Spatial and 3-D Audio Systems

Hüseyin Hacıhabiboğlu
Graduate School of Informatics, Middle East Technical University (METU), Turkey

INTRODUCTION

The advent of 3DTV, computer games, virtual and augmented reality applications made it necessary to design audio systems that are able to provide better quality audio with more spatial realism and envelopment. A variety of such systems exist that differ in terms of the principles that they operate. This article will present an overview of the existing spatial and 3-D audio systems and discuss their relative advantages and disadvantages within the perspective of possible usage scenarios.

BACKGROUND

The earliest system that was capable of recoding and reproducing sounds was the *paleophone*, developed by the French inventor Charles Cros. More well known is Edison's *phonograph*. Both devices allowed recording and reproducing sound by physical means (i.e. by recording a physical trace of the sound on a soft material and then using a stylus coupled to a diaphragm and a horn to obtain back the recorded sound). The advent of electroacoustic transducers made it possible to amplify recorded signals by electrical means. This made monophonic systems widely available by the early 20th century. Regardless of their reproduction quality, these systems lacked the spatial context, which every recorded soundscape will embody. Even if a symphonic orchestra is recorded and that recording is played back using a monophonic system, the perceived sound scene will be constrained to the aperture of a single loudspeaker.

Blumlein is credited with the invention of stereophonic sound in 1931 making it possible to record and partially reproduce the spatial context of a sound scene over a pair of loudspeakers. Innovative stereophonic recording and reproduction techniques have since been proposed and stereophony is still one of the most popular methods to render recorded or synthesized sound scenes with their spatial contexts.

Steinberg asserted in 1934 that the accurate reproduction of the auditory perspective requires at least three independently recorded and reproduced audio channels (Steinberg, 1934). While the early subjective experiments also support this assertion, it was not feasible to record, distribute and reproduce three-channels of audio using the technology at the time.

The basic mechanisms of human sound source localization and spatial perception were known as early as the late 19th century (Rayleigh, 1877). The renewed interest in spatial hearing in 1950s made it possible to design recording and reproduction systems that allow a listener to experience three-dimensional sound scenes over a pair of headphones. This technology is now known as *binaural audio*. Binaural audio has some stringent requirements on the processing and reproduction chain to deliver full 3-D audio.

Gerzon's work on spatial harmonic decomposition of a sound field and a special microphone to achieve such a decomposition resulted in Ambisonics (Gerzon, 1973). First-order Ambisonics typically uses 4-8 loudspeakers to deliver 3-D audio to a single listener positioned at the middle of the loudspeaker rig. The formulation of Ambisonics allows higher-order expansions of the sound field, which are used in higher-order Ambisonics (HOA).

Berkhout applied the principles of acoustic holography for the purpose of acoustic control (Berkhout, 1988). It was realized that the fundamental idea could also be applied to the physically accurate reconstruction of sound fields. Practical systems have since been developed and that technology is now known as wave-field synthesis (WFS).

The following section explains these systems in more detail, discusses the existing problems and presents the state-of-the-art.

DOI: 10.4018/978-1-4666-5888-2.ch594

SPATIAL AND 3-D AUDIO SYSTEMS

Stereophony

A typical stereophonic system requires only two loudspeakers separated by a base angle of 60° as observed at the listener position (e.g. loudspeakers are positioned ±30° to the left and right of the listener). In order to experience the spatial context accurately, the listener has to be seated at the *sweet spot* positioned on the midline separating the two loudspeakers. Due to its minimal hardware requirements it has been used in a variety of systems from radio and TV broadcasts to computer games. Stereophonic sound can be obtained either artificially or by recording sound scenes.

Artificial Stereophony

Artificial stereophony is based on the fact that the direction of the local sound field around a listener's head can be manipulated by adjusting the level and/or time differences between the left and the right channels.

Panning by adjusting the level differences is known as intensity stereophony. Two *panning laws* that relate the loudspeaker gains to the panning direction exist: the *sine panning law* and the *tangent panning law*. These laws relate the gains of the left and right loudspeakers, g_L and g_R, to the panning direction, θ, and the stereophonic base angle, θ_0, as shown in Figure 1. These laws can be used to calculate the gains of each loudspeaker, based on a simple trigonometric expression. This way, virtual sources can be positioned between the left and the right loudspeakers. The tangent panning law is known to be more robust to head rotation (Bernfeld, 1973).

There also exist panning laws that relate the combination of the level and the time differences to panning direction. This is known as *time-intensity stereophony*. While it can provide a slightly larger sweet spot, the generated stereo image is known to be less stable than intensity panning, a phenomenon known as *phasiness* (Lipschitz, 1986).

True Stereophony

Stereophonic sound can also be recorded with microphones using special recording techniques. These can be grouped into three: *coincident*, *near-coincident*, and *non-coincident stereo techniques*.

Coincident techniques use two co-located microphones that capture the level differences. Well-known examples are the X/Y and mid/side (MS) techniques. One flavor of X/Y technique uses two cardioid microphones facing 60° to the right and 60° to the left. Another flavor, known as Blumlein pair, uses two bidirectional microphones facing 45° to the right and 45° to the left, respectively. MS technique uses a cardioid microphone facing the front direction and a bidirectional microphone facing sideways. The left and the right channels can be obtained by a weighted addition of the signals recorded by these individual microphones.

Near-coincident techniques use closely positioned directional microphones. The most frequently used techniques are the NOS and the ORTF techniques. In both cases, two microphones with cardioid directivity patterns are used. In the NOS technique, the separation is 30 cm and the microphones face -45° and 45° with respect to the front direction. In the ORTF technique, the separation is 17 cm and the microphones face -55° and 55° with respect to the front direction. Near coincident arrays use both the level and the time differences between channels.

Non-coincident techniques use microphones separated by more than about 1 m. A well-known example is AB stereo, which uses two omnidirectional microphones. This way, only the time differences are

Figure 1. Artificial stereophony by intensity panning