SURROUND SOUND IN RADIO DRAMA

EERO ARO

Producer, Sound Designer
Finnish Broadcasting Company, Radio Theatre
Finland

Radio drama can benefit from spatial sound as it releases the listener from the position of a spectator. General production tools for multichannel audio can be used in radio drama, but the setting between the performance and the listener has very little in common with multichannel music recordings. In a radio play the point of listening can change with time. Auditory objects can move and the size and the character of the room may change, spaces may appear and disappear, and there can even be several layered rooms or spaces at the same time. All of this gets its final form in the mix.

SURROUND SOUND IN RADIO DRAMA

The picture of a radio play is larger than any cinema screen. It has no frames or edges, as it builds up in the imagination of the listener.

The majority of radio dramas are made in two channel stereo. Stereo can express the directions of sound sources and the depth and the width of the sound stage *in front of the listener*.

However, many radio drama people feel that two channel stereo is rather limited, as it is just mimicking the setting of a traditional stage theatre. In the classic setting of the theatre the performers are placed on a stage in front of the audience, who are sitting in their seats opposite to the stage. This setting allows the audience to follow the play without being taken *into* the play. It gives them a safe feeling, as they can *choose* whether they are taking part in the play or not. They *can* be involved in the play if they wish to.

In a well-made theatre play the spectator can identify him- or herself with one of the characters, usually the protagonist. As long as this character is visible on the stage, there is a connection between the spectator and what is happening on the stage. The spectator feels that he or she is *part of the play*. This connection is very fragile and is easily broken. Any distracting sound, light or off-stage movement can shatter this illusion.

However, radio drama is *not* stage theatre. A radio drama listener is required to make an active commitment, an "aural contract" with the play. [1] The goal of the radio play director is to make the listener stop *listening* and make him *hear*. In other words, to set him free from the position of a spectator. This is where spatial sound reproduction can come to help, as it can bring the listener inside the story by immersing him with sound. Spatial sound gives more room for the placement of the sounds and for the reproduction of the spatial characteristics than two channel stereo.

Of course there is a paradox in here; if a radio play takes place in the imagination of the listener, does the reproduction medium really matter at all? Wouldn't mono or stereo do as well, or headphones or just a pair of loudspeakers?

There is at least one good reason to use spatial audio for radio drama; multichannel listening causes less listening fatigue. This is because surround sound is closer to natural hearing than two channel stereo. With stereo speakers, the sound image (with all of the reverberation included) is located in front of the listener. With the headphones the sound image is localized inside the listeners head, not to mention the uncomfortable feeling of using headphones during a radio play with a long duration.

Clarity

If delivering *information* is the important thing, it is practical to use mono sound. However, if several people are talking in the studio at the same time, two channel stereo and surround sound help the listener to detect individual sound sources from the crowd, which in mono reproduction would be masked by louder sounds.

The illusion of space

Surround sound reproduction creates an illusion of an auditory space. The cues are given both by the ambient sounds themselves and in the placement of the sounds, but also in the spatial characteristics, such as the reverberation and the reflections of the room.

Directing listener's attention

The listener's attention can be guided to a certain direction, in surround sound this can also be to the side or behind the listener.

EXPANDING FROM TWO CHANNEL STEREO

Binaural techniques

Binaural techniques were used for radio drama in the 1970s. The BBC made a number of productions and dummy-heads were especially popular in Germany.

The good thing about binaural audio is that it delivers a three dimensional sound image with two channels only and the sound image can be very realistic. A downside is that the listener must use headphones. Traditional dummy-head techniques also have the problem that most listeners don't locate frontal sounds in front of them, but behind the head. There are solutions for these problems, but HRTF-processing or convolution panning are not yet everyday production techniques in radio drama, not to mention Ambiophonics or head-tracking in the listener's home. Radio also still needs to take care of the compatibility between mono, stereo and spatial reproduction. With binaural techniques the compatibility isn't very good. Binaural stereo tends to have a hole in the middle in loudspeaker listening.

If we look at binaural recording from the actor's point of view, there are certain problems. For an actor, a dummy-head represents a "one point microphone". Talking to a dummy-head needs extra training and experience as compared to mono and stereo microphones that the actors are used to. The actors need to place themselves and make their movements in relationship to the dummy head. They constantly need to keep in mind, where the "third listening person" is. The important contact between the actors is easily interfered with if they need to pay too much attention to the microphones. Best results are gained when the actors work in contact with each other, and it is the sound engineer who takes care of microphone placement and the recording.

The actors need to pay extra attention when they move on the left and the right side of the dummy-head. The sound image must not jump suddenly from the front to the back or vice versa, and that can happen very easily. Moreover, the speed of the movement must be different depending on the distance to the head.

Using non-dummy-head sound effects with the dummy head dialog recordings would require additional processing at the mix.

Matrixed surround systems

Matrixed systems were made for the cinema, in which dialog must be localized to the picture. To achieve this, the systems used bandwidth limiting of the (single) rear channel, time delay to the rear speakers and adaptive control of the soundfield. Many of the adjustments were left for the home listener.

Matrixed systems were not very ideal for radio drama, because it wasn't possible to place direct sounds at the rear. There was only one rear channel, so that sounds couldn't be placed at any particular direction at the back. Placing a voice in the rear channel sounded bad, as the basic frequencies and the sibilances of human speech were filtered out. Localization of a voice in the rear channel was undefined because of the rear channel delay. It wasn't possible to compensate for the rear channel delay either, because the home listener may have adjusted the delay in his decoder by as much as 50 milliseconds.

Advanced matrix systems, such as Dolby Pro Logic II, solved most of the problems, but unfortunately PLII came to the market a little bit too late. Discrete digital multichannel systems were already there, and they are much better suited for audio only applications.

In the 1980s and 90s there was an alternative for transmitting surround sound via FM radio, and that was UHJ Ambisonics. UHJ didn't suffer from the technical disadvantages of the original Dolby Surround, but unfortunately Ambisonic decoders were not commonly available.

Discrete channel systems and 5.1

Discrete 5.1 systems were quickly adopted by radio drama in many countries as soon as they became available. All five channels are full range and discrete. If necessary, the LFE channel can be used for special effects.

There are some problems with the ITU-R BS.775 speaker layout, as the angles between the speakers don't support stable phantom images anywhere else than in front of the listener. The layout creates a nice feeling of space, but not so good localization of virtual sound sources on the sides or at the back.

THE AUDITORY PERSPECTIVE

Realistic perspective (in the audience)

The realistic sound perspective directly compares with the real world. A realistic perspective usually has following properties:

- audience's point of listening (POL) [2]
- stationary auditory objects
- clear picture of the space in which the performance is taking place
- clear picture of where the boundaries of the room are
- the space does not change

A realistic perspective is typically used for classical music, radio and tv documentaries and news, as well as sports programmes.

Imaginary perspective (in the band)

An imaginary perspective does not compare with the real world, or it is scattered or broken into pieces in place or time. An imaginary perspective has following properties:

- the point of listening can be "in the band", but it can also change with time
- the point of listening at times may be "in the audience"
- the point of listening can move
- the size and the character of the room may change
- there can be several rooms or spaces at the same time
- different sounds or instruments can have individual spaces of different sizes and characters, which are layered with each other
- spaces can appear and disappear
- spaces can change with time, or dynamically as a function of the auditory objects
- virtual sound sources can move

What is described above is not big news to those, who have been listening to pop and rock music all of their lives. The imaginary sound perspective was born with the electronic instruments. An electric guitar or a synthesizer has no direction or space of its own, all that is artificially made. In pop music the different parts of a drum kit are very often mixed into different spaces. In electroacoustic music space has always been a compositional element.

Microperspective

Radio drama can sometimes use very strange and unrealistic spaces and even microperspectives. If tiny microphones are put inside small machines, or example a mechanical alarm clock, the dimensions of the original sound source are magnified when it is played back through loudspeakers. Unrealistic spaces can also be created with convolution filters. Small spaces can be modelled and used in the mix. Sometimes the point of listening can be inside a small space, such as a cupboard.

RECORDING

Dialogue

Radio drama dialogue is usually recorded with mono or stereo microphones. Special surround microphones or multichannel microphone setups aren't too practical for radio drama, except in some special cases.

One of the things to avoid is hiss and noise. Actors sometimes use a very soft voice, less than a whisper. To keep the studio air conditioning noise as quiet as possible, directional microphones must be used at close distance. When the actor is speaking into the microphone from one direction, a multichannel surround microphone isn't a very good solution, as it also captures noise from all directions. The noise will cause problems at mixing.

Coincident surround microphones, such as the Soundfield, have a one point character similar to the dummy head, which was discussed above. Coincident surround microphones can be a good solution in plays that take place in just a couple of rooms and that have a rather stationary setting. If this type of a microphone is used for dialog, then all placing and the movements of the actors must be planned very carefully in beforehand. A ready-made surround recording doesn't leave much room for changes at the mix and must be used as it is. On the other hand, coincident surround microphones are handy for on-location recording, where noise is a part of the ambience.

Spaced multichannel microphone arrays are problematic with actors in the studio. There should be a certain distance between the actors and the microphones to avoid comb filtering effects. Radio drama studios are usually too small in size for spaced arrays. Comb filtering will happen very easily, if the actor is moving close to a spaced array. Spaced arrays could possibly be used for recording a theatre play on stage, or when recording on location.

Sound effects and music

Spot effects and Foley sounds are best recorded in mono or two channel stereo, for the same reasons as dialogue recording. Spot effects should be kept as clean as possible from any other noises, to allow easy adjustments when mixing them into the final sound image.

Special surround recording techniques work well with ambient sounds such as city noise, traffic, wind, rain, thunder etc. The best option is to record surround sound effects separately from the dialogue and mix them later together. Music of course benefits from surround recording, depending on how the music will be used in the play. Having said that, "music for radio drama" is not the same as "listening music". Radio drama music is "just" material for the radio play.

MIXING

Creating surround sound for radio drama takes place at the mix. Using mono and stereo tracks for dialogue simplifies the mixing procedure a lot, because otherwise the number of tracks will grow very high. Several types of track combinations can end up to the same mix; mono, stereo and multitracked takes, B-format etc. Handling a large number of multichannel files gets complicated and assigning multichannel effects to the tracks needs concentration.

Some people think that a recording can use only one sound reproduction system at a time, either mono, stereo or multichannel. Of course this is not true, you can:

- use direct mono feeds to single loudspeakers (real sound sources)
- pan sounds between a pair of speakers (virtual sound sources)
- play back a two channel stereo recording from two adjacent speakers (virtual sound sources)
- play an acoustic 5.1 recording through all loudspeakers
- play an acoustic B-format recording through all loudspeakers

Very effective openings of sound atmospheres can be made by using first only one loudspeaker, then mixing into two channels and finally expanding into a surround image. A corresponding close-down is simply made the other way around. The mono point source doesn't need to be in front of the listener.

Figure 1 An expanding opening from mono through stereo to full surround

Hearing cannot really "know" which cues or sounds belong together. It just receives different sound cues from different directions and the total image builds up in our imagination.

COMPOSITION OF THE SOUND IMAGE

Simplicity

Imagination needs only a couple of sound cues to create an auditory illusion. A too dense sound scene clogs up the listener's ability to receive any more cues.

Using only essential sounds

Auditory perception can follow only a limited amount of cues at the same time. The number varies between three and seven cues depending on the case. If there are more sounds, hearing starts to group similar cues together, or starts to ignore cues that are considered less important. Selection of the next cue can happen instantly and the sounds that were considered important just a moment ago, are immediately forgotten.

Direction of the auditory objects

Auditory objects with important information should be placed in centre front. We turn our head automatically towards an important sound, if it first appears in some other direction than in the front. Thus, if the goal is to deliver information, it is better to put it in the centre front right away. In radio drama centre front is a natural direction for the voice-over and a narrator.

Sounds appearing from outside the field of vision represent something unknown and wake the listener's attention. Some of the sounds can mean a threat or a danger (the tiger), but some others mean surviving (game to be hunted). A voice appearing suddenly from behind may surprise the listener, but a cosy room ambience will create a safe feeling.

Pairwise intensity panning between two speakers creates rather unstable phantom images especially to the left and right and behind the listener. Ambisonic B-format panning does a slightly better job especially to the sides of the listener.

Directional balance

Two auditory objects in a soundfield tend to create a tension between them. The listener gets a feeling that a loud auditory object is pulling a softer sound towards itself. Several small auditory objects can add together and balance with one loud auditory object.

Care should be taken for the long term balance of the auditory objects both in the left-right dimension and in the front-to-back dimension. If "heavy" or "remarkable" (compressed or loud, especially at low frequencies) sounds remain for a long time on either side or behind the listener, the sound balance starts to have too much weight on just one side of the soundfield. Auditory objects should be placed in turns on both sides in such a way that the overall feeling is balanced.

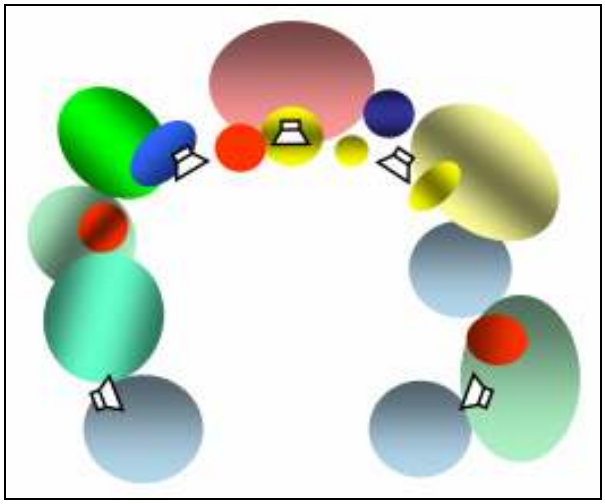

Figure 2 A well-balanced surround mix

Volume

When the sound needs to be really loud and/or the sound image should create an engulfing omnipresence, all six channels are available. Five speakers make much louder noise than two. This is especially effective with explosions and crashes etc. LFE can also be used.

A void space

An anechoic room is "acoustically empty". Hearing has no means to judge the size or the shape of a room if there isn't any sound, as there are no acoustic reflections either. In cinemas or radio plays an anechoic space symbolizes the time before we were born or the time after our death. A totally quiet place is frightening for many people. A void space isn't much used in cinema or radio plays, but digital audio certainly gives a possibility to do that.

Silence

In reality, there is always some quiet sound even in silent places. Where there is air, there is sound. In buildings there is some noise from the heating, air conditioning, water pipes or the draining. In towns or cities some of the traffic noise comes through the walls. The structures of the building make some noise when warming up or cooling. Recording the silence effect in surround asks for really high-quality equipment.

Depth

Auditory objects in the front are localized at different depths in the sound image, thanks to the reflections played back from the rear speakers.

To some extent it is also possible to place auditory objects *inside* the circle of the speakers. A phantom image that is panned between the left and right front speakers, is localized closer to the listener than the hard image from a centre speaker only. Otherwise there cannot be stable phantom images inside the array. Only an illusion of a fast movement through the middle area works well.

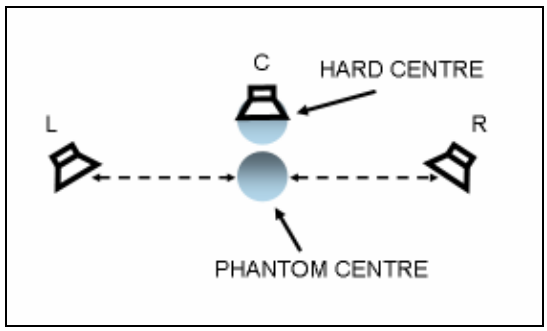

Figure 3 A phantom centre image is localized in front of the centre speaker

A divergence or focus control also brings an auditory object slightly closer to the listener, although the same controls change the width of the auditory object at the same time. Ambisonic tools have a diameter adjustment, which adjusts the distance of the auditory object from the listener.

Creating the illusion of space

The scenes of a radio play take place in various interiors and exteriors. Unfortunately the acoustics of radio drama studios cannot be adjusted or changed very much with screening or curtains. This is why scenes are often recorded in acoustically-damped studios and the feeling of space is created artificially.

The feeling of space is created by artificial or sampled reverberation, or by recording in the original acoustics.

The change of the size of the space works as a transitional element between the scenes, for example mono narration in one speaker only changing over to a following scene which is in full surround. Temporal transitions can also be realized with spatial changes.

A moving auditory object

A mono sound is moved by panning, or by making a recording in a real environment. The auditory object for this should be a figure type of sound. Without a moving picture it is difficult to understand that a background type of sound would be moving. Ideally the tone of the sound should change with the distance, and there should be a doppler effect also, and we should hear the reflections of the auditory object from the surrounding surfaces. Plain intensity panning cannot do this; there should also be EQ, pitch shifting, a bunch of delays, reverberation and dynamic panning of all of these effects!

Large moving auditory object

A classic example of the movement of a large auditory object is the giant spaceship in the beginning of the "Star Wars" movie gliding from behind the audience towards the front. A large auditory object moves like a "cloud" or a "blanket" that is pulled over the listener. It is easier to perceive the auditory object moving in front-to-back direction than sideways without the picture.

Subjective movement (we go with)

The subjective movement assumes that the point of listening is moving and the scenery around is staying in place. The easiest case is a movement towards the front, in which the auditory objects keep moving by the listener, from the front to the back on both sides. A backwards or sideways movement is difficult to understand with sound only. A sudden change of the direction will make the listener confused. A turn in the direction of a movement breaks the continuity that the listener is expecting.

The listener can only perceive a limited number of moving auditory objects at the same time. Also, hearing doesn't integrate moving points as lines in the same way as seeing does. Acceleration of a spaceship to the speed of light in a science fiction movie is usually represented by bright lines filling up the space. If each of these moving light spots had a corresponding sound, all we would hear would be white noise. For sound, it is better to use a small number of moving auditory objects at a rate of - say - half a second.

One of the most effective and interesting subjective movements in surround sound is moving from an acoustic environment to another, for example from a narrow corridor to a hall with a lot of people. The illusion doesn't work too well in two channel stereo, as you need the front-to-back dimension for it. This effect is best achieved by recording it acoustically.

Direction of movement

Our relationship to the left-right direction is cultural and is based in the direction of writing. In cultures, where people write from left to right, the same orientation means a change from "darkness to the light", from the "unknown to the home" or "to a commitment" or "from below upwards". A leftward movement leads to a "distance", "away from familiar places", towards an "adventure" or "against the stream". In stage theatre or cinema a bad person appears to the stage from the right and a likeable person from the left.

A very slow speed of the movement creates just a vague change in the sound. A fast movement is detected better. The easier it is to perceive an auditory objective as a figure, the slower the movement can be. A change in the speed of the movement is hard to understand in audio only applications. A change of the pitch gives the listener a better illusion of speeding up or slowing down.

Rotation

Rotating or tilting the complete soundfield can be used for correcting or changing the orientation of an acoustic B-format recording. For example a recording made in a shopping mall may have dominant sounds in same directions with the dialogue that was made in a studio. The interfering sounds can be turned into another direction by rotating the ambience recording at postproduction. Ambisonic B-format rotation is the easiest way to do this. With LCRLsRs-channels all of the channels would need to be panned separately and in sync with each other.

Figure 4 Rotate VST-plugin [3]

Rotation can also be used as an effect. When the speed of the rotation is fast, the sound image starts to lose its directional properties and the sound changes into a new sound. This happens at rates over 16-20 rpm.

A third application is to use rotation as a transition element between two scenes. This can be directly compared to a wipe transition in film cutting or from the rotating scene of stage theatre. A stage setting is turned to the side on a turntable and the next scene then fills its place. A corresponding wipe effect in radio drama can be made by rotating the soundfield.

Tools

A couple of digital audio workstations have integrated tools for some of the abovementioned spatial effects and there are a lot of plug-in programs available. There are also some multichannel EQs, reverbs and dynamic processors both in hardware and software. A large number of free multichannel VST-plugins exist but unfortunately there are just a handful of multichannel VST-Hosts. There is a need for integrated DAWs that would have the proper recording facilities, a waveform editor, multichannel mixing, multichannel plugin support and CD and DVD mastering. In professional production there is no time to jump back and forth between different pieces of software to do different processing jobs. Everything must take place within just one software application.

Sound effects

Much of the sound effects come from effect libraries, which are usually stereo, although there are multichannel sound libraries available. Basicly the surround sound image in a radio drama is built from mono and stereo elements. Stereo elements can be upmixed into four or five channels.

DELIVERING SURROUND SOUND TO THE HOME

An efficient way to broadcast surround sound to the home today is through satellite. The satellites have enough bandwidth for high data rates.

In terrestrial DVB the broadcasters want to use the bandwidth for as many programme services as possible and for other additional services. What is left of the bandwidth after that, can only deliver a stereo signal. A re-division of the DVB-T-multiplex would be needed for transmitting multichannel audio.

The bandwidth problem applies to Digital Radio too. Unfortunately the DAB was born at a time when compression techniques were at a high development boom. After a long consideration the broadcasting companies finally decided to use a compression technique that has subsequently been several times surpassed. A common bitrate for stereo programmes in DAB is today 192 or 224 kbit/s, when a decent 5.1 Dolby Digital signal would need 384 kbit/s. Joint stereo mode is also used, which effectively wipes out the Dolby Surround encoding from a stereo signal. At the moment, the best quality matrixed surround signal in broadcasting is delivered through FM stereo with Dolby Pro Logic II.

There are several multichannel formats for Internet distribution, for example Windows Audio Media and MP3 Surround. Both streaming and downloadable files are possible. Some companies have been experimenting with downloadable DTS.wav-files that can be burned onto a CD and played back through a home cinema system. The sound quality of all of these formats is very good and competes well with traditional FM-broadcasting.

DVD-V records are possibly the best way to deliver a high quality surround sound to the home.

---

[1] Alan Beck: Listening to radio plays: Fictional soundscapes.
http://interact.uoregon.edu/medialit/wfae/readings/beck
[2] Michel Chion: Audio-vision: Sound on Screen
Columbia University Press, New York 1990
[3] Dave Malham:
http://www.york.ac.uk/inst/mustech/3d_audio/vst/bproc_help.html