accurate synthesis of a full room effect, including the diffuse later reverberation. Some possible application fields of binaural and transaural synthesis are classified below in view of these considerations.

- *Studio recording and post-production.* A binaural mixing console can be used for producing spatial effects transgressing the limitations of conventional stereo recording techniques or for allowing the use of auxiliary microphones in live dummy-head recordings [18-23]. Another application of binaural synthesis is for improving the headphone reproduction of conventional stereo recordings [24, 25]. Real-time signal processing is needed, and continuous adjustment of source localization is desirable for mixing applications. A limited number of discrete spatialized room reflections can be simulated with a binaural mixing console including a delay line in each channel, at the cost of devoting an additional channel to each synthetic reflection. Using an external artificial reverberator to add the late diffuse reverberation requires precautions for maintaining binaural compatibility, and offers limited dynamic control flexibility because of the non-homogeneity of the processing equipment.

- *Virtual reality, computer music, telecommunications and advanced human-computer interfaces* [26-34]. Systems for these applications are essentially equivalent to binaural mixing consoles, with more stringent requirements on time-varying processing in order to ensure natural interactivity or the simulation of moving sources. As with binaural mixing consoles, only a limited number of binaural room reflections can be synthesized in real-time, yet with more flexibility and accuracy [26]. Some computer music applications mention an artificial reverberation module for rendering diffuse reflections [30, 32], but do not mention methods for ensuring accuracy, naturalness and binaural compatibility of the synthetic reverberation.

Meta Platforms, Inc.
Exhibit 1076
Page 005

- *Auralization in architectural acoustics and psychoacoustics* [12, 35-39]. In current auralization systems for architectural acoustics, real-time constraints are relaxed and the focus is on the accurate reproduction of the transformations undergone by the sound during its propagation in the room. The techniques used are again based on the principle of superposition, yet this superposition is performed through the process of computing a global binaural room impulse response (which is then stored for future analysis or convolution with anechoic signals), rather than mixing binaural elementary signal components (direct sound and reflections).

0.3 Factors which influence the fidelity of the reproduction

Since the perceptual requirements vary from an application to another, the psycho-experimental validation of models and practical compromises may need to be carried out for each specific application. However, some general elements are provided by experimental results already published [40-44], which have been confirmed in our experience.

The major challenges to binaural or transaural reproduction include:
- out-of-head localization of virtual sound sources in headphone simulation,
- minimization of front-back reversals and faithful reproduction of source elevation,
- accurate and natural-sounding reproduction of the room effect.

The major factors wich influence the success of the simulation include :
- the techniques used for measuring and modeling the HRTFs,
- headphone equalization and reproductibility of headphone donning,
- variations of the HRTFs between individuals,
- the listener's free-field localization performance,
- the possibility of tracking listener movements during reproduction,
- interference of non-auditory (e.g. visual) information or cognitive cues,
- the presence of a synthetic room effect in the simulation,
- the techniques used for synthesizing the room effect.

Intracranial localization and front-back reversals are key challenges to headphone applications, and may be influenced by all factors in the lengthy list above. It is difficult, with the current knowledge, to establish a hierarchy in the importance of these factors. Front-back reversals and the uncertainty of elevation judgements are natural limitations of our auditory system (due to the fact that only two probes are used to sample the sound field). These ambiguities are essentially resolved, in natural listening, by the directional dependence of diffraction effects on the pinnae (reproduced in binaural synthesis), and by the variations of sound pressure signals at the two ears in connection with head movements (see e.g. [8]).

This suggests that a decisive improvement can be obtained by using a 'headtracking' system in the simulation to monitor the listener's head position (particularly orientation) in real-time and compensate his movements in the binaural synthesis process, in order to make the apparent positions of sound sources in the virtual space independant of listener movements [27]. This technique is expected to enhance out-of-head localization and make the success of the simulation less dependent on other factors (e.g. the use of non-individual or modelled HRTFs, headphone equalization). Headtracking imposes strong real-time constraints on the time-variant binaural synthesis process

Meta Platforms, Inc.
Exhibit 1076
Page 006

to minimize the total latency of the directional compensation, which should be less than about 50 ms to 100 ms [27].

Although intracranial localization is not a limitation encountered with transaural listening on loudspeakers, our experience indicates that this technique imposes very strong constraints on the position and orientation of the listener's head. These constraints are somewhat relaxed if the virtual positions of primary sound sources are restricted to the frontal sector, in which case listener movements are not more restricted than in conventional stereo reproduction (see also [15-17]). Possible solutions to update the cross-talk cancellation process include headtracking or adaptive filtering techniques such as in echo cancellation applications (requiring a "microphone headset" to monitor signals at the listener's ears). The practical advantages of such techniques over headphone reproduction with headtracking have yet to be assessed. Finally, an additional limitation of transaural reproduction, as with any loudspeaker reproduction technique, arises from the reflections and reverberation in the listening room, which, in theory, should be cancelled electronically in order to guarantee a faithful reproduction (in this context, see e.g. [45]).

0.4 Cost effective implementation of binaural and transaural synthesis

The successful application of binaural and transaural synthesis depends not only on the fidelity of the reproduction, but also on the possibility of achieving efficient DSP implementation, especially in those contexts where real-time operation is required. This efficiency can be improved by optimization based on perceptually relevant models.

In the following, we first define the specification of the synthesis filters for conversion of a mono signal to binaural or transaural formats -or between these two formats-, making use of the phase properties of the HRTFs and the shuffler structure proposed by Cooper and Bauck [15]. We then proceed to the design and implementation of the synthesis filters for simulating free-field listening on headphones, in the time-varying context (allowing simulation of moving sources or headtracking). This includes a comparison of FIR and IIR implementations of the synthesis filters, with a close attention given to interpolation between directions and commutation of the filters.

Extension from free-field simulation to natural environments becomes rapidly impractical, as mentionned earlier, if each room reflection is to be reproduced in real time as an additionnal virtual sound source, because the processing becomes excessively heavy. This limitation is overcome by further modeling of early reflections and later reverberation, supported by physically and perceptually based models of room reverberation. The parameters of these models can be given by an analysis / synthesis procedure exploiting a measured or computed impulse response, so that the virtual environment can be made to accurately imitate the acoustics of an existing room. This then allows one to mix the processed signals with an actual binaural recording made in that room.

It is shown that the proposed combination of digital filter design techniques and artificial reverberation algorithms allows to substantially reduce the complexity of signal and control computations to be carried out in real time. It is concluded that, using recent programmable digital signal processors, the complete binaural or transaural synthesis (including reflections and reverberation) can be implemented with a single chip per source, while maintaining sufficient accuracy for many of the envisionned applications.

Meta Platforms, Inc.
Exhibit 1076
Page 007

1 SPECIFICATION OF DIRECTIONAL FILTERS FOR FREE-FIELD SIMULATION

1.1 Measurement and properties of HRTFs

The experimental aspects of HRTF measurements are discussed e.g. in [8, 46-48]. In the Spatialisateur project, a database of HRTFs has been collected from 20 subjects for research purposes [2]. For each subject, this database consists of 49 measurements for different azimuths and elevations: two horizontal planes (elevation 0° and 30°) were sampled every 15°, and an additional measurement is taken at 90° elevation. The HRTFs were measured with maximum-length sequences as a test signal at a sampling frequency of 48kHz, and the impulse responses were recovered by use of the fast Hadamard deconvolution algorithm [49-51]. The signal was emitted by a concentric two-driver loudspeaker, and recorded with electret microphones inserted at the entrance of the ear canals. HRTFs measured in the horizontal plane are presented on Fig. 2 and Fig. 3 for a particular subject. An additional binaural measurement was made with the test sequence feeding the earphones (the microphones being left at the same position inside the ear canals), and a measurement of the loudspeaker response was made with an omnidirectional microphone at the place of the listener's head.

The following properties can be exploited for designing the binaural and transaural synthesis filters [3, 4]:
a) A general property of rational linear filters (pole-zero filters): every stable filter can be decomposed into the series association of a minimum phase filter and an all-pass filter. This all-pass filter realizes an "excess phase", which is obtained by subtracting from the phase frequency response its minimum-phase part, derived from the logarithm of the magnitude frequency response using the inverse Hilbert transform [52]. This decomposition is illustrated on Fig. 3a.
b) A particular property of HRTFs: in the case of HRTFs, the all-pass filter is approximately equivalent to a pure delay (the excess phase of the HRTF is a linear function of frequency, at least below roughly 8 to 10 khz) [53, 15]. This delay is estimated by linear curve fitting on the excess-phase response between 1 kHz and 5 kHz. This method provides an estimation of the interaural delay that is somewhat more robust than methods based on threshold detection or on cross-correlating the left and right impulse responses (Fig. 3b).

1.2 Diffuse-field normalization of HRTFs and headphone calibration

Some normalization must be performed on the HRTFs in order to eliminate the effects of the transducers: loudspeaker, microphones and headphones. A faithful reproduction can theoretically be obtained by deconvolving each HRTF by the loudspeaker response and the headphone-to-microphone response for the corresponding ear. This deconvolution can be performed by division in the frequency domain, provided that the loudspeaker and headphone transfer functions be minimum phase (in the present case, it turns out that they can also be approximated by a minimum-phase filter cascaded with a pure delay). This normalization method, however, is not completely satisfying, since it leads to an HRTF database that will either be "pre-deconvolved" by the transfer functions of a particular pair of earphones, or will contain the effects of the microphones used for the mea-

Meta Platforms, Inc.
Exhibit 1076
Page 008

surements. This problem is similar to the equalization problem adressed in the design of dummy heads [8], for which two approaches have been proposed:

a) *Normalization with respect to a given reference direction.* This leads to a database of "monaural HRTFs" [8], where each HRTF is divided by the reference HRTF measured in the same ear. The reference direction is typically chosen to be the frontal direction in the horizontal plane (0° azimuth and 0° elevation) since many headphones are equalized with reference to this direction. This normalization eliminates all effects due to the transducers, and, more generally, all effects which are independent of the direction of sound incidence, including the ear canal resonance. It also eliminates the effects of possible dissymetries in the microphone frequency responses or in their placement inside the ear canals, restoring the natural symmetry which is expected from HRTFs due to the approximate symmetry of the head.

b) *Diffuse-field normalization.* This normalization has the same advantages as the monaural normalization, with the additional benefit of not priviledging a particular direction. The diffuse field HRTFs are derived from the power transfer functions which would be measured in the two ears in diffuse field conditions. In practice, the diffuse field HRTF is estimated in each ear by power-averaging the frequency responses of all the HRTFs measured in that ear and taking the square root of this average spectrum. If the measured directions do not sample the sphere uniformly, each HRTF can be given a weight in the avering process (proportional to a solid angle associated to the corresponding direction) [54].

According to [8], diffuse-field normalization was proposed by Theile in 1981 as a general solution to the problem of equalizing dummy-head recordings. The advantage of this solution is that it preserves the tone color of room reverberation as it is captured by microphones used in conventional stereophony (an omnidirectional microphone with a flat frequency response is naturally equalized to diffuse field). In listening tests, this appeared to be the best compromise for ensuring "loudspeaker compatibility" of dummy-head recordings: ensuring that the tone color is well rendered in conventional stereo reproduction using two loudspeakers, although the three-dimensional quality of the recording cannot be fully preserved (because of the cross-talk from each loudspeaker to the opposite ear).

It follows that diffuse-field normalization of the HRTF database leads to synthetic binaural signals with the following properties:
- timbre compatibility with conventional microphone recordings and with conventional stereo reproduction on loudspeakers,
- compatibility with dummy-head recordings equalized to diffuse field,
- compatibility with reproduction on diffuse-field calibrated headphones (i.e. equalization of the earphone-to-eardrum coupling can be avoided, unless optimal listener-specific equalization is desired). Diffuse-field calibration can be generally viewed as a desirable property for headphones, since it yields the optimum reproduction of room reverberation from conventional microphone recordings [55].

Meta Platforms, Inc.
Exhibit 1076
Page 009

1.3 Pre-processing of measured HRTFs

In view of the above properties, the following pre-processing and normalization procedure can be proposed, to be applied to each measured HRTF.

1.3.1 {magnitude, excess phase} representation of transfer functions

Every HRTF is split in the frequency domain into a minimum-phase component and an all-pass component:

$$H(f) = \underbrace{|H(f)| \cdot \exp(j \cdot \varphi(f))}_{\text{<----- minimum phase -->}} \cdot \underbrace{\exp(j \cdot \Phi(f))}_{\text{<-- all-pass -->}} \tag{1}$$

where f is the frequency, $\varphi(f)$ denotes the minimum phase and $\Phi(f)$ denotes the excess phase (which is approximately a linear function of frequency in the present case). The magnitude frequency spectrum and the excess phase spectrum are sufficient to completely characterize the transfer function H(f) because the magnitude $|H|$ and the minimum-phase Φ are uniquely related by the Hilbert transform. Thus we can use the notation:

$$H = \{|H|, \Phi\} \tag{2}$$

(illustrated by Fig. 3a), where $\Phi = 0$ if H is a minimum-phase function.

Since the product of two minimum-phase functions is a minimum-phase function and the product of two all-pass functions is an all-pass function, all convolution or deconvolution operations are computed separately on the minimum-phase component of H (by multiplying or dividing magnitude spectra) and on the all-pass component of H (by adding or subtracting excess-phase spectra, or rather delays in the present case). Thus, a convolution operation appears in usual polar notation:

$$H_1 \cdot H_2 = \{|H_1| \cdot |H_2|, \Phi_1 + \Phi_2\} \tag{3}$$

However, $|H|$ and Φ are not the modulus and argument of a given frequency response, but represent two independent tranfer functions associated in cascade. It follows that all subsequent approximation or modeling procedures can be performed *independently* on $|H|$ or Φ. One such approximation consists in replacing Φ by a linear phase response, as described above. Further modeling will be applied to $|H|$ below and in part 2 of this paper.

1.3.2 Diffuse-field normalization

The diffuse-field HRTF, noted H_o, is only defined by its magnitude spectrum and does not contain any phase information. Yet the diffuse-field normalization of H can be performed by simply dividing the magnitude spectrum $|H|$ by H_o and keeping the excess-phase Φ unchanged, yielding the diffuse-filed normalized HRTF:

$$H_n = \{|H|/H_o, \Phi\} \tag{4}$$

This is equivalent to assuming that H_o is a minimum-phase function. This is not a limitation since H will be restored if, in a subsequent process, $|H_n|$ is multiplied by H_o (for instance through minimum-phase diffuse-field equalization of the headphones). Unlike monaural normalization, diffuse-

Meta Platforms, Inc.
Exhibit 1076
Page 010

field normalization does not affect the excess-phase of the HRTF and, as a consequence, cannot compensate time delay misalignments caused by possible differences in the placement of the two microphones in the ear canals during the HRTF measurements. This can be corrected by monaural normalization of the excess phase Φ with reference to the frontal direction.

Fig. 4a shows the diffuse-field HRTF for the same ear as in the previous figures and Fig. 4b shows the effect of the diffuse-field normalization on the measured HRTFs of Fig. 2. The overall effect is flattening the spectra, and the elimination of the ear canal resonance is clearly visible. This reduction of information to only the directionally dependent features can only be an advantage for subsequent data reduction or modeling of the HRTFs.

1.4 Transaural stereo vs conventional stereo

The theory developped by Cooper and Bauck [15] allows to design cost-effective cross-talk cancellers for faithful reproduction of dummy-head recordings or processed binaural signals on a conventional stereo loudspeaker setup, preserving the three-dimensional quality of these recordings. This theory is rewritten below to introduce a family of synthesis filters for conversion of mono signals to binaural and transaural formats, and between these two formats [3, 4].

In the loudspeaker listening situation, shown on Fig. 5a, the acoustic propagation from the two loudspeakers (signals Z_l and Z_r) to the two ears (signals Y_l and Y_r) can be characterized by an "acoustic transfer matrix" of 4 transfer functions:

$$\begin{bmatrix} Y_l \\ Y_r \end{bmatrix} = \begin{bmatrix} H_{ll} & H_{rl} \\ H_{lr} & H_{rr} \end{bmatrix} \cdot \begin{bmatrix} Z_l \\ Z_r \end{bmatrix} \tag{5}$$

Given a binaural signal Y, a pair of appropriate loudspeaker signals Z must be derived, such that the signal captured by the ears is identical to Y. The solution is obtained by passing Y through the inverse of the above propagation matrix, which defines the "lattice form" of the transaural cross-talk canceller [15], shown on Fig.5a:

$$\begin{bmatrix} Z_l \\ Z_r \end{bmatrix} = \frac{\begin{bmatrix} H_{rr} & -H_{rl} \\ -H_{lr} & H_{ll} \end{bmatrix}}{(H_{ll} \cdot H_{rr}) - (H_{lr} \cdot H_{rl})} \cdot \begin{bmatrix} Y_l \\ Y_r \end{bmatrix} \tag{6}$$

In theory, this cross-talk cancelling method does not require a symmetrical loudspeaker setup or an anechoic reproduction environment. It is only necessary that the four transfer functions be measured in the same situation as during the reproduction, hence the constraint on the listener's position. However, if the symmetry of the listener's head and of the reproduction system are assumed, the realization of the transaural converter can be significantly simplified, as described below.

If, additionally, the listener and loudspeakers are sufficiently far from walls and obstacles, so that no strong reflections reach the listener within the first 5 to 10 ms following the direct sounds from

Meta Platforms, Inc.
Exhibit 1076
Page 011

the loudspeakers, the effects of listening room reflections on the apparent directions of virtual sound sources are minimized [15]. The transaural converter does not have to cancel these reflections, and, assuming symmetry, it can be designed from two HRTFs measured in free field, corresponding to the direction of one of the two loudspeakers. Taking the left channel for reference, and using the notations $L_0 = H_{ll} = H_{rr}$ and $R_0 = H_{lr} = H_{rl}$, the acoustic transfer matrix of Eq. (5) can be written:

$$\begin{bmatrix} H_{ll} & H_{rl} \\ H_{lr} & H_{rr} \end{bmatrix} = \begin{bmatrix} L_0 & R_0 \\ R_0 & L_0 \end{bmatrix} = \frac{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}{\sqrt{2}} \cdot \begin{bmatrix} M_0 & 0 \\ 0 & S_0 \end{bmatrix} \cdot \frac{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}{\sqrt{2}} \tag{7}$$

where $M_0 = L_0 + R_0$ and $S_0 = L_0 - R_0$. In this expression, the acoustic transfer matrix is diagonalized through transformation by the "MS matrix" familiar to sound engineers, also known as "Blumlein Shuffler" [56]. This acoustic transfer matrix can be simulated using two shufflers and two digital filters, as shown on Fig. 5b(d), which yields a cost-efficient realization of a simple system for natural headphone reproduction of conventional stereo recordings, since such recordings are usually intended for reproduction on lousdpeakers.

Since the MS matrix (or shuffler) is its own inverse, the inverse of the acoustic transfer matrix can be realized by inverting the transfer functions M_0 and S_0, which yields the "shuffler structure" of the cross-talk canceller, as proposed by Cooper and Bauck, shown on Fig. 5b(b). Although M_0 and S_0 are not minimum-phase, it turns out that they are of "joint minimum phase", i.e. they have the same excess-phase response, and this excess-phase is approximately a pure delay [15]. This delay can be ignored, making the transaural cross-talk canceller of Fig. 5b(b) realizable. If the binaural signal is equalized to diffuse field, all that is necessary is to use diffuse-field normalized versions of the HRTFs L_0 and R_0 in the design of the cross-talk canceller.

Finally, the series association of the binaural synthesis filter of Fig. 5b(a) and the cross-talk canceller of Fig. 5b(b) can be simplified into the "transaural panpot" of Fig. 5b(c). Thus it can be seen that no more than two digital filters are necessary to realize any of the four binaural or transaural synthesis or conversion filters. Techniques for cost-efficient design of these filters, applicable in these four cases, will be examined in the next part of this paper, with illustrations in the binaural synthesis case.

2 DESIGN AND IMPLEMENTATION OF THE DIRECTIONAL FILTERS

It is widely admitted that a duration of about 5 ms is sufficient for the storage of a measured HRTF with no loss of data [9, 57]. At a sampling rate of 50 kHz, this represents a 250 tap FIR filter and requires, to convolve the left and right channels simultaneously in real time, a processing power of 25 million operations (multiply-accumulates) per second (approximately the capacity of the most recent commercially available programmable digital signal processors).

Impulse responses of 5-ms duration lead to a large amount of data, and to a heavy processing cost for the binaural synthesis. This justifies efforts for modeling the HRTFs, in order to extract more

Meta Platforms, Inc.
Exhibit 1076
Page 012

compact representations and realize more efficient implementations of the binaural synthesis filters, based on the hypothesis that not all of the information contained in a measured HRTF is necessary for faithful reproduction of localization. Three main approaches can be proposed for modeling HRTFs:

- *Physical modeling.* Analytical models of the HRTFs can be derived from simplified mathematical descriptions of head geometry. One such approach is the "spectral stereo" theory developped by Cooper and Bauck [58, 15], which assumes a spherical head shape and yields "universal" approximations of the HRTFs, devoid of listener-specific idiosyncrasies. However, since the mathematical model does not include the pinnae, these approximations are only valid up to about 4 or 5 kHz. A stereo microphone pickup based on a similar model has been marketted recently [59].

- *Perceptual modeling and data analysis.* Many studies have been carried-out in order to define universal HRTFs or perceptually relevant features of the HRTFs, from statistical analysis of measured data (averaging or multidimensional analysis) or experimental testing (see e.g. [8, 60-62]).

- *Signal modeling.* Even when the above approaches are used to reduce or modify the data representation, they only produce a new specification of the directional filters, from which one must derive a signal model in order to implement the binaural synthesis. Only the signal modeling problem is examined here, and no previous processing of the data is assumed other than the normalization described in part 1. Our goal is to reduce the computational cost to a minimum, comparing FIR and IIR designs and implementations of the digital filters. This comparison is then extended to time-varying implementations (to allow head-tracking or the simulation of moving sources), whith a discussion of methods for interpolating between filters.

2.1 FIR design and implementation

2.1.1 FIR filter design

In 1989, Persterer and Richter [18, 19] described a binaural mixing console capable of running 8 HRTF filter pairs simultaneously at a sample rate of 48 kHz, with "near-continuous" adjustment of the localization using rotary controls (azimuth and elevation), and an FIR filter length of 3.5 ms (168 samples). More recently, experiments with filter lengths as short as 1.5 ms have been reported [41, 3, 4, 57]. Sandvald and Hammershoi [57] describe an experiment indicating that this reduction is virtually unnoticeable in A/B comparisons and argue that the best window for designing the FIR filters is simply the rectangular one.

Other experiments have shown that minimum-phase approximation of the HRTFs does not influence localization judgements [61, 44], which confirms the validity of the approximation proposed in section 1.1. The proposed FIR design method is thus to the following:

1. conversion of the HRTF to {magnitude, excess phase} representation (section 1.3.1);
2. approximation of excess phase by a pure delay as described in section 1.1;
3. diffuse-field (or monaural) normalization of the magnitude frequency response (section 1.3.2);
4. minimum-phase reconstruction of the FIR filter response, using the inverse Hilbert transform and the inverse Fourier transform.

Meta Platforms, Inc.
Exhibit 1076
Page 013

Fig. 6a indicates that it seems indeed possible to reduce the filter length down to 1.5 ms or even 1 ms. Most of the reduction results from taking out the delay component, and some extra reduction results from the normalization, which substantially reduces the ear canal resonance (around 4 kHz). The effect of rectangular windowing to 1 ms is shown on Fig, 6b(a). Because the frequency response is displayed on a logarithmic frequency axis, the smearing effect of the time-domain rectangular window appears mostly in the lower frequency range. Flattening the frequency response by normalization can be seen to reduce the consequences of this smearing effect, allowing the use of a shorter time window.

This reduction of filter order of course applies to the transaural → binaural (loudspeaker-to-headphone) converter of Fig. 5b(d), but does not apply to the transaural synthesis filters of Fig. 5b(b) and 5b(c), because the recursive nature of the cross-talk cancelling process introduces resonances which typically take 5 to 10 ms to die out.

2.1.2 FIR implementation

A question that arises for implementing the FIR filter is the choice between time domain (direct form) convolution or frequency domain (e.g. overlap-save) convolution. The direct form convolution requires N operations (multiply-adds) per output sample for a N-tap filter. With the overlap-save algorithm, a block of N output samples is computed using two $2N$-point real FFTs and $2N$ complex multiplies [52], and typically introduces a processing delay of $4N$ samples. Assuming that the cost of a N-point real FFT is approximately $(3/2)\cdot N\cdot\log_2(N)$ [63], the cost of the frequency-domain convolution can be evaluated to $6\cdot\log_2(N)+10$ operations per sample. If the architecture of the signal processor used allows optimal coding of the frequency-domain convolution algorithm, it becomes more efficient than the direct-form implementation for N at least equal to 64 (in which case the processing delay is 256 samples, or 5.3 ms with a sampling rate of 48 kHz). Thus, in view of the above design considerations, the implementation of the FIR filter in the frequency domain does not seem necessary for free-field binaural synthesis.

If a minimum-phase design is used, a delay line must be inserted in cascade with the FIR filter. At a sampling rate of 50 kHz, rounding the delay length to the nearest sample leads to a worst-case error of 20 microseconds on the interaural delay. Assuming that the interaural delay is a linear function of the azimuth reaching about 0.7 ms at 90° (Fig. 3b), this error corresponds to about a 2.7° azimuth error in the horizontal plane, which is smaller than the localization blur of an average listener (about 3.6° for frontal directions according to [8]). If a more accurate control is desired or if a relatively low sample rate is used, a possibility is to insert an additional first-order all-pass filter to realize a delay length between 0 and 1 sample, with a reasonably good approximation for frequencies up to 1/4 of the sampling rate [64].

2.2 IIR design and implementation

2.2.1 IIR filter design

Experiments with IIR designs of HRTF filters where carried out in the early 1980's by Kendall et al. [65]. More recently, Mc Cabe and Furlong [66] described the realization of a "transaural pan-

Meta Platforms, Inc.
Exhibit 1076
Page 014

pot" using two IIR filters of order 20, in a combination different from that of Fig. 5b(c). To realize a listener-independent implementation, they derived the specification of the filters from the "spectral stereo" theory [58, 15]. Using the structure of Fig. 5b(c), Poncet and Jot experimented with two IIR filters of order 10. Although the HRTF data they used were 6th-octave spectra derived from published graphs in [8, 53, 15], informal tests showed good lateralization for all subjects, yet non satisfactory results for virtual sources in the back of the listener [3, 4].

More recently, we have been experimenting with IIR designs of listener-specific HRTF filters [2, 5, 6], typically of order 16 or 20 with the same number of poles and zeros. Convincing 360° localization in the horizontal plane can be maintained with a model order as low as 10 for some subjects. Lately, Sandvald and Hammershoi reported an experiment comparing FIR and IIR designs of HRTFs and indicating that, for a given processing cost, an IIR approximation was easier to detect than the FIR approximation [57].

Methods for approximating acoustic transfer functions with IIR filters have been studied in [64, 67-69]. J.O. Smith [64] gives an extensive review of possible approaches, according to different error criteria and specifications, and including a frequency domain error-weighting method. In the present application, a least-squares approximation method yielding a minimum-phase solution for a given magnitude spectrum is sufficient. Such a method is implemented in the Matlab mathematical software [70]: the 'yulewalk' algorithm described in [71] (also used in [57]).

The IIR design procedure currently used in our project follows the steps described in section 2.1.1 up to step 3. In step 4, instead of reconstructing a minimum-phase impulse response, we can directly use the 'yulewalk' algorithm to obtain the IIR filter coefficients, for a given order. For the same HRTF as in the previous section and an order of 16 (i.e. 32 coefficients), this yields the approximation shown on Fig. 6b(b). The result in the lower frequency range is clearly disappointing compared to the FIR design of Fig. 6b(a). However, the low-frequency fit can be improved by the following "error-weighting" method, adapted from [64]:

4. *"constant Q smoothing"*: smoothing of the magnitude frequency spectrum using a Hann window with its width proportional to frequency;
5. *"warping"*: resampling of the magnitude spectrum on a warped frequency scale, according to the bilinear transform $z \rightarrow (z + r)/(1 + r \cdot z)$, with the "warping parameter" r taken between -1 and 1;
6. approximation of the warped magnitude spectrum using the 'yulewalk' algorithm;
7. *"dewarping"*: bilinear transform $z \rightarrow (z - r)/(1 - r \cdot z)$ applied to the transfer function $B(z)/A(z)$ returned by the 'yulewalk' algorithm.

The "warping" step, for positive values or the parameter r, has the effect of oversampling the specified magnitude spectrum at low frequencies and undersampling it at high frequencies. The value of r can be chosen so that the warped frequency scale approximates the Bark scale, i.e. the modeling effort is evenly spread across the critical bands [64]. The fact that the high frequency spectrum is undersampled makes it necessary to smooth the spectrum first if warping is used, with a sufficiently wide smoothing window. The result obtained by warping with $r = 0.4$, preceded by 12th-octave (half-tone) smoothing, is shown on Fig. 6b(c). It can be seen that the high-frequency accuracy has been traded in for a better low-frequency accuracy, which is slightly better than with the FIR design, although the high frequency approximation is not as good.

Meta Platforms, Inc.
Exhibit 1076
Page 015

In [57], a method is proposed with the same purpose of improving the low-frequency fit for a given IIR model order (it consists in designing a higher-order model and *a posteriori* reducing the order by discarding selected pole-zero pairs in the high frequency range). It is possible that the warping method proposed in [64] yields better-controlled results since the smoothing and warping are introduced before the least-squares approximation, but this remains to be verified.

2.2.2 IIR implementation

The implementation of digital filters requires more care in the IIR case than in the FIR case because of the limited word length available to represent computed values and filter coefficients (see e.g. [72-75]). The main problems encountered with fixed-point arithmetic are: (a) inaccuracies due to coefficient quantization, (b) round-off or truncation noise, (c) saturation, (d) limit cycles. These problems are largely aleviated in floating-point architectures, except for the round-off or truncation noise (which remains because the mantissa is still truncated or rounded after a multiplication of two floating-point values). The solutions considered here are the cascade or parallel association of second-order sections and the lattice structure. Other possibly higher-quality solutions (e.g. the normalized ladder [76, 73]) are not considered here due to prohibitive computational cost.

For a IIR filter with N poles and N zeros, the implementation with second order sections in cascade or in parallel requires $2 \cdot N$ operations per ouput sample on a floating-point architecture. For the cascade biquad implementation in fixed-point arithmetic, scaling requirements impose an additional coefficient per second-order section, which leads to $(5/2) \cdot N$ operations per sample. For high-quality audio applications, the use of the "direct form I" on a 24-bit precision architecture is recommended [72, 74]. Round-off noise propagation and scaling problems can be minimized by the pole-zero pairing strategy used to define each section, and the subsequent ordering of the second-order sections to form the complete filter (see e.g. [74, 75]).

The lattice implementation of a N-pole, N-zero filter requires $4 \cdot N$ operations per sample on a programmable digital signal processor (using the 2-multiply form). However, a VLSI implementation can take advantage of the 1-multiply form of the lattice cell (1 multiply and 3 additions), which leads to a similar efficiency as second-order sections. The lattice structure is known to be less sensitive to round-off noise, coefficient quantization and gain scaling problems than the second order section. Furthermore, the filter stability check is simpler with the lattice than with the second-order section: all that is necessary is that its coefficients be smaller than 1 [76].

2.3 Time-variant implementation for moving sources and headtracking

With the recent advent of real-time virtual acoustic displays, the problem of realizing a binaural synthesis system capable of achieving smooth interpolations between the measured HRTF directions has received more attention than in early binaural mixing console designs for recording applications. The currently existing systems designed for virtual reality applications with headtracking use FIR designs of the binaural synthesis filters [27, 77, 30] (with implementation in the frequency domain in the case of [30]).

Meta Platforms, Inc.
Exhibit 1076
Page 016

2.3.1 Interpolation vs commutation

Two different processes can be distinguished in relation to the implementation of time varying digital filters for audio applications.

- *Interpolation* is the process of synthesizing an intermediate transfer function from a database of predefined filters. In our application, this database contains directional filter pairs derived from the HRTF measurements through the FIR or IIR design procedures described in the previous sections. An interpolation process is necessary to simulate a static localization of the source in a direction which was not sampled during the HRTF measurements.

- *Commutation* is the process of updating the filter coefficients while the filter is running. The signal processor should realize this update "instantly" between the computations of two successive ouput samples (typically by flipping a dual-port memory). Updating the coefficients of a digital filter can produce noticeable "clicks". This problem is more severe with IIR filters because updating the coefficients produces a mismatch between the internal variables of the filter's recursive part and the new coefficient set, which has the same effect as an impulsive input signal: a transient response at the output. The role of the commutation process in our application is to allow smooth transitions from a source position to another without noticeable artifacts. Although this is often realized by synthesizing intermediate coefficient sets, we will not use the name "interpolation" here for this technique, because the intermediate filters do not necessarily need to be valid as *static* directional filters.

The interpolation process does not need to be performed in real time by the signal processor if the resolution of the HRTF database is finer than the minimum audible difference. However, if smooth transitions are desired, the necessary commutation process must run in real time. According to results reported in [8] on the localization blur of an average listener, a conservative database is obtained with 5° resolution in the horizontal plane and 10° resolution in the median plane. This assumption leads to a database of the kind recently collected by Gardner and Martin [48] (with 710 directions spanning elevations from -40° to +90°). Assuming a conservative 75-tap minimum-phase FIR design and a 24-bit coefficient wordlength, the resulting database occupies 320 kbytes of memory if both ears are stored. If only one ear is stored (assuming symmetry) and an order 16 IIR design is used, this memory requirement is reduced to 80 kbytes, but still represents a large amount of fast on-chip memory. As a consequence, it is desirable to implement an efficient real-time interpolation procedure, in order to reduce the size of the directional filter database.

In a typical "synchronus" operation mode, a new target direction is received at regular time intervals and the update of the digital filter coefficients can be implemented by the following cyclic procedure:
1. read a new pair of filters corresponding to the new target direction (or compute this new pair of filters by an interpolation method) from the on-chip HRTF database;
2. initialize the commutation process: compute the elementary increments for all the coefficients used to control the commutation;
3. increment and update the filter coefficient sets at regular time intervals until the next target direction is received, and then return to step 1.

Meta Platforms, Inc.
Exhibit 1076
Page 017

In this typical procedure, two periodic cycles appear on two different time scales, with one (the commutation cycle) included in the other (the interpolation cycle). The two corresponding periods will be called below the *interpolation period* and the *commutation period.* The interpolation period is dictated by perceptual and hardware constraints, and can typically be of 10 to 40 ms [27]. The commutation period is smaller than the interpolation period, but it is desirable to make it as long as possible to reduce the computational cost, while maintaining transitions free of artifacts.

2.3.2 "Ideal" interpolation

Since the interpolation does not need to be performed in real time if on-chip memory is sufficient, it is worth considering an "ideal" directional interpolation scheme which could be used off-line to construct a fine-resolution HRTF database from coarse-resolution HRTF measurements. Interpolation can be viewed as a filter design problem for which the representation adopted in section 1.3.3 can be followed. According to this representation, the magnitude spectrum and the excess-phase spectrum (i.e. the delay) of the intermediate filter can be modeled independently. In view of Fig. 3b, a piecewise linear interpolation of the delay seems sufficient. As for the magnitude spectrum, a natural approach is to perform a linear interpolation on a dB scale, since this is our preferred representation for predicting the perceptual accuracy of a filter design method in sections 2.1 and 2.2 (this representation, however, is based on models of monaural hearing and may require confirmation in the case of binaural hearing). Interpolating on the magnitude at each frequency is actually not the ideal way of interpolating between two spectra, since it eliminates transformations along the dimension of frequency (implying that frequency shifts of sharp peaks or notches may not be well interpolated).

Piecewise linear interpolation is not the most general interpolation method either. Theoretically, due to the periodic nature of the directional representation, functions of the azimuth (like the interaural delay shown on Fig. 3b) should be interpolated by Fourier series decomposition. This method is used in [3] to interpolate the interaural level difference from sparse measured data taken from [8] (with 30° steps), yielding a global analytical expression of the interaural gain as a function of the azimuth: the gain values are not read and interpolated from a stored table of gain values, but this table is replaced by the list of coefficients of the Fourier series, which has the same size unless the variations of the gain with azimuth are intentionally "smoothed". The extension of this method to include both azimuth and elevation is given by spherical harmonic decomposition, and this interpolation method can be applied to the delay as well as to the magnitude spectrum. This is actually a classical way of representing the directivity of a source or a transducer [78], applied here to the ear. Piecewise linear approximation differs from these methods by the fact that it is a *local* interpolation method, which makes it well suited to real-time implementation.

2.3.3 Time-variant FIR implementation and minimum-phase reconstruction

A straightforward approach to interpolating digital filters consists in computing an intermediate set of coefficients by linear interpolation between the sets of coefficients of the nearest filters. In the case of directional synthesis, the nearest filters correspond to the nearest directions and a 2-dimensional interpolation process is needed (azimuth and elevation). In a FIR implementation, the interpolation is then performed separately for each ear by linear combination of the impulse responses coresponding to the three nearest directions (or to the four nearest directions, as chosen in [27]).

Meta Platforms, Inc.
Exhibit 1076
Page 018

This method is equivalent to linearly combining the ouput signals of the nearest directional filters in the database, which simulates a situation where the listener is surrounded by loudspeakers (placed at all the positions where the HRTF measurements were taken), and a conventional amplitude "panpot" is used to control the perceived direction of the sound event. This method is often used in multichannel loudspeaker reproduction systems, e.g. for computer music [79].

According to the previous section, an alternative "filter design" approach consists in associating a variable delay line and a minimum phase FIR filter in cascade. The FIR filter should be designed, for every new target direction, by minimum-phase reconstruction from a magnitude spectrum obtained by linear combination of the log-magnitude spectra of the three closest HRTFs. This reconstruction looks costly since it involves an inverse Hilbert transform followed by an inverse Fourier transform. However, an interesting simplification arises from the linearity of the Hilbert transform: since the minimum phase is the inverse Hilbert transform of the log-magnitude, it is equal to the same linear combination applied to the minimum phases of the closest filters. Thus, in a frequency-domain FIR implementation, the "ideal" interpolation scheme of section 2.3.2 can be implemented, for a reasonable cost, by storing in the HRTF database *the complex logarithm of each minimum-phase frequency response* (i.e. its log-magnitude and its phase).

For a time-domain implementation of the FIR filter, a more practical method consists in interpolating by linear combination on the coefficients of the minimum-phase impulse responses. As shown on Fig. 7a(b), this yields surprisingly good interpolated designs. In contrast, on Fig. 7a(a), the same interpolation method applied to mixed-phase impulse responses yields unsatisfying interpolations of the spectra, due to "comb-filtering" effects caused by mixing responses with different delays. As reported in [77], these effects can be noticed by the listener when a source emitting a wide-band stationnary signal is simulated on headphones, and either the source or the listener (with a headtracker) is moving. The smooth linear interpolation obtained from minimum-phase impulse responses on Fig. 7a(b) can again be traced back to the linearity of the Hilbert transform. According to the previous paragraph, the only difference between the two methods (applied to minimum-phase responses) is that the "ideal" method interpolates on the complex logarithm of the frequency response, while the practical time-domain method is equivalent to interpolating on the frequency response itself. Thus, if the original filters have "close" frequency responses, it is not surprising that the two interpolation methods yield similar results.

Commutation and cost evaluation. It is natural, in the time-domain FIR implementation, to use linear combination of the coefficients both for interpolation and commutation. Typically, the commutation period can be taken equal to the FIR length N, which leads to a commutation cost equal to 1 operation per sample. Initializing the commutation (computing the increments) requires $2 \cdot N$ operations, and a 3-point interpolation requires $2 \cdot N$ operations also (the cost of searching the 3 nearest directions is not included in this evaluation). Thus, assuming that the 4 initial commutation periods in each interpolation period are used only for computing the interpolation and initialization, while the following commutation periods are actually used for updating filter coefficients, the cost of making the FIR filter time-variant is evaluated to only one extra operation per output sample.

Meta Platforms, Inc.
Exhibit 1076
Page 019

2.3.4 Time-variant IIR implementation

The technique of linearly interpolating the coefficient sets of the nearest filters in the database can also be considered for implementing time-varying IIR filters. This technique yields different results depending on the filter structure used. With second-order sections or lattice filters, the intermediate filters obtained by linear interpolation from two stable filter coefficient sets are guaranteed to be stable because the stability domain, for both structures, is a convex subset of all possible coefficient sets [75].

A difficulty arising with the decomposition in second-order sections is that the sections can be interchanged with no modification of the global transfer function. Since the global interpolation process can be viewed as independent interpolation processes applied to the different second-order sections, the pairing of poles and zeros and the ordering of the sections in the two original filters has a strong influence on the resulting interpolated filters. Thus, a pairing and ordering algorithm must be applied globally to the whole database of filters, in order to optimize the regularity of the interpolation from a filter to another (this problem is similar to the "formant labeling" problem encountered in speech synthesis). The algorithm we have designed, described in [80], is based on pairing the sections according to the frequencies of poles and zeros. Since the IIR filter design procedure of section 2.2.1 does not return only complex poles and zeros, some filters are "completed" with an additional complex pair of neutral (low magnitude) poles or zeros, while real roots are assigned to additional second-order sections. The interpolation plot of Fig.7b(a) was obtained using this pairing method and linear interpolation on the coefficients of a cascade of 10 second-order sections. The coefficients were obtained by order 16 IIR approximations of the 7 HRTFs measured at azimuths 0° to 90° in the horizontal plane, using the design method described in section 2.2.1. The pairing algorithm added two second-order sections for real poles and zeros.

It may be argued that the perceptual validity of the linear interpolation could possibly be improved by using different parameters to control the interpolation. Such control parameters should be simple combinations of the coefficients of the chosen filter structure, to limit the increase in computational cost involved by converting these control parameters into filter coefficients. Four possible choices of control parameters have been more particularly considered in this study:
- the direct form coefficients of the second-order sections in cascade
- the magnitudes and log-frequencies of the poles and zeros
- the reflection coefficients (coefficients of the lattice filter), noted k_i ($1 \leq i \leq N$)
- the log area ratios [81, 82, 76]: $\log[(1-k_i)/(1+k_i)]$

The first two control parameters, which can be used for the cascade association of second-order sections, are closely related and yield very similar interpolation plots in practice. The reflection coefficients and the log area ratios, which are adequate for the lattice structure, also yield similar interpolation plots, with a much more regular behaviour than with the second-order section coefficients, as illustrated in Fig. 7b(b). The advantages of the log area ratios for interpolating between spectra are well-known in speech processing, where time-varying IIR filters, often implemented in lattice form, are used to reproduce the speech formants. The denomination "log area ratio" comes from an analogy between the vocal tract, modelled as a tube with N sections, and the all-pole lattice structure, where the ratios $(1-k_i)/(1+k_i)$ are analogous to the ratios between adjacent sections of the tube [76]. The log area ratios were initially proposed as a transformation yelding

Meta Platforms, Inc.
Exhibit 1076
Page 020

improved quantization properties for coding the reflection coefficients k_i, since they can be related to a spectral sensitivity measure derived from the log-magnitude spectrum [82]. Their application to the interpolation problem thus yields perceptually uniform spectral transformations, as illustrated by Fig. 7b(b).

Unfortunately, there is no simple relation from the log area ratios to the coefficients of the second-order sections: the reflection coefficients are related to the coefficients of the global direct form filter by the Levinson algorithm [76], but a prohibitive amount of computation remains necessary for computing the poles and zeros from these direct form coefficients. Other control parameters applicable to second order sections are provided by the formulas used in the design of parametric equalizers [83-85]: e.g. bandwidth, cutoff frequencies, etc... However, there is no general conversion method to compute such parameters from any filter coefficient set.

Commutation:

Although the direct form coefficients of the second-order sections are not adequate for controlling the interpolation process, they may be adequate for controlling the commutation process. The commutation of digital filters is a problem generally encountered in the design of parametric equalizers for studio applications (e.g. digital mixing consoles), about which little has been written in the litterature. In general, the stability of the time-variant filter is not guaranteed even if the intermediate coefficient sets all yield stable filters, but it can be shown that this problem is not encountered with the second-order section [86, 75].

Mourjopoulos et al. [85] examined the commutation problem with typical equalizers and typical audio signals, taking masking effects into account for evaluating the audibility of the commutation artifacts ("clicks"). From their study, it appears that a commutation period of 5 to 10 ms is adequate for typical applications. However, this figure turned out not to be applicable to our problem, due to the large order of the filters and because the coefficients are given by individual modeling of each filter instead of a direct relation to the controlled parameter (the direction). This makes our problem more similar to speech analysis / synthesis applications, where the preferred synthesis algorithm is typically the lattice structure with the commutation smoothed on a sample-by-sample basis (i.e. with the commutation period equal to the sampling period).

An alternate method, proposed in [87], consists in updating alternately the coefficients of two filters which run simulateously. At every interpolation period, one of the two filters receives the new target coefficient set and the input signal is switched to this filter, while the other filter is left to decay with the previous coefficient set. This method works if the interpolation period is longer than the response of the filters, which is the case in our application. The method used in our time-varying IIR implementation is a variation of this technique. Although it requires running two filters simultaneously, it compares favorably with a lattice implementation since is involves no additional commutation cost.

Meta Platforms, Inc.
Exhibit 1076
Page 021

2.4 Conclusion: FIR vs IIR implementation

There are great benefits in exploiting the phase properties of HRTFs for designing binaural or transaural synthesis filters: since transfer functions can be manipulated under the form of minimum-phase filters cascaded with pure delays, normalization and modeling procedures are simplified, and efficient implementations are obtained. This applies to the four synthesis or conversion applications illustrated on Fig. 5b, where we always end up designing two minimum-phase filters.

The main advantage of IIR filter design is that it offers a general solution allowing a controlled fit at low frequencies even with low filter orders, because reducing the order does not mean a restriction on the impulse response length. As seen in section 2.1.1, FIR filters may be equally efficient for binaural synthesis, especially if the high-frequency content is important. This is particularly true if diffuse-field or monaural normalization is used, because the compensation of the ear canal resonance yields a shorter impulse response, making 1-ms FIR lengths possible. However, this is not the case for transaural synthesis, because the cross-talk cancellation process leads to longer impulse responses.

Time-varying implementations. The benefits of manipulating minimum-phase filters become even clearer in the design of a binaural processor capable of simulating dynamic mouvements of the source or the listener, as seen in section 2.3.3. The FIR implementation becomes particularly attractive in this context, because it can be made time-variant at a negligible cost, using the straightforward method of interpolating on the coefficients of the impulse response. As illustrated by Fig. 7a(b), this yields a relevant method for interpolating between directions, allowing to minimize the on-chip memory storage required for the HRTF database. In contrast, the cost of a high-quality time-varying IIR implementation is twice that of a static IIR implementation with second-order sections. Thus, the IIR approach is competitive only if the IIR model yields better approximations than a FIR model with an order 4 times as high. This is likely for a transaural panpot, but unlikely for binaural synthesis, in view of the results presented here.

Processing cost evaluation. Following from these evaluations, which call for further confirmation by psycho-experimental validation, the cost of a binaural or transaural synthesis or conversion filter for high-quality audio applications should not excede 100 to 150 operations per output sample (for processing the two channels at a 50 kHz sample rate). This is enough for realizing a static or time-variant FIR binaural synthesis filter, or a time-variant IIR binaural or transaural filter. In a static IIR implementation, this cost can be divided by 2, or the accuracy of the IIR model can be improved. This implies that a digital signal processor capable of 500 operations per output sample at a sample rate of 50 kHz can process 3 or 4 sources simultaneously. The implementation of the delay lines is not included in this evaluation, since most DSPs can perform memory moves and address calculations in parallel with arithmetic operations. Alternately, to simulate a more natural listening experience, one DSP can be used to synthesize a direct sound plus the first 2 or 3 room reflections that follow. However, we can substantially reduce the complexity by taking advantage of the fact that each reflection is not perceived as an isolated event. This will allow us to replace the 2 or 3 reflections by a full synthetic room effect, still without exceeding the capacity of one processor.

Meta Platforms, Inc.
Exhibit 1076
Page 022

3 REAL-TIME BINAURAL SIMULATION OF ROOM REVERBERATION

3.1 Convolution vs artificial reverberation

As mentionned in the introduction of this paper, the most straightforward method of simulating room reverberation consists in convolving the source signal with a binaural impulse response measured in an enclosure or computed from a geometrical and physical description of the room boundaries. Since an impulse response measured in a large room can typically be 2 or 3 seconds long, the time-domain convolution may require more than a hundred thousand operations per output sample at a 48 kHz sample rate. For the same task, the frequency-domain convolution only requires about 100 operations per sample (according to the evaluation in section 2.1.2), but will introduce a processing delay of about 10 seconds. This solution is adequate as a fast off-line computation method for auralization in room acoustics, but not for real-time studio applications or virtual acoustic displays. Possible solutions for real-time artificial reverberation with no processing delay include the following three approaches:

- *Hybrid time-domain and frequency-domain convolution.* By combining a time-domain FIR filter for the beginning of the impulse response and frequency-domain convolution for the later response (itself divided in sections of exponentially increasing length), Gardner recently designed a zero-delay convolution algorithm which requires approximately $35 \cdot \log_2(N)$-150 operations per sample for a N-point response [63]. If N = 128k points, this leads to about 450 operations, which is within the reach of recent DSPs at a 48 kHz sampling rate (provided that efficient coding of the algorithm on the chip is possible).

- *Artificial reverberation using feedback delay networks.* Following the pioneering work of Schroeder in the early 1960's, the conventional approach to artificial reverberation has been to combine a FIR filter to simulate early reflections and a recursive delay network to simulate the later diffuse reverberation [88-94, 3]. This approach is widely used in the design of studio reverberators and assisted reverberation systems for concert halls [95, 96]. A difficulty with this method is to ensure the naturalness of the artificial reverberation, and accurate control of the reverberation time as a function of frequency. This difficulty can be overcome, however, and the feedback delay network can be made to accurately imitate the reverberation of a medium or large room [3, 97].

- *Hybrid approach combining multirate filter banks with parametric modeling.* In this approach [98], the room response is modelled in each octave band by a FIR filter in cascade with a recursive comb filter. The efficiency of the FIR section is improved by subsampling each subband signal. The method allows accurate control of the reverberation time in octave bands and accurate match to the beginning of a measured room response, but does not provide direct control of the time structure of early reflections. Little is known, at the time of this writing, on the processing cost of the algorithm and on the naturalness of the synthetic reverberation obtained by this recent method.

An important concern for many applications is the possibility of efficiently controlling the artificial room effect in real time, which implies a description on a higher level than a mere storage of the sampled impulse response. Both physical and perceptual studies of room acoustics indicate that the control formalism should give direct access to the distribution of early reflections in time, direction and amplitude (or spectrum). In contrast, the later reverberation lends itself to a statistical descrip-

Meta Platforms, Inc.
Exhibit 1076
Page 023

tion, i.e. an exponentially decaying gaussian process which can be characterized by two functions of frequency: its spectrum and its reverberation time. A recursive algorithm allows more efficient control of such an exponentially decaying process than a feedforward (convolution) algorithm. This statistical model of the late reverberation, however, supposes a large modal overlap in the frequency domain, which is realized above the "Schroeder frequency" of the room. This limit frequency is about 100 Hz in a typical room (and lower in a concert hall), but will be higher in a small relatively reverberant room [11, 3].

3.2 Stereo reverberator

Artificial reverberation by delay networks can bring the following advantages:
- direct real-time control of both the early reflection distribution and the late reverberation,
- long reverberation times (up to infinite) do not imply an increase in processing cost,
- a multichannel decorrelated output can be obtained for no increase in processing cost.

This is illustrated on Fig. 8, which shows a structure for realizing a stereo reverberator for conventional studio applications [3]. The reverberator is made of 2 modules: an early reflection generator and a reverberant filter. In a typical mixing configuration, the reverberator is a peripheral equipment whose input signal is gently low-pass filtered, and the direct path is realized by the mixing console, which provides a "panpot" to localize the sound event in the stereo image.

The early reflection module is a FIR filter providing P delayed and scaled copies of the input signal. These are shared alternately between the left and right output channels and constitute "first-order reflections". These delayed signals are also sent separately to the P input channels of the reverberant filter, which then synthesizes reflections of order 2, 3, etc... through successive trips around a feedback loop made of an energy preserving (i.e. unitary) matrix A and a bank of delay lines τ_i ($1 \leq i \leq P$) of different lengths. This structure is essentially derived from earlier work by Stautner and Puckette [91], with the following additional refinements [3, 94]:

- *Reverberation time control.* The reverberation time is controlled by associating an absorbent filter $\alpha_i(f)$ to each one of the delay lines τ_i. The frequency response of each absorbant filter is uniquely specified to ensure exact control of the reverberation time as a function of frequency $Tr(f)$: its absorption, expressed in decibels as a function of frequency, is proportional to the associated delay length and inversely proportional to the reverberation time.

$$20 \cdot \log|\alpha_i(f)| = -60 \cdot \tau_i / Tr(f) \tag{8}$$

This has the effect of ensuring *locally uniform damping* of the normal modes of the reverberator: all modes in the vicinity of a given frequency have the same decay time, equal to the reverberation time specified at that frequency. This method is valid irrespective of the unitary feedback matrix A used, and eliminates the possibility of "ringing modes" in the late reverberation decay, which is a crucial criterion for the naturalness of the synthetic reverberation (see [3] or [94] for a more general theoretical study).

- *Frequency and time density.* The second condition for synthesizing a natural-sounding room effect is to ensure a sufficient number of normal modes per Hz and a sufficient number of "reflec-

Meta Platforms, Inc.
Exhibit 1076
Page 024

tions" per second. These constraints can be seen as necessary for satisfying the statistical model of the reverberation, i.e. approaching a gaussian amplitude distribution in both the frequency domain and the time domain. The modal density is equal to the total length of the delay lines τ_i, which should be at least 1/4 of the reverberation time to ensure sufficient modal overlap in the frequency domain. The time density then depends on the number P of delay channels and the "density" of the feedback matrix $\boldsymbol{A}$. For a given P, the maximum time density is obtained with feedback matrices having a "unit crest factor" (meaning that all coefficients have the same absolute value), because this ensures maximum increase of the reflection density through each trip of the signals around the feedback loop. Cost-effective unitary matrices with real coefficients and unit (or low) crest factor can be derived e.g. from the class of Householder matrices, requiring only $2 \cdot P$ operations or $P \cdot \log_2(P)$ operations (instead of P^2 operations) to compute the mixing [3].

- *Early reflections*. The early reflections are controllable separately in time and amplitude and the early / late reverberation energy ratio can be controled using the coefficients b_i and c_i. The durations of the delay lines τ_i can be chosen so that "first-order reflections" precede all reflections of a larger order. Finally, the settings of the early reflection parameters do not affect the tone color of the synthetic late reverberation. This is another important naturalness criterion, which is satisfied here because the delayed inputs to the reverberant filter are transmitted on independent channels (contrary to some reverberator designs where they are mixed in order to increase the time density of the late reverberation, thereby causing comb filter colorations). Several early reflection modules can be connected in this manner to one reverberant filter in order to simulate a situation with several sources at different positions in the same room.

Assuming that the absorbent filters and the low-pass filter are second-order IIR filters (requiring 5 operations each), the total number of operations per output sample for the reverberator of Fig. 8 is $P \cdot (\log_2(P)+8)+7$. This implementation allows reverberation time control in three separate frequency bands with adjustable cross-over frequencies. A 12-channel version of this algorithm yields sufficient time and frequency densities for high-quality artificial reverberation, with a total of approximately 150 operations per output sample.

In practice, data and memory manipulations can increase this cost by 50% to 100%, depending on the architecture of the programmable processor used (particularly the number of internal accumulators and registers). As an illustration, a carefully handcoded version of this algorithm with first-order absorbent filters, theoretically requiring about 110 operations, occupied a full Motorola DSP56000 with a 20.5 MHz clock, at a 48 kHz sample rate (i.e. a capacity of about 200 operations per output sample) [3]. Since some existing DSPs are more than twice as fast, this basic algorithm can be substantially extended in order to ensure binaural compatibility and more accurate synthesis of room reverberation, still without exceeding the capacity of one processor.

3.3 Binaural room simulator

In this section, we describe extensions of the above reverberator with the purpose of synthesizing a two-channel recording made in a virtual reverberant environment, using a conventional stereo microphone pickup or a dummy (or live) head. These extensions will be presented in two steps: the first step provides compatibility with binaural or stereo recordings for virtually no increase in complexity; the second step yields a more refined model allowing to reproduce spectral effects which

Meta Platforms, Inc.
Exhibit 1076
Page 025

depend on the directions of incidence of the early reflections, the source directivity and the absorption properties of the room boundaries.

Recent studies (e.g. [43]) suggest that some modifications of the temporal and spatial distribution of early reflections are not perceptible. Following this hypothesis, a possible approach towards simplifying the description of the early room effect would be to investigate the perceptual effects of elementary modifications in this description. The goal could be to evaluate the minimum audible difference for each of the parameters which characterize the distribution of early reflections (number of reflections and date, energy, direction of each reflection). We explore here an alternative approach, where the parameters of the distribution remain unchanged, but we attempt to simplify the model for reproducing each reflection, without affecting the overall perceptual effect. This approach has direct consequences in terms of processing cost reduction.

3.3.1 Binaural reproduction of early reflections with a "stereophonic" model

A straightforward extension for transforming the stereo configuration of Fig. 8 into a binaural spatial processor consists in replacing the stereo panpot by a binaural panpot (requiring an extra 100 to 150 operations per sample according to the evaluation in section 2.4). From the discussion in section 1.2, we know that this modification will be valid if the binaural panpot is diffuse-field normalized, in order to ensure timbre compatibility with the synthetic reverberation. For no additional cost, we can control the early reflections as left-right pairs in order to define $P/2$ "stereo reflections", each with a date, an amplitude, and a direction (the direction being controlled by the delay and amplitude difference between the two channels). Additionally, a 2-by-2 matrix can be added at the output of the reverberant filter to control the correlation between the two channels (i.e., over headphones, the interaural cross-correlation coefficient, or IACC, defined e.g. in [8]). A derivation of this binaural reverberation model and an adjustement of its parameters are detailed below [3].

The binaural information in a pair of HRTFs can be separated as follows (see e.g. [8]):
- interaural differences of delay and amplitude,
- spectral cues due to the diffraction of the incoming sound wave by the head and the pinnae.

This can be understood as a hierarchization of binaural cues. A complete and accurate reproduction of these cues is necessary in free field to convey accurate localization. However, the current knowledge of spatial hearing (notably the "precedence effect") suggests that when the direct sound from the source is followed by a group of reflections, reducing the accuracy in the directional reproduction of each early reflection could remain unnoticed, as long as the direction of the direct sound is faithfully rendered. This reduction could be applied on the order of the FIR or IIR filters used to model the HRTF pair for each reflection. If the order of these filters is reduced to zero, we are left only with frequency-independent interaural delay and amplitude differences.

This model can be called "stereophonic" because it is equivalent to a conventional stereo recording realized with a non-coincident pair of microphones (assuming frequency-independent directivity). In this stereophonic model, we can adjust the spacing and directivity of the microphones in order to reproduce the prevalent binaural cues:

Meta Platforms, Inc.
Exhibit 1076
Page 026

- *Interaural delay difference.* For frequencies above about 1 kHz, existing measurements of the interaural delay show approximately a linear function of the azimuth increasing from 0 ms at 0° azimuth to about 0,7 ms at 90° azimuth [8, 15]. This is approximately verified when the interaural delay is derived from the excess phase of the HRTFs, as described in section 1.1 (Fig. 3b).

- *Monaural gains.* To evaluate the monaural gains, denoted g_l and g_r in the following, we are led to adopt a frequency-domain weighting window, in order to define a "mean energy" of each HRTF. The interaural gain calculated in this way proves to be hardly modified whether we select a frequency band $[f_1, f_2]$ of [1, 8 khz], [1, 5 khz], or [2, 8 khz]. As a consequence, in accordance with section 1.1, the chosen limits are f_1 = 1 khz and f_2 = 5 khz. The monaural gains for the direction i are given by the following expressions, where $(H_{l,i}, H_{r,i})$ denotes the HRTF pair and Av denotes averaging over frequencies:

$$g_{l,i} = \sqrt{Av_{f_1}^{f_2}|H_{l,i}(f)|^2} \qquad g_{r,i} = \sqrt{Av_{f_1}^{f_2}|H_{r,i}(f)|^2} \tag{9}$$

3.3.2 Separating spatial and temporal aspects: an average directional filter

Implementing the directional filters with the above "stereophonic" model yields a considerable processing cost reduction, compared to the full binaural synthesis of each early reflection using the same kind of directional filter as for the direct sound. However, although this model essentially preserves the lateralization cues in each reflection, it ignores the spectral variations introduced by the head and the pinnae, as well as the frequency-dependent effects of wall reflections and source directivity. We now present an approximate model for reintroducing this missing spectral information [5, 54].

The contribution of the early reflections to the perceived spectrum depends on:
a) the temporal distribution of the reflections and their respective amplitudes e_i,
b) the "history" of each reflection: the spectrum of each reflection, within the geometrical acoutics theory, can be viewed as the result of a cascade of elementary filters associated to the radiation of the sound source in the direction of emission of the sound ray, to the absorption by the walls and the air and to the HRTFs of the listener for the direction of incidence of the reflection (e.g. [35]).

To clarify the derivation below, the directivity of the source and the absorption coefficients of the walls will first be considered independent of frequency, i.e. they will only result in a pure attenuation for each reflection. Under these assumptions, the information from (a) is exactly that which would be recovered by an omnidirectional microphone, while the information from (b) is strictly directional and corresponds to the difference between the information recovered by the listener's ears and that recovered with an omnidirectional microphone. As illustrated in Fig. 9a, we can further decompose this difference by first considering the information recovered with a pair of conventional microphones (according to the above stereophonic model) and then considering the additional spectral information due to the head and the pinnae.

Given the temporal integration properties of the ear, we may assume that the perception of the different spectra of the individual reflections is, in part, described by a common spectrum for the

Meta Platforms, Inc.
Exhibit 1076
Page 027

group of reflections, irrespective of their temporal distribution. This suggests that the binaural reproduction of a dense group of reflections can be simplified without affecting the listener's perception, by using a common filter for all reflections.

If all the directional filters (L_i, R_i) which appear at the final (binaural) processing level of Fig. 9a had the same transfer function, it would be equivalent to implement the process as shown in Fig. 9b, with a common pair of filters (L, R). To ensure that the grouping of the directional filters remains inaudible in a more general case, it is proposed that, for each ear, the total energy conveyed by the reflections should remain unchanged for all frequencies. For N reflections with amplitudes e_i and directions corresponding respectively to the HRTF pairs ($H_{l,i}$, $H_{r,i}$), the total energy arriving at each ear may be written:

$$E_l = \sum_{i=1}^{N} [(e_i)^2 \cdot |H_{l,i}(f)|^2] \qquad E_r = \sum_{i=1}^{N} [(e_i)^2 \cdot |H_{r,i}(f)|^2] \tag{10}$$

In the realization of Fig. 9b, the transfer functions of the two filters L and R must be normalized in order to take into acount the total energy introduced by the gains e_i and $g_{l,i}$. The required transfer functions for the pair of filters (L, R) on Fig. 9b are thus given by:

$$|L(f)|^2 = \frac{\sum_{i=1}^{N} [(e_i)^2 \cdot |H_{l,i}(f)|^2]}{\sum_{i=1}^{N} [(e_i)^2 \cdot (g_{l,i})^2]} \qquad |R(f)|^2 = \frac{\sum_{i=1}^{N} [(e_i)^2 \cdot |H_{r,i}(f)|^2]}{\sum_{i=1}^{N} [(e_i)^2 \cdot (g_{r,i})^2]} \tag{11}$$

Considering the calculation of the monaural gains $g_{l,i}$ and $g_{r,i}$ given by Eq. (9), this amounts to dividing each total energy spectrum, $E_l(f)$ and $E_r(f)$, by its own average energy calculated in the frequency band $[f_1, f_2]$. Note that it is the same operation that yields the filter pairs (L_i, R_i) of Fig. 9a from the HRTF pairs ($H_{l,i}$, $H_{r,i}$).

Finally, we can reintroduce, in this pair of "average filters", the frequency dependence of the acoustical characteristics of the sound source and the walls. In Fig. 9a, this implies replacing each of the gains e_i by a filter equivalent to the cascade of elementary acoustic filters undergone by the reflection prior to its arrival at the listener. This can also be realized by transferring the spectral variations in this filter into the filters L_i and R_i and making $(e_i)^2$ equal to the average energy of the reflection in the frequency band $[f_1, f_2]$, as measured with an omnidirectional microphone. The grouping of the individual filters (L_i, R_i) into the global filters (L, R) of Fig. 9b can then be realized just as described above. The frequency responses of the filters (L, R) finally incorporate the following spectral effects, averaged for each ear over all reflections:
- the spectral dependance of the source directivity,
- the variation of the absorption caracteristics of the boundaries and the air with frequency,
- the spectral effects related to the directional distribution of the reflections.

Meta Platforms, Inc.
Exhibit 1076
Page 028

It should be noted that this method does not imply any simplification in the geometrical modeling (by a source image or ray-tracing algorithm) of the sound propagation in the room. The proposed approximation is only introduced at the final binaural processing (or "auralization") stage.

3.3.3 Application to later diffuse reflections

Let us now consider a section of a room impulse response where the reflections create a diffuse field: at any time, a large number of reflections are received by the listener from directions equi-distributed in space and with the same spectrum $e(f)$. It is then unnecessary to synthesize the sound field by controlling the direction of each reflection. Rather, the signal captured by an omnidirectional microphone can be described as a gaussian noise characterized by the spectrum $e(f)$, which depends on the source radiation and wall absorption properties. The signals at the two ears can be described as two partially correlated gaussian noises, and this correlation is measured by the IACC coefficient. The spectra of these two noises are given by the product of $e(f)$ with the left and right diffuse-field HRTFs, and the latter are equal to 1 if the HRTF database is diffuse-field normalized. The diffuse-field normalized binaural signal model for diffuse reflections is thus simply characterized by the spectrum $e(f)$ of the diffuse sound field, as measured by an omnidirectional microphone, and the interaural cross-correlation coefficient IACC.

At the opposite extremity of the acoustic channel, a similar approach can be followed for isolating the contribution of the source in the incoming spectrum $e(f)$, if we now assume a diffuse field in the whole room. In this case, the sound field is fed uniformly by all directions of emission from the source. By the reciprocity principle, it follows that the diffuse-field radiation transfer function of the source (i.e the power average of the spectra emitted by the source in all directions) adequately describes the contribution of the source to the sound field perceived by the listener. This average spectrum can thus describe, more particularly, the contribution of the source to the late reverberation process in a room [10]. Actually, the late reverberation process in a room is completely described, when an ideally diffuse sound field is reached, by an initial spectrum equal to the product of the diffuse field transfer functions of the source and the receiver, and by the reverberation time as a function of frequency.

3.3.4 Structure of the binaural room simulator

Finally, starting from the stereo configuration of Fig.8, the binaural spatial processor is constructed by the following extensions:

- *Direct sound:* The stereo panpot is replaced by a binaural panpot. To improve the accuracy of the simulation, an additional filter should be inserted to equalize the source signal and simulate the air absorption.

- *First-order reflections:* The early reflections are individually controlled as left-right pairs, with, for every odd i: $\Delta t_i = t_{i+1}-t_i$, $b_i = e_i \cdot g_{l,i}$ and $b_{i+1} = e_i \cdot g_{r,i}$. Thus, the date, amplitude, and direction (lateralization) of a binaural reflection are defined by the 4 values (t_i, t_{i+1}, b_i, b_{i+1}). An improved model of the early reflections is obtained by inserting an average binaural filter at the output of the early reflection generator, as defined in section 3.3.2 by Eq. (11).

Meta Platforms, Inc.
Exhibit 1076
Page 029

- *Later reflections and reverberation:* Since the two output signals of the reverberant filter are uncorrelated, the IACC coefficient can be controlled by inserting the following matrix at the ouput of the reverberant filter [35, 3]:

$$\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{where } \theta = \text{Arc}\sin(\text{IACC}) / 2. \tag{12}$$

A more accurate late reverberation model is obtained by inserting, still at the output of the reverberant filter, an additionnal "diffuse" average filter whose essential role is to reproduce the power spectrum of the sound source, averaged over all directions. The other effects (air and wall absorption) are described by the reverberation time $Tr(f)$. No additional processing is necessary for binaural compatibility, provided that the binaural panpot and the early "oriented" average filter be diffuse-field normalized.

The stereo reverberator can thus be transformed into a binaural reverberator for virtually no additional processing cost. To improve the accuracy of the room simulation, two additionnal spectral correctors are introduced: the "oriented" average filter for early reflections and the "diffuse" average filter for the late reverberation. The task of adjusting or designing the "diffuse" average filter is simple, since it only involves the sound source, due to the assumption of a diffuse field. The specification of the "oriented" average filter, according to section 3.3.2, is more involved. This can be a limitation in real-time applications where movements of the source or the listener can modify the direction and/or the spectrum of each early reflection. An approximate solution to this problem is to use the "diffuse" average filter for both the early and late reflections.

This leads to 3 possible models for reproducing the early reflections:
1. individual spectral and binaural processing of each reflection,
2. individual gain control and lateralization, plus "oriented" average filter,
3. individual gain control and lateralization, plus "diffuse" average filter.

To compare these three models, a preliminary psychoacoustical experiment was carried out, in which subjects could instantly switch between the different models of rendering the early reflections on headphones, using listener-specific HRTFs [54, 5]. The artificial room effects comprised 7 discrete reflections, differing only in amplitude and direction, followed by an exponentially decaying reverberation. The analysis led to the following conclusions:
- the use of the "oriented" average filter could not be distinguished from the individual binaural simulation of each early reflection, except in the case of a delayed lateral reflection (later than 40 ms) emerging from the distribution.
- the use of the "diffuse" average filter for the early reflections cannot be detected when the reflections come from the back of the listener or if they are more than 3 dB below the direct sound.

These results indicate that, for many applications, it is sufficient to provide a global equalization of the room effect taking into account the power radiated by the source, while the discrete early reflections are individually "lateralized" by use of a simplified binaural panpot reproducing only the two prevalent binaural cues (interaural gain and delay difference). This means that the binaural room simulator and the stereo room simulator are identical from the point of view of the algorithm used, and differ only in the gain and delay panning laws used to lateralize the early reflections. The

Meta Platforms, Inc.
Exhibit 1076
Page 030

same algorithm can thus simulate a variety of conventional stereo sound pickup techniques differing by the distance, orientation, and directivities of the microphones.

3.4 A practical application

In this section, we describe the application of the above models to the design of a two-channel Spatialisateur. In this project, the artificial reverberation algorithm of Fig. 8 was extended in order to provide a finer decomposition of the time structure of the impulse response, as shown on Fig 10a. This decomposition arises from psycho-experimental studies on the objective and perceptual characterisation of room acoustical quality, and includes an intermediate packet of reflections, denoted *R2* [99]. This intermediate packet is realized by the "second-order reflections" produced by the reverberator shown on Fig. 10b.

Temporal effects. This reverberator is an extension of the algorithm of Fig. 8, where an intermediate processing stage, comprising the matrix *M2* and the bank of delay lines *T2*, has been inserted between the early reflection generator and the reverberant filter. The matrices *M2* and *M3* have maximum "density" (in the sense of section 3.2) and an absorbent filter is associated to each delay line in *T3* to control the reverberation time $Tr(f)$ according to Eq. (8). The delay lengths in *T1* and *T2* can be set so that the different sections in the response fit in the lime limits defined by Fig. 10a. If P denotes the number of channels (i.e. the dimension of the matrices *M2* and *M3* and the number of delay lines in *T2* and *T3*), the number of operations for implementaing this reverberator is approximately $2 \cdot P \cdot (\log_2(P)+3)$ per output sample. A high-quality version is obtained for $P=8$ (about 100 operations) and a simpler version, adequate for simulating relatively small rooms, is obtained for $P=4$ (about 40 operations).

Directional effects. The direct sound and the multi-channel reverberator outputs are converted to binaural format (diffuse-field normalized). In this model, the direct sound (*DS*) and the early reflections (*R1*) are directionnal, while the later reflections (*R2*, *Rev*) are assumed to be diffuse. These later reflections just go through a P-by-2 matrix allowing to control the IACC according to Eq. (12). The reflections *R1* are transformed by a localization module into Q "stereo reflections" with individual arrival times, amplitudes, and directions. The direct sound goes through a binaural localization filter (or "panpot"). Assuming 150 operations for the binaural panpot, the total cost of the directionnal synthesis, is $2 \cdot (Q+P+2)+150$, i.e. about 190 operations if $P=8$ and $Q=4$.

Spectral effects. Each one of the 4 components (*DS*, *R1*, *R2*, *Rev*) goes through a 2-channel level and spectrum corrector, which combines effects associated to the room (sound attenuation during propagation and reflections) and the source (position and radiation). These four spectral correctors can be implemented as parametric equalizers [83-85]. In our implementation, using second order IIR filters, the spectral corrections, along with the reverberation time, can be adjusted in three independent frequency bands, with adjustable cross-over frequencies. Since 7 filters are necessary, this entails a additionnal processing cost of 35 operations per output sample.

Output filter. The two output signals can be processed, if necessary, through an output equalization stage. This is either a cross-talk canceller of the type shown on Fig. 5b(b), or a headphone equalizer, whose transfer function is given, for each ear, by the magnitude ratio of the diffuse-field HRTF over the measured headphone frequency response (see part 1). If the binaural output signal

Meta Platforms, Inc.
Exhibit 1076
Page 031

is played on diffuse-field calibrated headphones, the output filter is not necessary, unless a listener-specific calibration is desired. In our implementation, this output filter is implemented with two IIR filters or order 16, which involves about 70 operations per sample.

This spatial processor imitates a two-channel recording made in a natural environment. By selecting different panpot methods, the same algorithm can be configured for binaural recording or conventional stereo recording techniques (e.g. AB, XY, MS...). To imitate a recording in a real environment, an analysis/synthesis procedure can be used to set the spectral correctors and the reverberation time according to an impulse response measured with an omnidirectional microphone. The same can be obtained from a computer model of a virtual room, yielding at the same time the parameters of the early reflection distribution. For maximum accuracy, if real-time control is not necessary, the four spectral correctors and the absorbent filters used to control the reverberation time can be approximated using IIR or FIR design methods (as in [3, 97]).

Computational requirements

The only significant additional processing cost in transforming the stereo spatial processor into a binaural spatial processor lies in the binaural synthesis of the direct sound. The computational requirements, expressed in operations (multiply-adds) per sample, can be decomposed as follows:
- 180 operations for the stereo reverberator,
- 100 to 150 operation for the binaural synthesis filter (according to section 2.4),
- 70 operations for the output filter, if necessary (transaural mode or headphone calibration).

We thus reach a total of 330 operations per output sample (or 400 in transaural mode). Although this evaluation only takes arithmetic operations into account, it appears that a recent DSP allows to implement the binaural Spatialisateur in real time. Furthermore, this is a rather elaborate algorithm designed to satisfy demanding studio or computer music applications. For some applications, a simpler reverberation algorithm and a less accurate reproduction of the direct sound are acceptable, so that two or more spatial processors could run simultaneously on the same DSP. This evaluation is valid for time-varying implementations, allowing the use of a headtracker and continuous movements of the sound source relative to the listener.

Control interface

In our current implementation, the control interface allows the listener to use his/her personal HRTFs or to select instantly, among five different "heads", the set of HRTFs which yield the most natural auditory sensation. The direction of the sound source can be continuously varied in real time, as well as the room effect, which can be controlled using the following parameters:
- the distribution of the early reflections in time, amplitude and direction,
- the time limits of the three temporal sections of the room effect (see Fig. 10a),
- the energies of these three sections and the reverberation time, with their spectral variation.

Alternatively, a "perceptual" control interface can be used, derived from psychoacoustical research carried out at IRCAM on the perception of room acoustical quality in concert halls and opera houses [99]. This research has led to the definition of mutually independent perceptual factors, nine of which are currently implemented in the Spatialisateur:

Meta Platforms, Inc.
Exhibit 1076
Page 032

- *source proximity*	early sound: energy of direct sound and early room effect
- *brilliance* and *warmth*	variation of early sound with frequency
- *room presence*	late sound: energy of later reflections and reverberation
- *running reverberance*	early decay time
- *envelopment*	energy of early room effect relative to direct sound
- *late reverberance*	late decay time
- *liveness* and *intimacy*	variation of late decay time with frequency

4 CONCLUSION

We have described techniques for modeling and implementing head-related transfer functions (HRTFs) under the form of digital filters, taking advantage of the minimum-phase approximation and the diffuse-field normalization of the HRTFs. Cost-efficient implementations are obtained by IIR design methods applicable to both binaural or transaural synthesis. FIR design is also effective in the case of the binaural "panpot" because the filter length can be reduced to 1.5 ms (or even 1 ms), and because a simple but relevant interpolation process can be implemented to allow head-tracking or dynamic movements of sound sources. With the proposed methods, the binaural or transaural "panpot" requires less than 100 to 150 operations (multiply-adds) per sample, at a 48 kHz sample rate. This cost can be divided by two with IIR designs in "static" applications (including the transaural "cross-talk canceller" or a headphone equalizer, if needed).

To synthesize complex fields with room reflections and reverberation, an extension is proposed from the classic binaural mixing console design (where each reflection is synthesized as an additional virtual sound source). The spatial processor described here includes a cost-efficient auralization method to synthesize early reflections, preserving the prevalent binaural cues (interaural delay and time difference) individually for each reflection, while spectral cues are grouped in an "average" binaural filter. The computational cost of the binaural reverberator, including a feedback delay network to synthesize the later diffuse reverberation, is less than twice that of the binaural panpot. The total cost for spatial processing of one sound source can thus be handled by a single programmable digital signal processor, allowing real-time binaural or transaural synthesis of source localization as well as room reflections and reverberation, and yielding faithful control of the subjective distance.

Due to the cost-efficiency of the algorithm, it is possible to envision a new generation of digital mixing consoles including a room simulator in each channel, allowing unrestricted binaural synthesis of three-dimensional sound images, and also applicable to conventional stereo recording. For studio and computer music applications, this evolution may call for a new kind of perceptually based user interfaces to control the room simulation. Similarly, auditory displays as used in multimedia or virtual reality applications can be extended to include the convincing reproduction of room reflections and reverberation. In some applications, additional cost savings can be achieved by further simplifying the room simulation model, or by sharing the late reverberation module between several virtual sound sources (assumed to be in the same room). Finally, auralization systems for use in architectural acoustics may evolve towards real-time operation, allowing instant monitoring of modifications in source or listener location, room geometry or wall materials.

Meta Platforms, Inc.
Exhibit 1076
Page 033

5 ACKNOWLEDGEMENTS

The authors would like to thank Martine Marin, Andre Gilloire and Mireille Mein for collaboration in the measurement and design of the binaural filters. Also, for ideas and discussions on the implementation of digital filters: Frederic Bimbot, Jean Laroche, Jacques Prado, Eric Moulines and Nicolas Moreau (of Telecom Paris), as well as Philippe Depalle and Xavier Rodet (of IRCAM). Support was received from Studer Digitec and Association Nationale pour la Recherche Technique at earlier stages of this work, under the supervision of Antoine Chaigne.

Commercial applications of artificial reverberation methods described in part 3 of this paper are covered by French patent no. 92 02528, assigned to France Telecom [100] (extensions pending).

6 REFERENCES

[1] G. Bloch, G. Assayag, O. Warusfel, J.-P. Jullien, "Spatializer: from room acoustics to virtual acoustics", *Proc. Int. Computer Music Conference*, 1992.
[2] M. Marin, "Etude de la localisation en restitution des sons. Application à l'amélioration des systèmes de téléconférence", rapport CNET NT/LAA/TSS, 1993.
[3] J.-M. Jot, "Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs", doctoral dissertation, Télécom Paris, 1992.
[4] F. Poncet, "Simulation de la localisation de sources sonores dans l'espace", rapport Télécom Paris, Dépt. Signal, 1992. (partly covered in [3])
[5] J.-M. Jot, O. Warusfel, E. Kahle, M. Mein, "Binaural concert hall simulation in real time", presented at the IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz), Oct 1993. (paper available from the authors)
[6] M. Marin, A. Gilloire, J.-F. Lacoume, O. Warusfel, J.-M. Jot, "Environnement de simulation pour l'évaluation psychoacoustique des systèmes de prise et de restitution du son dans un contexte de téléconférence", *Proc. 3rd French Congress on Acoustics (Toulouse)*, April 1994.
[7] J.-M. Jot, "Spatialisateur: the *Spat~* processor and its library of Max objects - Introduction", IRCAM documentation, Oct. 1994.
[8] J. Blauert, *Spatial Hearing: the psychophysics of human sound localization*, Cambridge MIT Press, 1983.
[9] C. Pösselt, J. Shroter, M. Opitz, P.L. Dyvenyi, J. Blauert, "Generation of binaural signals for research and home entertainment", *Proc. 13th Int. Conf. Acoustics*, vol. 1, pp. B1-6, 1986.
[10] O. Warusfel, "Etude des paramètres liés à la prise de son pour les applications d'acoustique virtuelle", *Proc. 1rst French Congress on Acoustics*, vol. 2, pp. 877-880, 1990.
[11] H. Kuttruff, *Room Acoustics*, 2nd edition. Applied Science Publishers Ltd, London, 1979.
[12] M. Kleiner, B.-I. Dalenbäck, P. Svensson, "Auralization - An overview", *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861-875, 1993.
[13] M.R. Schroeder, B.S. Atal, "Computer simulation of sound transmission in rooms", *IEEE Conv. Record*, pt. 7, pp. 150-155, 1963
[14] M.R. Schroeder, "Digital simulation of sound transmission in reverberant spaces", J. Acoust. Soc. Am., vol. 47, no. 2, pp. 424-431, 1970.

Meta Platforms, Inc.
Exhibit 1076
Page 034

[15] D.H. Cooper, J.L. Bauck, "Prospects for transaural recording", *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3-19, 1989.
[16] G.S. Kendall, "Directional sound processing in stereo reproduction", *Proc. Int. Computer Music Conference*, pp. 261-264, 1992.
[17] J.L. Bauck, D.H. Cooper, "Generalized transaural stereo", *Proc. 93rd AES Convention (San Francisco)*, preprint 3401, 1992.
[18] A. Persterer, "A very high performance digital audio processing system", *Proc. 13th Int. Conf. Acoustics (Belgrade)*, 1989.
[19] F. Richter, A. Persterer, "Design and applications of a creative audio processor", *Proc. 86th AES Convention (Hamburg)*, preprint 2782, 1989.
[20] H. W. Gierlich, K. Genuit, "Processing artificial-head recordings", *J. Audio Eng. Soc.*, vol. 37, no. 1/2, 1989.
[21] W. Pompetzki, "Binaural recording and reproduction for documentation and evaluation", *Proc. AES 8th International Conference*, pp. 225-229, 1990.
[22] D. Griesinger, "Binaural techniques for music reproduction", *Proc. 8th A.E.S. International Conference*, pp. 197-207, 1990.
[23] M. Wöhr, G. Theile, H.-J. Goeres, A. Persterer, "Room-related balancing technique: a method for optimizing recording quality", *J. Audio Eng. Soc.*, vol. 39, no. 9, pp. 623-631, 1991
[24] A. Persterer, "Binaural reproduction of an 'ideal control room' for headphone reproduction", *Proc. 90th AES Convention (Paris)*, preprint 3062, 1991.
[25] P. Rubak, "Headphone signal processing system for out-of-head localization", *Proc. 90th AES Convention (Paris)*, preprint 3063, 1991.
[26] S. H. Foster, E.M. Wenzel, R.M.Taylor, "Real-time synthesis of complex acoustic environments", *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz)*, 1991.
[27] E.M. Wenzel, S. H. Foster, F.L. Wightman, D.J. Kistler, "Realtime digital synthesis of localized auditory cues over headphones", *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz)*, 1991.
[28] H. Lehnert, "Real-time generation of interactive virtual auditory environments", *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz)*, 1993.
[29] D.R. Begault, "The composition of auditory space: recent developements in headphone music", *Leonardo*, vol. 23, no. 1, pp. 45-52, 1990
[30] R. Bidlack, D. Blaszczak, G.S. Kendall, "An implementation of a 3D binaural audio system within an integrated virtual reality environment", *Proc. Int. Computer Music Conference*, pp. 442-445, 1993.
[31] S.T. Pope, L.E. Fahlén, " The use of 3-D audio in a synthetic environment: an aural renderer for a distributed virtual reality system", *Proc. Int. Computer Music Conference*, pp. 146-149, 1993.
[32] J. Huopaniemi, M. Karjalainen, V. Välimäki, T. Huotilainen, "Virtual instruments in virtual rooms - a real time binaural room simulation environment for physical models of musical instruments", *Proc. Int. Computer Music Conference*, pp. 455-462, 1994.
[33] M. Cohen, N. Koizumi, "Exocentric control of audio imaging in binaural telecommunication", *IEICE Trans. Fundamentals Elec. Comm. Comp. Sc.*, vol. E75-A, no. 2, pp. 164-170, 1992.
[34] D.R. Begault, "Multichannel spatial auditory display for speech communication", *J. Audio Eng. Soc.*, vol. 42, no. 10, Oct. 1994.

Meta Platforms, Inc.
Exhibit 1076
Page 035

[35] J. Martin, D. Van Maercke, J.-P. Vian, "Binaural simulation of concert halls: a new approach for the binaural reverberation process", *J. Acou. Soc. Am.*, vol. 94, no. 6, 1993.
[36] K.H. Kuttruff, "Auralization of impulse responses modeled on the basis of ray-tracing results", *J. Audio Eng. Soc.*, vol. 41, no. 11, 1993.
[37] W. Ahnert, R. Feistel, "EARS auralization software", *J. Audio Eng. Soc.*, vol. 41, no. 11, 1993.
[38] K. Genuit, "Sound analysis and synthesis in consideration of the binaural signal processing", *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz)*, 1991.
[39] G.S. Kendall, "PinnaWorks: a NeXT application for three-dimensional sound processing in real time", *Proc. Int. Computer Music Conference*, pp. 118-121, 1993.
[40] F.L. Wightman, D.J.Kistler, "Headphone simulation of free-field listening II: psycho-physical validation", *J. Acou. Soc. Am.*, vol. 85, pp. 868-878, 1989.
[41] D.R. Begault, "Challenges to the successful implementation of 3-D sound", *J. Audio Eng. Soc.*, vol. 39, no. 11, pp. 864-870, 1991.
[42] D.R. Begault, "Perceptual effect of synthetic reverberation on three-dimensional audio syatems", *J. Audio Eng. Soc.*, vol. 40, no. 11, pp. 895-904, 1992.
[43] D.R. Begault, "Binaural auralization and perceptual veridicality", *Proc. 93rd AES Conv. (San Francisco)*, preprint 3421, 1992.
[44] D. Hammershoi, J. Sandvad, " Binaural auralization. Simulating free-field conditions by headphones", *Proc. 96th AES Convention (Amsterdam)*, preprint 3863, 1994.
[45] J.N. Mourjopoulos, "Digital equalization of room acoustics", *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884-900, Nov. 1994.
[46] F.L. Wightman, D.J.Kistler, "Headphone simulation of free-field listening I: stimulus synthesis", *J. Acou. Soc. Am.*, vol. 85, pp. 858-867, 1989.
[47] D. Hammershoi, H. Moller, "Head-related transfer functions: measurements on 40 human subjects", *Proc. 92nd AES Convention (Amsterdam)*, preprint 3289, 1994.
[48] W.G. Gardner, K. Martin, "HRTF measurements on a KEMAR dummy-head microphone", Tech. report. #280, MIT Media Lab Perceptual Computing, 1994.
[49] J-P. Jullien, A. Gilloire, A. Saliou, "Mesure des réponses impulsionnelles en acoustique", Note technique CNET NT/LAA/TSS/181, Nov. 1984.
[50] J. Borish, J.B. Angel, "An efficient algorithm for measuring the impulse response using pseudo-random noise", *J. Audio Eng. Soc.*, vol. 31, pp. 478-488, 1983.
[51] D.D. Rife, J. Vanderkooy, "Transfer-function measurement with maximum-length sequences", *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419-444, 1989.
[52] A.V. Oppenheim, R.W. Shafer, *Digital Signal Processing*, Prentice Hall, 1975.
[53] S. Mehrgardt, V. Mellert, "Transformation characteristics of the external human ear", *J. Acou. Soc. Am.*, vol. 61, no. 6, pp. 1567-1576, 1977.
[54] M. Mein, "Perception de l'information binaurale liée aux réflexions précoces dans une salle. Application à la simulation de la qualité acoustique", mémoire de DEA, Univ. du Maine, Le Mans, Sept. 1993.
[55] G. Theile, "On the standardization of the frequency response of high-quality studio headphones", *J. Audio Eng. Soc.*, vol. 34, no. 12, Dec. 1986.
[56] M.A. Gerzon, "Applications of Blumlein shuffling to stereo microphone techniques", *J. Audio Eng. Soc.*, vol. 42, no. 6, pp. 435-453, 1994.
[57] J. Sandvad, D. Hammershoi, " Binaural auralization. Comparison of FIR and IIR filter representation of HIRs", *Proc. 96th AES Convention (Amsterdam)*, preprint 3862, 1994.

Meta Platforms, Inc.
Exhibit 1076
Page 036

[58] D.H. Cooper, "Problems with shadowless stereo theory: asymptotic spectral status", *J. Audio Eng. Soc.*, vol. 35, pp. 629-642, 1989.
[59] G. Thiele, "On the naturalness of two-channel stereo sound", *J. Audio Eng. Soc.*, vol. 39, no. 10, pp. 761-767, 1991.
[60] G.S. Kendall, W.L. Martens, M.D. Wilde, "A spatial sound processor for loudspeaker and headphone reproduction", *Proc. 8th AES Int. Conf.*, pp. 209-221, 1990.
[61] D.J.Kistler, F.L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", *J. Acou. Soc. Am.*, vol. 91, pp. 1637-1647, 1992.
[62] H.L. Han, "Measuring a dummy-head in search of pinna cues", *J. Audio Eng. Soc.*, vol. 42, no. 1/2, pp. 15-37, 1994.
[63] W.G. Gardner, "Efficient convolution without input/output delay", *Proc. 97th AES Convention (San Francisco)*, 1994.
[64] J.O. Smith, "Techniques for digital filter design and system identification with application to the violin", Ph.D. dissertation, CCRMA, Dept of Music, Stanford Univ., June 1983.
[65] G.S. Kendall, W.L. Martens, "Simulating the cues of spatial hearing in natural environments", *Proc. Int. Computer Music Conference*, pp. 111-125, 1984.
[66] C.J. McCabe, D.J. Furlong, "Spectral stereo surround sound pan-pot", *Proc. 90th A.E.S. Conv. (Paris).*, preprint 3067, 1991.
[67] O. Warusfel, "Modélisation paramétrique de phénomÄnes acoustiques simples", doctoral dissertation, Univ. P. M. Curie, 1985.
[68] J.N. Mourjopoulos, "On the variation and invertibility of room impulse response functions", *J. Sound Vib.*, vol. 102, no. 2, pp. 217-228, 1985.
[69] J.N. Mourjopoulos, "Digital equalization methods for audio systems", *Proc. 84th A.E.S. Conv.*, preprint 2598, 1988.
[70] T.P. Krauss, L. Shure, J.N. Little, *Signal processing toolbox for use with Matlab*, The MathWorks, Inc., Feb. 1994.
[71] B. Friedlander, B. Porat, "The modified Yule-Walker method of ARMA Spectral estimation", *IEEE Trans. Aerospace Electronic Systems*, vol. 20, no. 2, pp. 158-173, March 1984.
[72] J. Dattorro, "The implementation of recursive digital filters for high-fidelity audio", *J. Audio Eng. Soc.*, vol. 36, no. 11, pp. 851-878, 1988.
[73] D.C. Massie, "An engineering study of the four-multiply normalized ladder filter", *J. Audio Eng. Soc.*, vol. 41, no. 7/8, pp. 564-582, 1993.
[74] R. Wilson, "Filter topologies", *J. Audio Eng. Soc.*, vol. 41, no. 9, pp. 667-678, 1993.
[75] J. Laroche, "Traitement des signaux audio-fréquences", Télécom Paris, Dept Signal, 1995.
[76] J.D. Markel, A.M. Gray, *Linear prediction of speech*, Springer-Verlag, Berlin, 1976.
[77] E.M. Wenzel, S. H. Foster, "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis", *Proc. IEEE Workshop on Applications of Digital Signal Processing to Audio and Acoustics (New Paltz)*, 1993.
[78] E. Skudrzyk, *The Foundations of Acoustics*, Springer-Verlag, NewYork, 1971.
[79] J. Chowning, "The simulation of moving sound sources", *J. Audio Eng. Soc.*, vol. 19, no. 1, pp. 2-6, 1971.
[80] V. Larcher, "Interpolation de filtres audio-numériques pour application à la reproduction spatiale des sons sur écouteurs", rapport Télécom Paris, Dépt. Signal, 1995.
[81] J.R. Haskew, J.M. Kelly, R.M. Kelly, T.H. McKinney, "Results of a study of the linear prediction vocoder", *IEEE Trans. Communications*, vol. 21, no. 9, Sept. 1973.

Meta Platforms, Inc.
Exhibit 1076
Page 037

[82] R. Viswanathan, J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems", *IEEE Trans. Acou. Speech Signal Process.*, vol 23, pp. 309-321, 1975.
[83] P.A. Regalia, S.K. Mitra, "Tunable digital frequency response equalization filters", *IEEE Trans. Acou. Speech Signal Process.*, vol 35, no. 1, 1975.
[84] P. Dutilleux, "Simple to operate digital time-varying filters", *Proc. 86th A.E.S. Conv.*, preprint 2757, 1989.
[85] J.N. Mourjopoulos, E.D. Kyriakis-Bitzaros, C.E. Goutis, "Theory and real-time implementation of time-varying digital audio filters", *J. Audio Eng. Soc.*, vol. 38, no. 7/8, pp. 523-536, 1990.
[86] Y. Grenier, "Modélisation de signaux non stationnaires", doctoral dissertation (Doctorat d'Etat), Université d'Orsay, Oct 1984.
[87] W. Verhelst, P. Nilens, "A modified-superposition speech synthesizer and its applications", *Proc. IEEE Int. Conf. Acou. Speech and Signal Proc. (Tokyo)*, pp. 2007-2010, 1986.
[88] M.R. Schroeder "Natural sounding artificial reverberation", J. Audio Eng. Soc., vol.10, no. 3, pp. 219-223, 1962.
[89] M.A. Gerzon, "Unitary (Energy preserving) multichannel networks with feedbacks", *Electronics Letters*, vol. V, no. 12-11, 1976.
[90] J.A. Moorer, "About this reverberation business", *Computer Music Journal*, vol. 3, no. 2, pp. 13-18, 1979.
[91] J. Stautner, M. Puckette, "Designing multi-channel reverberators", *Computer Music Journal*, vol. 6, no. 1, pp. 52-65, 1982.
[92] J.O. Smith, "A new approach to digital reverberation using closed waveguide networks", *Proc. Int. Computer Music Conference*, pp. 47-63, 1985.
[93] G.S. Kendall et al., "Image model reverberation from recirculating delays", *Proc. 81st AES Conv. (Los Angeles)*, preprint 2408, 1986.
[94] J.-M. Jot, A. Chaigne "Digital delay networks for designing artificial reverberators", *Proc. 90th AES Conv. (Paris)*, preprint 3030, 1991.
[95] D. Griesinger, "Practical processors and programs for digital reverberation", *Proc. 7th A.E.S. International Conference*, pp. 187-195, 1989.
[96] D. Griesinger, "Improving room acoustics through time-variant synthetic reverberation", *Proc. 90th AES Conv. (Paris)*, preprint 3014, 1991.
[97] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation", *Proc. IEEE Int. Conf. Acou. Speech and Signal Proc. (San Francisco)*, paper no. 675, 1992.
[98] M. Schoenle, N. Fliege, U. Zolder, "Parametric approximation of room impulse responses by multirate systems", *Proc. IEEE Int. Conf. Acou. Speech and Signal Proc.* , vol. I, pp. 153-156, 1993.
[99] J.-P. Jullien et al., "Some results on the objective characterisation of room acoustical quality in both laboratory and real environments", *Proc. Inst. of Acoustics*, vol. XIV, no. 2, 1992.
[100] J.-M. Jot, A. Chaigne, *Spatialisation artificielle audio-numérique*, French patent no. 92 02528, March 1992.

Meta Platforms, Inc.
Exhibit 1076
Page 038

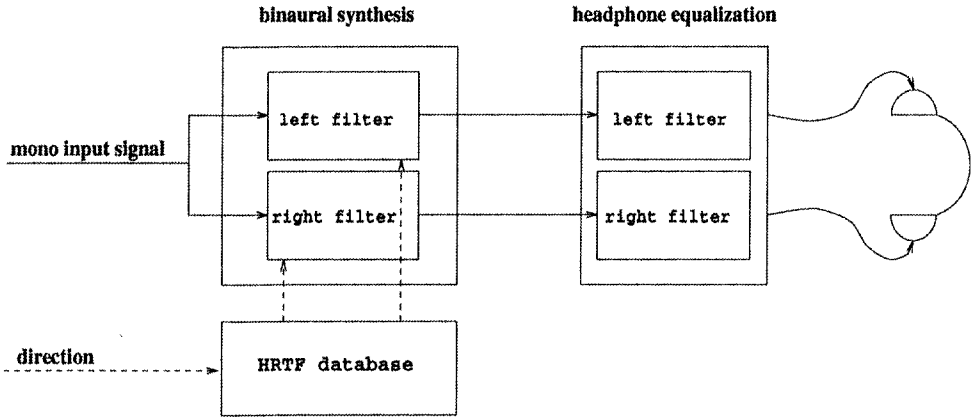


Fig. 1: Principle of binaural synthesis

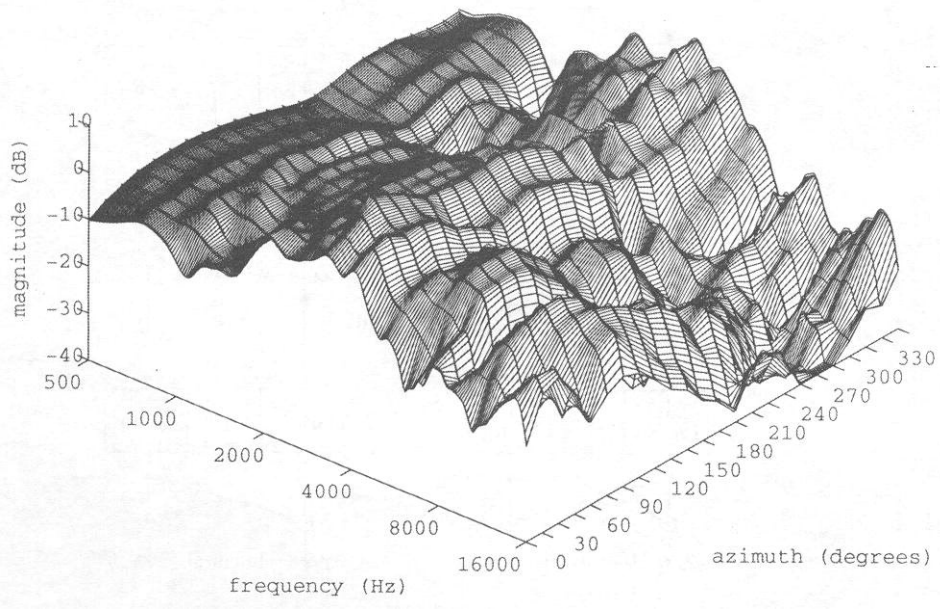


Fig. 2: Measured HRTFs in the horizontal plane (0° elevation).
The spectra are smoothed to a half-tone frequency resolution.

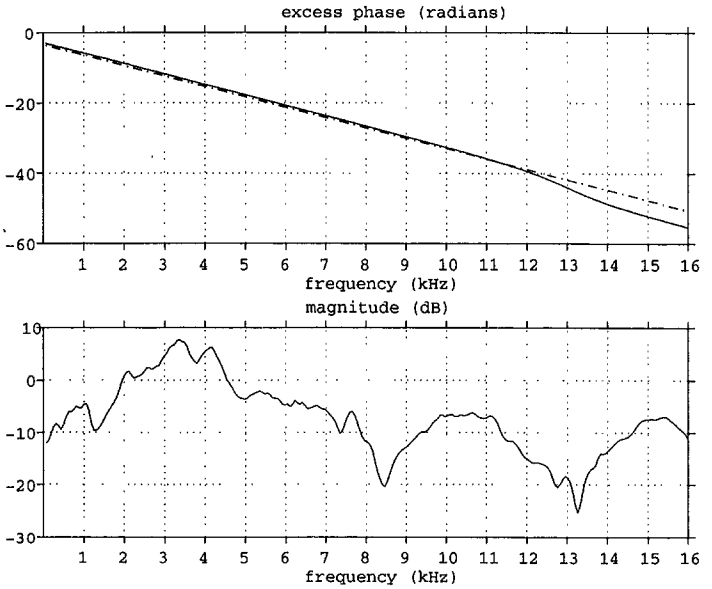


Fig. 3a: Measured HRTF at 0° elevation and 30° azimuth, showing linear approximation of excess-phase (magnitude normalized by louspeaker response).

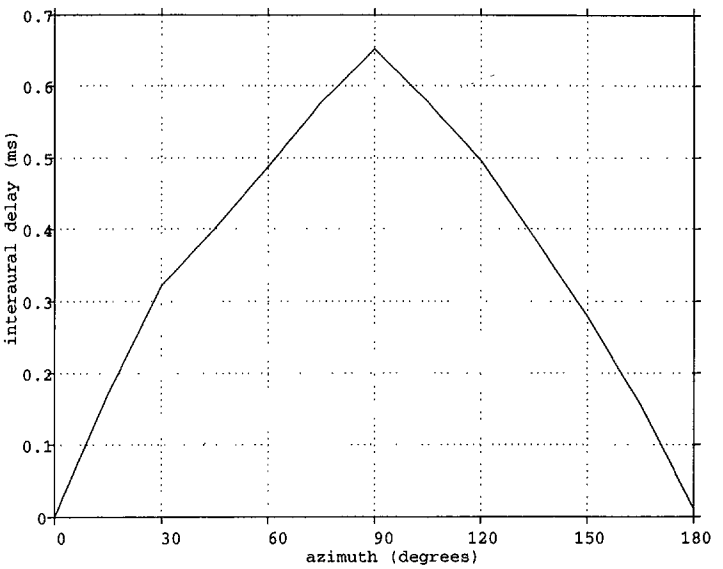


Fig. 3b: Measured interaural delay at 0° elevation, derived from linear approximation of excess-phase response.

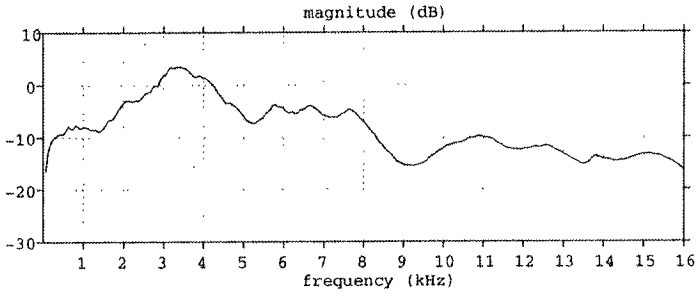


Fig. 4a: Estimated diffuse-field HRTF, computed by power-averaging of the measured HRTFs (normalized by loudspeaker response).

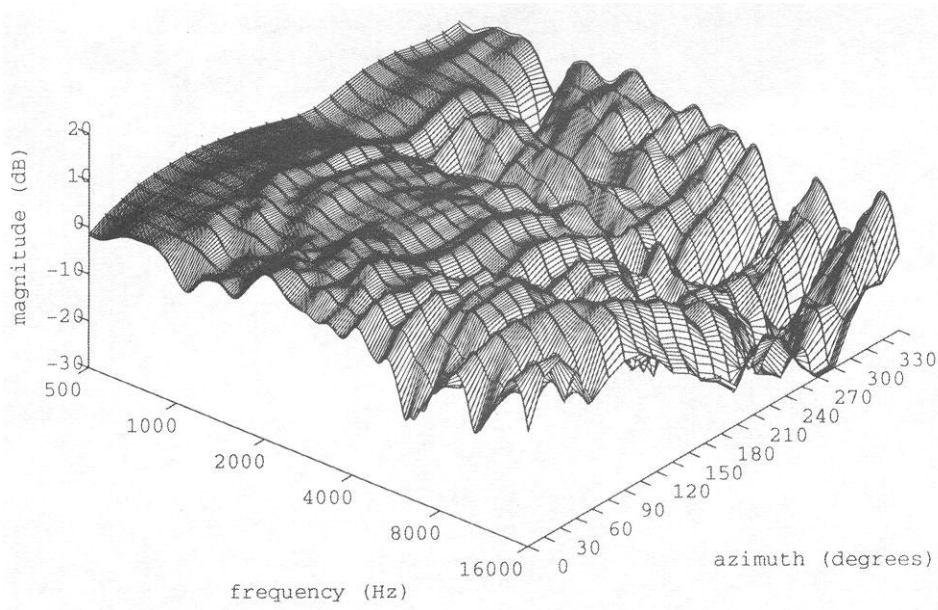


Fig. 4b: Diffuse-field normalized HRTFs in the horizontal plane (0° elevation).

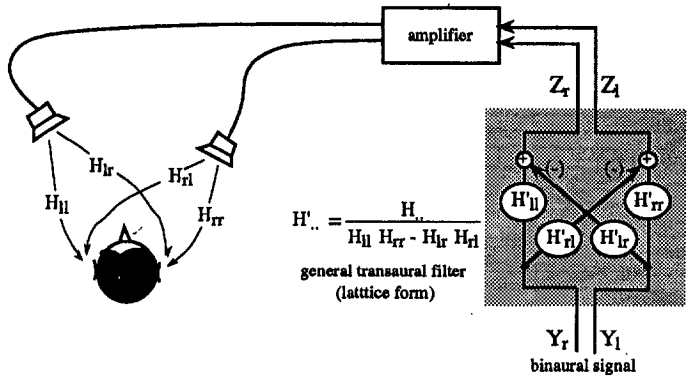


Fig. 5a: The four acoustic transfer functions in loudspeaker listening and the lattice form of the transaural cross-talk canceller.

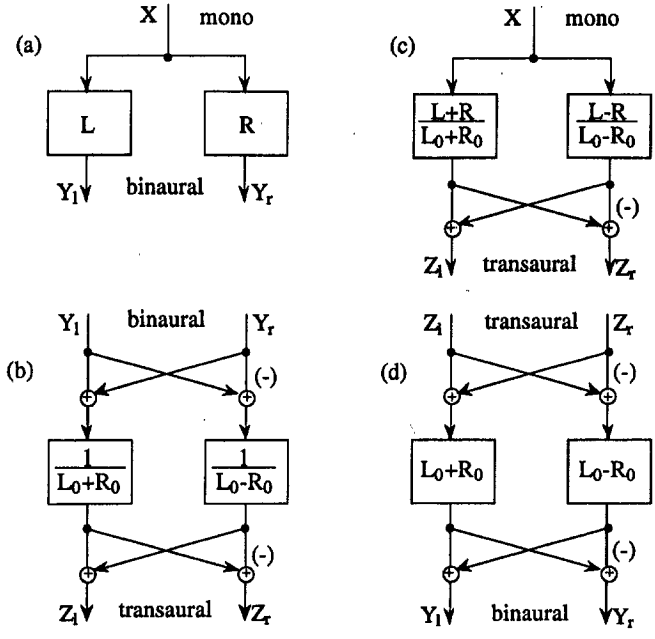


Fig. 5b: Binaural and transaural synthesis filters and format converters based on the shuffler structure from Cooper and Bauck.

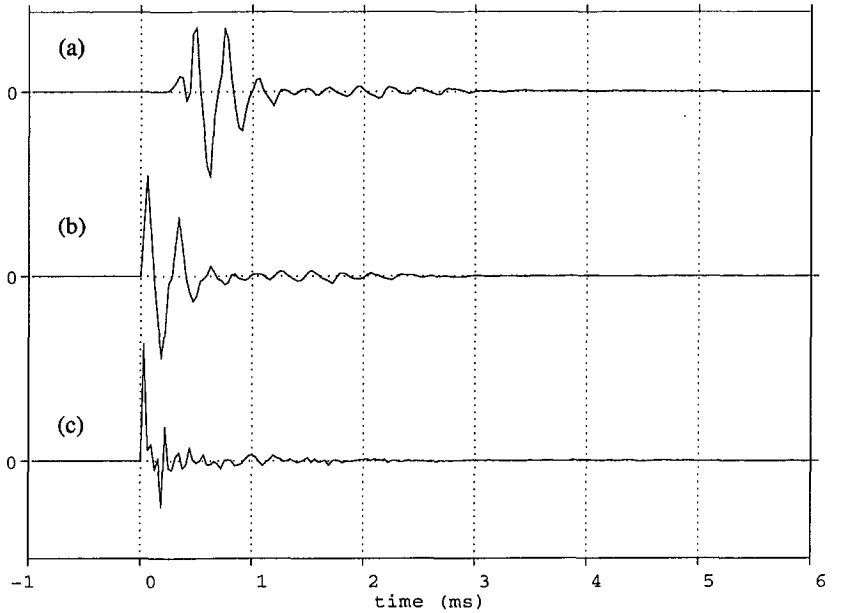
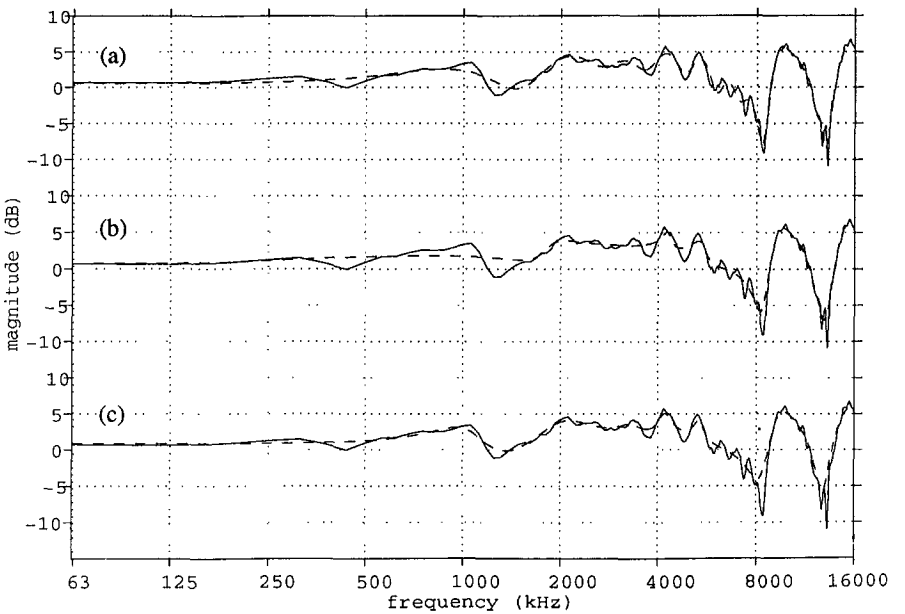


Fig. 6a: Time-domain (FIR) response (0° elevation, 30° azimuth). (a) Original measure. (b) Minimum-phase. (c) Minimum-phase with diffuse-field normalization.
Fig. 6b (below) Full line: diffuse-field normalized HRTF. Broken line: (a) 1-ms FIR design. (b) IIR design, order 16, without warping. (c) IIR design, order 16, with warping (r=0.4).



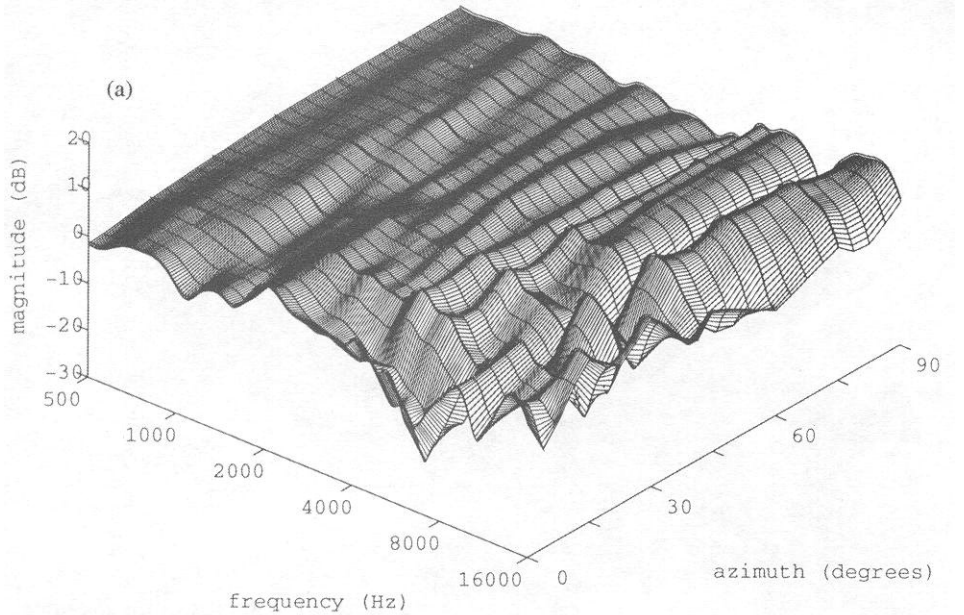
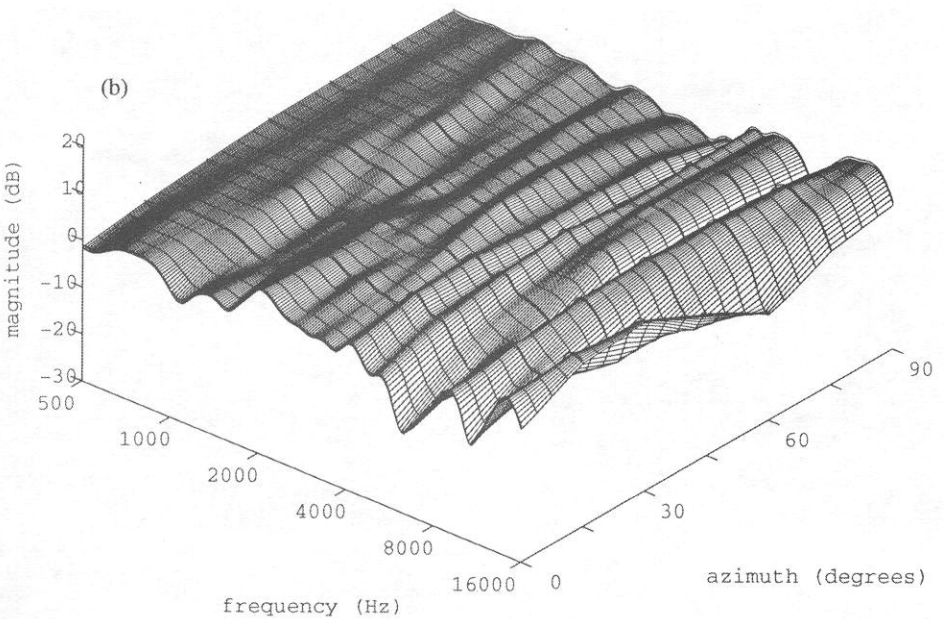


Fig. 7a: (a) Linear interpolation on impulse response coefficients with 3 intermediate steps.
(b) Same interpolation computed from minimum-phase impulse responses.
Original measured HRTFs are shown with a thicker line.



Meta Platforms, Inc.
Exhibit 1076
Page 044

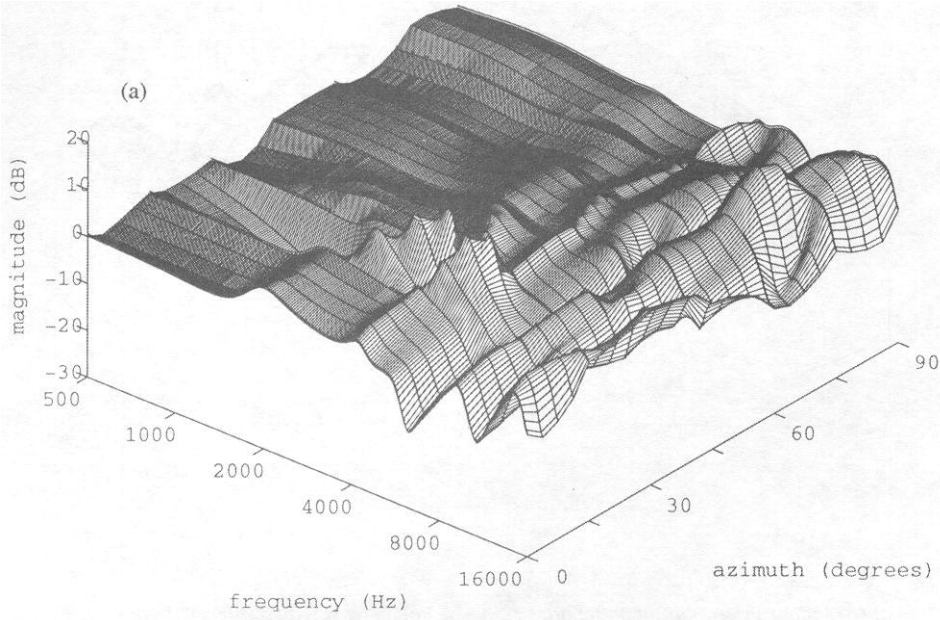
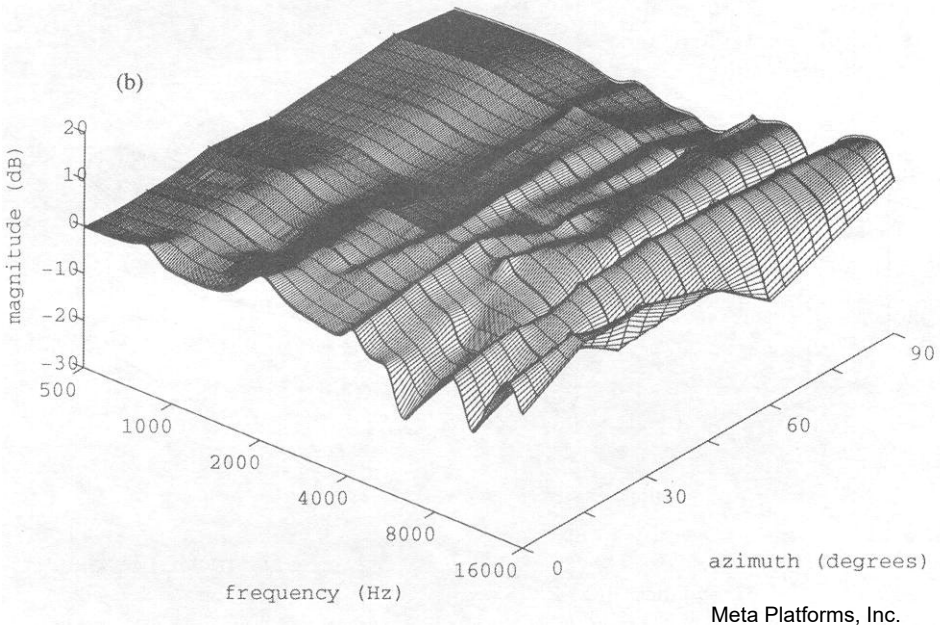


Fig. 7b: (a) Linear interpolation on coefficients of cascade second-order sections, with 3 intermediate steps(original measured HRTFs shown with thicker line). (b) Same interpolation on log area ratio coefficients.



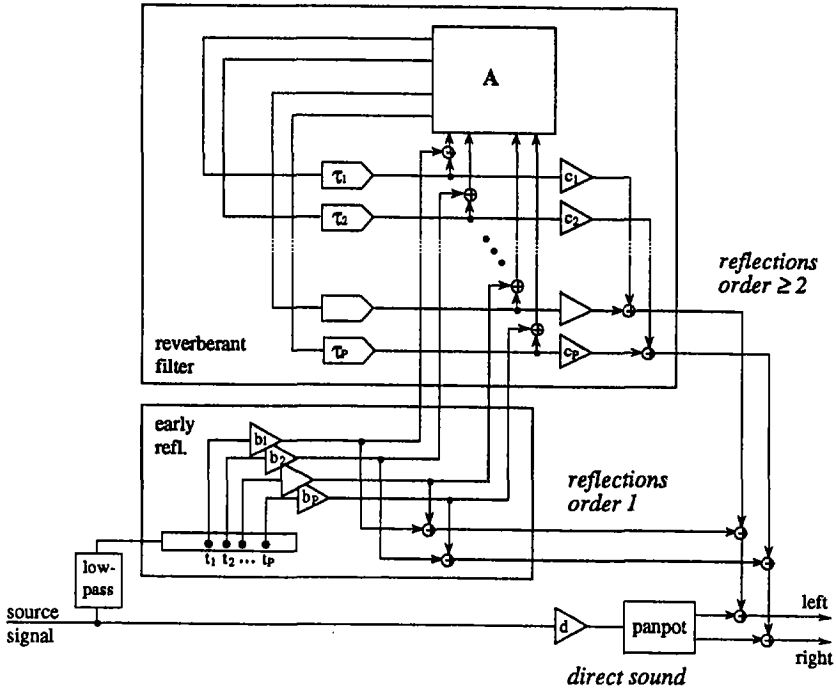


Fig. 8a: Stereo reverberator algorithm.

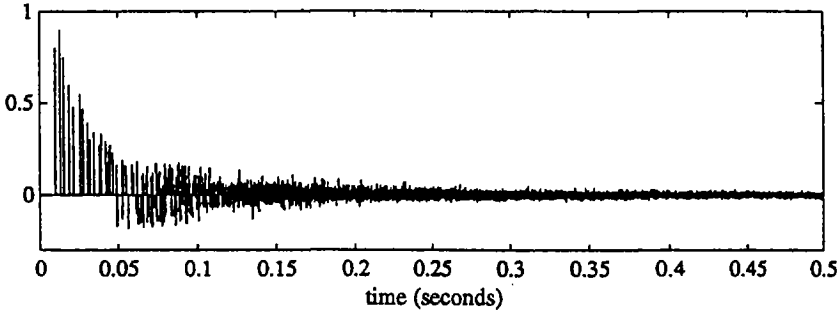


Fig. 8b: Example impulse response of the above reverberator for P=16. Modal density = 1s. Frequency-dependent controls set flat. Sum of left and right channels.

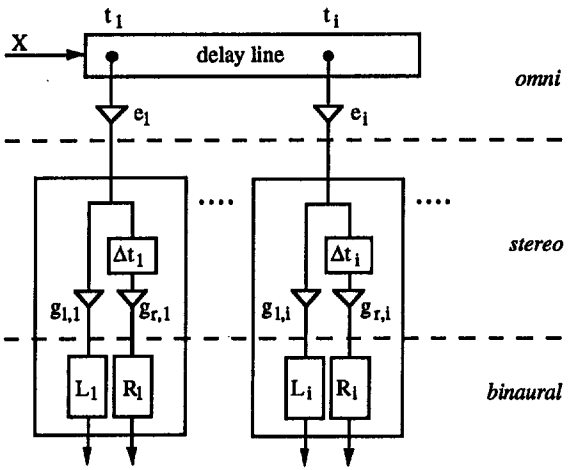


Fig. 9a: Simulation of a group of reflections, illustrating the three levels of reproduction: omnidirectional microphone, conventional stereo recording, binaural recording.

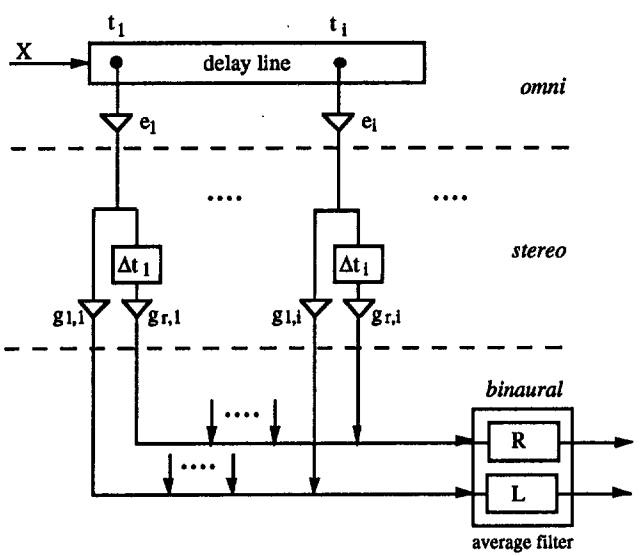


Fig. 9b: Common directional filtering of a group of reflections by use of an average directional filter.

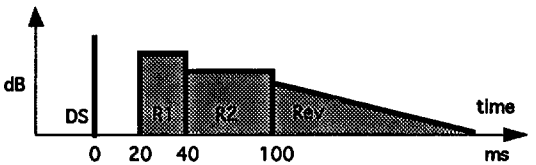


Fig. 10a: Temporal decomposition - generic room effect

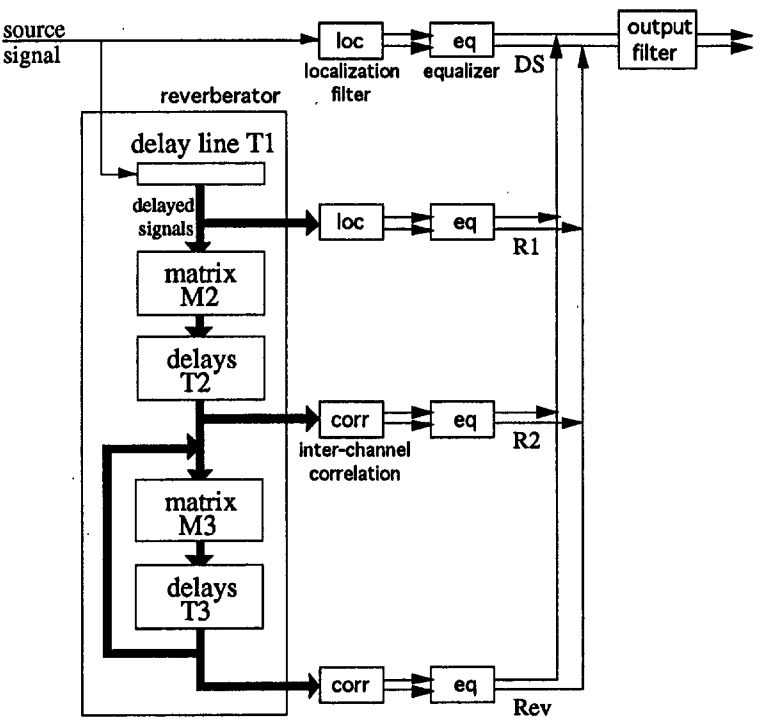


Fig. 10b: Structure of Spatialisateur algorithm for two-channel reproduction of room reflections and reverberation (binaural, transaural or conventional stereophony).