Binaural Synthesis based on Spherical Acoustics

氏名（本籍地）	Salvador Castañeda, César Daniel
学 位 の 種 類	博　士（情報科学）
学 位 記 番 号	情　博　第　６２２　号
学位授与年月日	平成２８年　９月２６日
学位授与の要件	学位規則第４条第１項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）システム情報科学専攻
学位論文題目	Binaural Synthesis Based on Spherical Acoustics
論文審査委員	（主査）東北大学教　授　鈴木　陽一
	東北大学教　授　田中　和幸
	東北大学教　授　伊藤　彰則
	東北大学准教授　坂本　修一

論 文 内 容 の 要 旨

Future media, information and communication technology will require to create the perceptual illusion of being immersed in a remote environment, whether real or imaginary, so as to enable more realistic telepresence applications. The creation of different blends between real and virtual environments is being recognized as a central stage of media tools that aim to support the synchronous collaboration among people in different locations, the safe spatial navigation and training in hazardous work environments, and the affective therapeutic experience in metal health treatments, among many other applications.

Different modes of perception, such as vision, touch and hearing, are involved when experiencing the world around us. Modern immersive systems need therefore to include displays for all the sensory modalities, so as to achieve high levels of sense of realism (the perceived correspondence between a technology-mediated experience and a similar experience not mediated by technology, in experimental conditions), sense of presence (the feeling of being immersed in a distinct environment), and sense of naturalness (the degree of consistency between the technology-mediated experience and a similar experience in real life).

Hearing plays an important role in enhancing the sense of spatial presence, fundamentally because the spatial features of sound increase the awareness of the physical surroundings and their localized events. Spatial sound also enhances the sense of social presence because it contributes to the perception of the semantic and emotional components of daily experience. The sense of naturalness, on the other hand, can also be enhanced by the spatial features of sound since they contribute to the perception of self-motion, a key phenomenon that needs to be considered when simulating the experience of being able to move around freely while interacting with a technology-mediated environment.

The spatial nature of hearing likewise allows paying attention to events outside the current visual focus and beyond the distances that can be reached with our hands. This fact is being exploited to develop virtual auditory displays for spatial navigation in situations of little or no visual attention, without reducing the necessary amount of interaction. Such kind of systems particularly seek to enable visually impaired individuals to travel through familiar and unfamiliar environments without the assistance of guides.

The development of technologies for synthesizing the spatial features of sound is therefore becoming increasingly important, not only for conveying the illusion of being immersed in a distinct environment, but also for displaying and distributing spatial information to the users when their other modes of perception are already highly loaded. Such technologies are in general referred to as spatial sound systems, and they typically involve the use of transducer arrays for capturing and presenting the acoustic variables associated with the auditory scenes.

The current study revolves around a class of spatial sound systems that seek to convey high-definition and personal listening experiences by synthesizing the sound pressure at the ears of the listeners, namely, the binaural signals. In particular, this thesis concerns the design of a binaural system for presenting sounds not only at far distances, but also in the region of space immediately surrounding our bodies where objects can be grasped and manipulated, that is, in the peripersonal space. The main purpose of this investigation is the formulation of methods of mathematical physics that will enable the future development of binaural systems for presenting sounds at far and near distances.

Spherical acoustics is the name adopted in this thesis for referring to a body of physical and mathematical work, one of the aims of which is to allow for the stable representation or encoding of spatial sound information at scalable directional resolutions. This kind of representation in turn facilitates the spatial analysis and signal processing required for instance in binaural systems.

Binaural systems based on spherical acoustics have so far been focused on presenting sounds at far distances. However, there is an increasing interest on developing virtual auditory displays for presenting sounds in the peripersonal space.

The objective of this thesis is to extend or generalize the existing spherical acoustic models to enable the development of stable spatial sound recording and optimal binaural synthesis techniques, not only for sounds at far distances, but also at near distances from the listeners' ears.

For practical purposes, the synthesis of sound on a horizontal plane at the height of the listeners' ears is also addressed in this study due to its importance in many applications. In connection with the kind of transducer array used to synthesize the spatial features of sound, three-dimensional synthesis is associated with spherical geometries, while the synthesis of sound restricted to the horizontal-plane is associated with circular geometries.

Chapter 1 introduces the motivation of this study. A survey on related systems for the synthesis of binaural signals by spherical microphone arrays is carried out, highlighting the use of the spherical harmonic functions for the analysis of the captured sound pressure field. Lastly, the specific objectives of this thesis are presented.

Chapter 2 overviews the theory of spherical acoustics, which comprises the methods of mathematical physics in spherical geometries that are of interest in binaural systems. This chapter presents an overview of the theory of spherical acoustics from three equivalent approaches. The first approach emphasizes the physical nature of this theory and lies on the solutions to the acoustic wave equation in spherical coordinates. The second approach highlights the mathematical aspect of the theory, which arises from the fact that the angular portion of the solutions to the wave equation defines an orthonormal basis of functions on the sphere that is used to represent spatial sound information in general. Finally, the third approach brings up a general scheme for processing of spatial information in the represented or encoded domain, which constitutes a convenient and intuitive framework for the design and implementation of binaural systems in spherical geometries.

Chapter 3 presents a model for the synthesis of the spatial sound information that is contained in a external model of spatial hearing. This model is called the head-related transfer function. Synthesis is considered at far and near distances, and for spherical and circular geometries. Binaural presentation would ideally require knowing the head-related transfer functions (HRTFs) for all points in three-dimensional space. However, measuring or calculating these HRTF datasets is a complex and time consuming task. This chapter presents models for the synthesis of HRTFs for arbitrary sound source positions. The input to these models are the values of the HRTFs on a boundary at a given initial distance. These models constitute a useful result, since spherical datasets are typically obtained for only one far distance. Moreover, prior knowledge about the HRTFs at the target positions is not required when applying these models. An existing model for synthesis on the three-dimensional space from values on a spherical boundary is first reviewed. For practical purposes, two existing models for synthesis on the horizontal plane from values on a circular boundary are also reviewed. The first contribution of this chapter is an additional horizontal-plane model that outperforms the existing ones. The second contribution is a general method for limiting the angular bandwidth of the continuous models, which can be applied to all existing models, disregarding whether they are based on spherical or circular boundaries.

Chapter 4 presents a model for the capture of spatial sound from far and near distances by using spherical and circular microphone arrays. This chapter addresses the problem of estimating the sound pressure signals outside a spherical microphone array mounted on the surface of an acoustically rigid spherical baffle. The pressure signals outside the array are estimated from the microphone array signals. For this purpose, a continuous-space model is first formulated for estimating the sound pressure field outside a rigid sphere, once the measurements of sound pressure on its whole surface are available. The presence of the spherical baffle constitutes an acoustically rigid boundary; it provides higher stability to the estimation as opposed to the use of open spherical arrays that does not use a baffle. The continuous-space model is used to derive a discrete-space model by angular-band limitation and sampling. The resulting discrete-space model constitutes a theoretical basis for estimating the sound field outside the spherical microphone array.

Chapter 5 presents a general model for binaural synthesis at far and near distances. This model allows to edit and combine distinct types of spatial sound information defined on the sphere. For instance, microphone array recordings and datasets of head-related transfer functions. The model therefore relies on the models formulated in previous chapters. More specifically, the model combines distance-varying filters and boundary-matching filters.

論文審査結果の要旨

3 次元音空間情報収音・再生技術は高臨場感情報通信のための基盤技術であり，その中でも，音源から耳までの音の伝搬特性を表す頭部伝達関数に基づくバイノーラル音空間合成技術は個々人に対して精度の高い音空間を簡便なシステムで提示可能という大きな利点を有している。特に，球座標上の調和関数による球面調和解析に基づき，球状マイクロホンアレイを用いて収録した音空間情報を球面調和解析した後に聴取者個々人の頭部伝達関数に合わせて合成する手法は，全方位にわたって高い空間解像度が実現でき，かつ，様々な再生系にも適用可能な手法として有効性が高い。しかし，球面調和解析が方向にのみ着目した分析法であることから，従来手法では頭部伝達関数の音源距離依存性を反映させた音空間合成が困難であった。さらに，空間上で離散的に得られる音空間情報に連続的な球面調和解析を適用することで生じる歪が合成音空間に悪影響を及ぼしていた。著者は，これらの問題点に対し，球面調和解析の有効性に着目しつつ，数学的，物理的な解析の結果に基づいた定式化により解決するための研究を進めてきた。本論文は，その成果をまとめたもので，全編 6 章からなる。

第 1 章は序論である。

第 2 章では，球面調和解析に基づく音空間解析法について概説するとともに，その問題点について述べている。

第 3 章では，任意の距離での頭部伝達関数の合成法に取り組み，波動方程式の距離に関する項が Hankel 関数で表現できることに着目し，特に，人間の音空間知覚精度の高い水平面の頭部伝達関数を対象に，頭部伝達関数の角度依存成分と距離依存成分を分離して分析することで，全方向・距離にわたって頭部伝達関数を合成する新しい手法の理論的検討，および，定式化を行っている。これは独創的，かつ，実用性に優れた成果と高く評価できる。

第 4 章では，Kirchhoff-Helmholtz 積分方程式を用いた従来手法において音空間合成精度に影響を及ぼしていた，収音時と再生時における境界面の不整合を解決するため，両者の取り得る状態を体系的に分類した結果に基づいて，各状態をマッチングさせる新しいフィルタ設計法を提案している。これは，球面調和解析を用いた全ての音空間合成法に適用可能で，有効性も高い優れた提案と評価できる。

第 5 章では，第 3 章，第 4 章で提案した手法に基づき，収録した音空間に含まれる音源の距離情報を方向ごとに任意に操作可能な新しい音空間合成法を与えている。これは頭部伝達関数を利用した様々なバイノーラル音空間合成手法にも適用可能な優れた提案として評価できる。

第 6 章は結論である。

以上要するに本論文は，高い臨場感を持つ音空間情報を高精度に収音し，聴取者の両耳へ全方向・距離にわたって高精度に再生する新しい音信号処理方式を定式化したもので，システム情報科学ならびに音情報科学の発展に寄与するところが少なくない。

よって，本論文は，博士（情報科学）の学位論文として合格と認める。