

INVITED REVIEW

Single box surround sound

Tony Hooley

1 Ltd, St. John's Innovation Centre, Cowley Road, Cambridge, CB4 0WS England

Abstract: In this tutorial paper, the author introduces a full physical surround-sound system in a single equipment box, focusing on its background, novel technology, and its application. Working on the principles of phased-array antennas commonly used for electromagnetic waves, but adapted for the wide-bandwidth requirements of audio acoustics, Digital Delay Arrays (DDA) simultaneously produce multiple independently-steered and -focused beams of sound each potentially carrying different audio programme material. Utilising the available reflective surfaces (e.g. ceiling and walls) in nearly all domestic listening environments, these distinct beams may each be arranged to reach the listeners from different directions, thus producing surround-sound. The basic signal processing requirements as well as several refinements are described, along with a discussion of the major design parameters of practical uniform array antennas, with extensions to non-uniform and non-planar array structures.

Keywords: Beam-steering, Array-antenna, Surround-sound, Digital Delay Array

PACS number: 43.60.Fg [doi:10.1250/ast.27.354]

1. FOREWORD

In this tutorial paper, the author introduces a full physical surround-sound system in a single equipment box, focusing on its background, novel technology, and its application. Working on the principles of phased-array antennas commonly used for electromagnetic waves, but adapted for the wide-bandwidth requirements of audio acoustics, Digital Delay Arrays (DDA) simultaneously produce multiple independently-steered and -focused beams of sound each potentially carrying different audio programme material. Utilising the available reflective surfaces (e.g. ceiling and walls) in nearly all domestic listening environments, these distinct beams may each be arranged to reach the listeners from different directions.

2. INTRODUCTION

A statistic from several major manufacturers of (low-end) multi-box (i.e. 5+) surround systems is that >60% of all buyers never connect the rear speakers, either because they are unable practically to site them in their listening rooms (lack of space, aesthetics), or because they can find no acceptable route for the cabling, or both. A Digital Delay Array (**DDA**) eliminates these problems and can work straight out of the box, sitting on top of, beneath or within the users' video display screen. Recent developments enable the entire DDA (array and electronics) to be built into and fully integrated with the display screen itself, so providing a **zero-box** surround-sound solution with no cabling whatsoever.

A DDA is a wide-band acoustic "phased-array" antenna and electronic driver system, capable of producing one or more steerable and focusable beams of sound in the audible range. Because of the wide-bandwidth requirements of high fidelity audio signals (i.e. 8 or more octaves), phase shifters are impractical and digital delay lines are used instead, so the term *phased-array antenna* is better replaced by digital-delay-array antenna.

One excellent practical implementation (see e.g. Fig. 1) comprises a 2D planar triangular array of small (say 15–30 mm diam) wideband acoustic output transducers on a 40 mm pitch uniformly covering an area of, for example, 800 mm × 500 mm, with over 200 transducers in total, each driven by a dedicated power amplifier and fed from a dedicated synthesised signal channel. Suitable transducers might usefully cover the range from 50 Hz or below right up to 20 kHz. A signal processing system (nowadays digital by default) takes one or more audio input signals and for each input signal generates a unique version of that input for each transducer, delayed and amplitude attenuated by a specific amount which is a function of the position of that transducer in the array.

Where there is more than one input signal, these transducer-unique versions, one for each input signal, are summed to produce the nett drive signal for that transducer and applied to the transducer's dedicated power amplifier (see Fig. 2).

By suitable choice of the aforementioned delays and gains, the transducer array can be made to synthesise directed **beams** of sound (one for each input signal), these

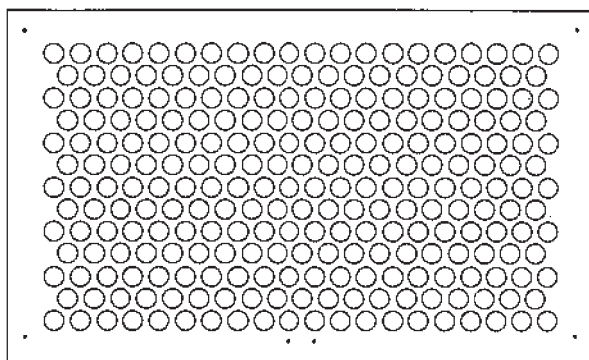


Fig. 1 A basic DDA array with 254 transducers on a uniform triangular grid.

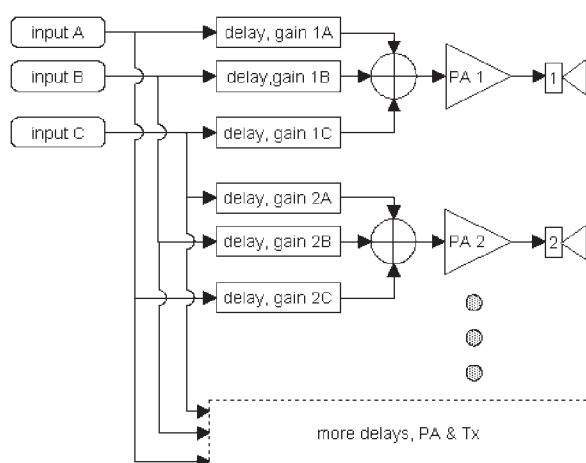


Fig. 2 Each input signal (here A, B, C) is uniquely delayed and gain-adjusted before being summed with the others for each PA/transducer in the array.

beams being independently steerable in $\sim 2\pi$ steradians of space in front of the transducer array, and each focusable at any distance in front of or behind (i.e. virtual focal point) the array.

The beams of sound may be bounced zero or more times off walls and ceilings of a listening room so as to reach the listeners from in-front, the front-sides, or behind on both sides (see Fig. 3). With five such beams suitably steered and focused, a full surround-sound experience can be produced. The perception of the listeners is that the sound beams arrive from the directions of their final bounce point en route to the listeners, or where there is no bounce, then from the array itself, as indeed they do. Because of this physical fact, the surround-sound image stays locked to the room geometry no matter how the listeners move their heads, or indeed move around in the greater part of the listening room.

Phased-array antennas as such, are not new in audio, with a discussion by Jordan [1] appearing as long ago as 1971 in the UK, and a significant patent being filed by Yanagawa *et al.* [2] in 1992 in Japan. However, the

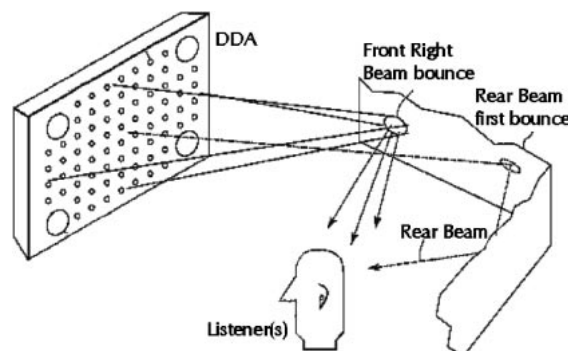


Fig. 3 Synthesised sound beams (Front Right, and Rear Right) bounced off one and two walls to a listener.

additional concepts (multiple simultaneous discrete beams/channels, and reflections off walls and ceilings) needed to produce a Digital Delay Array for surround sound reproduction came somewhat later, Hooley *et al.* [3]

3. PSYCHOACOUSTICS

To understand how surround-sound of any kind can work at all, it is necessary to look at the human sound direction perception mechanisms. For left-right direction discrimination the ear-brain system mainly uses two parameters depending on signal frequency: relative time-delay; and relative amplitude. For front-back disambiguation the relative *and* absolute frequency response is used. This latter effect is related to the Head-Related-Transfer-Function (HRTF).

Because DDA actually delivers most of the acoustic bandwidth from the appropriate directions, it does not need to modify the frequency response of signals (using HRTF “tricks”) to fool the brain. However, to adequately steer very low frequencies requires quite large arrays (see below) which are impractical in most domestic environments. Here DDA tacitly uses the brain’s own psychoacoustic tricks, by delivering the lowest frequencies *directly* to the listener without bounce-paths, but suitably *delayed* to arrive synchronously with any (steered and bounced) higher harmonics. Our experimental findings are that the brain just puts the various components back together again and determines the “true” direction, from the bounce point of the correlated *higher harmonics*. We hypothesise that this occurs because human direction perception is relatively poor at low frequencies in the absence of higher (correlated) harmonics.

Two other psychoacoustic effects indirectly of relevance are the *Haas Effect* and the *Precedence Effect*: Haas [4] described how the brain perceives later arriving copies of a sound, as less-loud (by up to 10 dB) than the first-arriving sound. Wallach [5] and others [6,7] described how the brain is able to *ignore* the direction of a later-arriving copy of a sound, even if up to 10 dB *louder* than the initial

sound, and even if from a different direction, so long as the delay was not too great ($< \sim 35$ ms). Evolutionarily speaking, these effects are a useful adaptation as they help the ear-brain system to de-clutter received sound signals and determine “true” origin in reverberant environments such as caves, chasms *etc.* in the presence of strong *later-arriving* sounds.

Unfortunately with DDA we generally require the brain to **attend** (not ignore) only to any *later* arriving copy of a sound (i.e. after it has travelled some way around the room via one or more bounce points) over time intervals of up to 12–30 ms or more; the Haas & Precedence Effects essentially work against DDA presentation of sound fields, assisting the brain to correctly recognise and identify the DDA itself as the true causal source of the sounds, and to ignore the bounced and thus delayed sound images from the walls and ceiling. However, we have determined that there is a second effect essentially the *inverse* of the Haas/Precedence Effects; *viz.*, if the *later* arriving (within ~ 10 –35 ms) sound copy is *louder* by 12 dB to 15 dB or more than the earlier arriving sound, then the brain completely *ignores* the *early* arriving sound, and hears, and *directionalises*, *only* the later copy. This is important for DDA because no practical array sound-beam has zero sidelobes, and such sidelobes can reach the listener directly, and well before the bounced beam. DDA surround-sound works precisely because of this psychoacoustic effect, as far as we know not previously described.

4. PRINCIPAL COMPONENTS OF A DIGITAL DELAY ARRAY

At its simplest, a DDA is comprised of:

- a number N of acoustic transducers (“Tx”) arrayed usually in a 1D or 2D planar arrangement, each with its own power amplifier (“PA”) at the input of which is a summing junction;
- and an adjustable signal delay for each transducer for each of n DDA input signals, the outputs of which are summed for each transducer at the PA inputs (Fig. 2).

Thus there are $n \times N$ required signal delays, feeding into N summing junctions, PAs and Tx.

The values of delay to be used for each input/Tx pair are simply calculated (see Fig. 4):

- a focal point location F_j ($1 \leq j \leq n$) in 3D-space relative to the array centre for the sound-beam corresponding to the j th input is defined (it may be at infinity, but its direction is then specified);
- the distance d_j^{\max} from F_j to the array Tx farthest from F_j is calculated;
- the distance d_{ij} from F_j to the i th Tx is calculated (for $1 \leq i \leq N$);
- a time delay $t_{ij} = (d_j^{\max} - d_{ij})/c_0$ is calculated, where c_0 is the sound velocity in air.

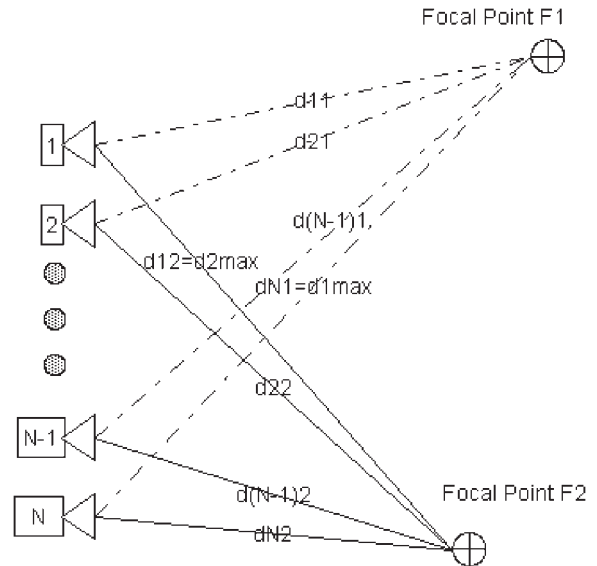


Fig. 4 Beam focal points and transducer path lengths.

The nett effect is that signals from each of the n inputs from all transducers arrive simultaneously at each of the n focal points. In practice more complex focal arrangements are possible with e.g. deliberate astigmatism, which can be useful in some circumstances.

Such a DDA would be capable of delivering multiple independent sound-beams to widely separated focal points. However, if these had been arranged such that the beams ultimately converged on a listener via zero or more bounces, the different input signals would not in general arrive synchronously because of the possibly widely different total path-lengths encountered. Thus the first refinement to the above scheme is to add *per-channel-delays* (n.b. distinct from the *per-Tx* delays required for focussing/beam-forming) with the shortest path delayed the longest etc. to provide channel synchronism.

A second refinement is to apply a unique gain-adjust for each and every one of the $n \times N$ delayed signals (see Fig. 2) prior to PA input summation, as a function of Tx position in the array. The primary purpose of this is to produce array **apodisation** (USA *apodization*), generally a tapering-off of aperture-illumination (Tx strength) towards the array edges. This has the effect of greatly reducing beam sidelobes, at the expense of some beam-width increase, and reduction in maximum power capability. A secondary purpose is to provide transducer calibration: — real mass-production transducers have significant gain (sensitivity) variation and a single calibration step during manufacture (preferably post array-assembly) can eliminate this to first order.

A third refinement is to provide *per-channel* equalisation (EQ). Because in general in a DDA surround-sound setup, each channel will arrive via a different physical path, possibly via widely differing reflectance characteristics on

each of several bounces, such an EQ process can rebalance the channels at the listeners' location. Such EQ filters can simultaneously correct for imperfections in the inherent Tx frequency response too. Typically these EQ filters are adjusted during a set-up procedure, which can be completely automatic.

As mentioned previously, practical domestic arrays are generally too small in physical extent to steer low frequency sounds with precision. So a fourth refinement is to separate out with filters the low frequency components below some frequency F_{\min} from all input channels, prior to the channel-EQ and beam-forming-delay stages. These n channel-LF components are then all added together, and delayed by an amount T_j related to the difference in distance to the listener via the *longest* reflected path D_{\max} as measured along the beam trajectory, and the *shortest* (direct, unreflected) path D_{\min} , so $T_j = (D_{\max} - D_{\min})/c_0$. This summed and delayed form of the n channel-LF components is then fed to the whole array in parallel without any appodisation, by feeding this signal in at each of the PA summing junctions. In this way, the LF components below some minimum steering frequency F_{\min} are dispatched directly to the listeners at such times as to arrive in sync with their related higher-frequency components.

In a similar manner, and not unlike a conventional home-theatre system, in a fifth refinement, the frequency components of all channels that are too low to be *reproducible* at all by the array transducers are filtered out and applied separately to a sub-woofer, again with suitable delays to achieve time sync at the listening position.

In practical DDA implementations it is often beneficial (e.g. for cost &/or power efficiency) to use digital PAs, with digital inputs. In this case it is then useful to introduce sample-rate-converters (SRC) near the n channel inputs to the DDA, if those inputs are also digital. The SRCs isolate the digital PAs from jitter effects from the digital inputs, and allow the PAs to operate at a fixed clock rate even when switching between different input digital sampling rates (e.g. 38 Kbps, 44 Kbps, 96 Kbps etc). It is similarly useful to use output-filterless digital PA technology that can drive the transducers directly, as the PA-Tx interconnections are short and entirely enclosed within the DDA casework.

In audio engineering it is usual to strive for *linear* transducers and PAs, primarily for fidelity of reproduction. In DDAs linearity takes on a new importance, because the DDA relies on the *Principle of Superposition* to enable the simultaneous creation of multiple different wavefronts from the DDA array, i.e. one for each beam or channel. Significant nonlinearity in this area of the DDA (i.e. after the summing junctions at the PA inputs) can introduce *inter-beam crosstalk*.

There are further refinements that can be added to a DDA processing system. These include: embedding the Dolby/DTS digital multi-channel decoder; adding adjustable lip-sync delays (neither of which are specifically DDA features); allowing two different sound channels to be beamed *directly* at and indeed focussed separately on each of two listeners, giving 15 dB or more separation in a normal domestic environment; adding an auto-setup routine, usually in conjunction with a microphone (plugged in as necessary) — there are several workable schemes whereby the acoustic properties of the listening room are probed by using the beam scanning features of the DDA to determine optimal bounce locations for the various beams, as well as channel EQ parameters — available devices can completely set themselves up with essentially no user input.

5. FUNDAMENTALS OF ARRAY STRUCTURES

We will now look at the factors that determine practically useful DDA array shapes. We will first consider 1D and 2D arrays (which in practice are essentially the same thing, as real transducers have significant 2D extent), and then discuss 3D arrays. Most of this discussion applies equally to *receiving* arrays (of microphones) as well as to our topic here, *transmitting* arrays. Unless otherwise stated we assume below that all arrays are mounted flush into a plane surface of greater extent than the array, with a closed back so that we may ignore rear-radiation from the transducers. For clarity, we ignore secondary effects due to the finite frontal area of this surface, and the finite volume of the space enclosing the transducers.

The single most important factor is the gross *array-size* L in any given direction, as this determines how *narrow* a beam may be formed *parallel* to that direction, as a function of frequency. In the limiting case, a point source ($L = 0$) is *omnidirectional* at all frequencies and there is no beaming at all, so no meaningful beamwidth. A useful rule of thumb is that an array of size L will start to have useful directionality at wavelengths less than L , or equivalently at frequencies greater than $F_L = c_0/L$. The exact details can be found in any good acoustics book and don't concern us here. Suffice to say, for wavelength w or frequency $F_w = c_0/w$, if we have $L \gg w$ then we will have good control over beam width and direction. So we need *big* arrays to steer low frequency sounds well.

The second key factor is the *spacing* s between individual array element centres. Because an array may be thought of as a discrete spatially sampled version of a continuous radiating surface, all the familiar rules about time sampling-rate, signal-bandwidth, Nyquist frequency and so on apply spatially: as a consequence, for a uniform array (i.e. constant s) then for output sound waves with wavelength $w_a < s$ the array will produce *alias beams*,

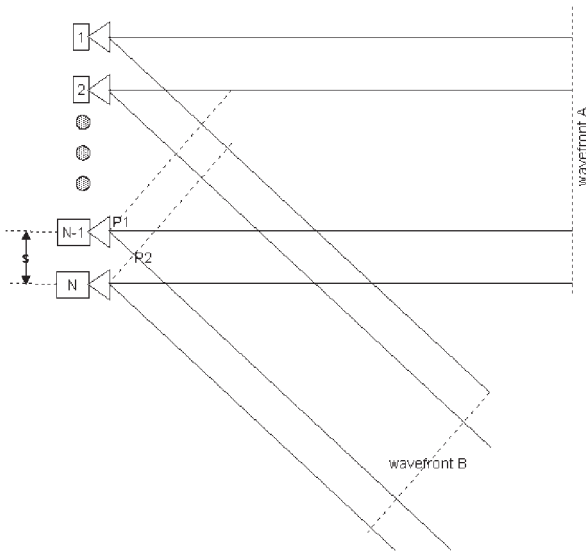


Fig. 5 When transducer delays are all equal producing a beam dead-ahead (wavefront A), at a frequency where distance P1.P2 ($\leq s$) is one wavelength then a second (alias) beam is produced in the direction of wavefront B.

because of the spatial *undersampling*. When the array is steered dead-ahead, orthogonal to the array-plane, then at the frequency $F_N = c_0/s$ an alias beam first appears, at right angles to the main beam and in the plane of the array. As F increases beyond this Nyquist limit F_N , the alias beam creeps outwards further into the half-space in front of the array (see Fig. 5), additional alias beams appearing as F increases further. When the array is steered to other angles, the first alias beam appears at an even lower frequency than F_N . So we need **closely-spaced** transducers to steer high frequencies without alias beams.

The third factor is the transducer radiating size d . A practical upper limit is $d \leq s$ as otherwise transducer mechanical overlap occurs which is very difficult to manage. In an ideal world, transducers would perfectly tile the array with no gaps as this would maximise the radiating surface area thus maximising radiation efficiency. However, the individual transducers have their own directional characteristics and the larger they are the more they concentrate their energy on-axis, this effect increasing with frequency. No matter what steering delay patterns are fed into the DDA, the array cannot steer more energy to any location than the transducers themselves are radiating to that location. Practical DDAs for home audio require steering angles up to at least 60 deg off-axis, and bandwidths up to at least 15 kHz.

Although the effects of transducer radiation patterns can to some extent be offset by beam EQ, if the maximum power to be delivered to such an extreme angle, at the highest operating frequency, is to be say no more than 3 dB down on the on-axis power, then the transducers in the array need to be equivalently non-directional. In practice up to 40 mm

diam transducers can perform adequately. So we need **small-diameter** transducers for wide beam-angle capability.

The array beam-shape and the array *aperture-illumination* function $I(x, y)$ (the radiated acoustic intensity as a function of position in the plane of the array) are related by the Fourier transform. Thus a Gaussian $I(x, y)$ will produce a Gaussian beam-shape free of sidelobes, whereas a rectangular $I(x, y)$ will produce a *sinc* function beamshape with many significant sidelobes. Just as in Fourier time-series signal-processing where one *windows* finite data sets with a smooth time-window-function, to reduce end effects and their artefacts in the FT, so one can usefully *window* a DDA array with a spatial-window-function $I(x, y)$; here the process is called *appodisation*. Truncated raised-cosine appodisation functions work very effectively in practice, though as with signal processing, too much appodisation, while creating clean mostly sidelobe-free beamshapes, also causes widening of the main beam, and reduction in maximum emitted array power.

It is important to understand *why* any such array has a useful directional effect. At its simplest, the array transducers are driven with signals each delayed so as to *constructively interfere* at the focal point of the array. This ensures that the intensity here is N times what it would be were only one array transducer to be driven (ignoring small effects due to Tx-focal point distance differences). However, at other positions the same distance from the array the signals from the different Tx will in general arrive in *random phase* giving an average intensity of $N^{1/2}$. At certain points at this distance the individual Tx signals will more or less *destructively interfere* creating deep nulls in intensity, which give the characteristic lobular beam pattern of array antennae. Thus the average in-beam to background-floor intensity ratio is expected to be $\sim N^{1/2}$. Thus for good “signal to noise” ratio one needs a **large number** N of transducers.

6. NON-UNIFORM ARRAYS (NUA)

In all the examples above the transducers were uniformly spaced with Tx centre separation $= s$. At frequencies above the (spatial) Nyquist frequency F_N the reason alias beams appear is that there is more than one solution to the beam-forming equation (because of the periodicity of sine/cosine functions and the regularity of the Tx spacings), and signals from all of the transducers add in-phase in *more* than one direction in space, viz. at the intended beam focal point and at similar focal points in each alias beam (see Fig. 5). Note that all of these alias beams have the *same nominal strength* as the intended beam and so cannot easily be ignored (see Fig. 6).

As it is generally impractical for cost and complexity reasons, to implement domestic DDA arrays with $F_N > \sim 20$ kHz, because of the small spacing and large transducer

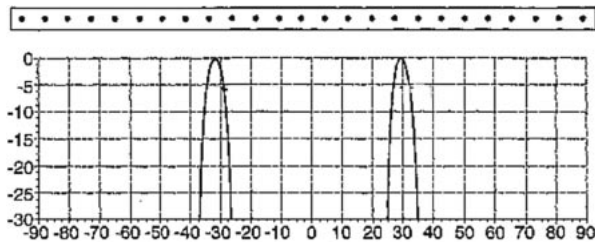


Fig. 6 A *uniform* 1D array (top) and the full-power alias beams produced above its Nyquist frequency. The lower plot shows beam intensity [dB] vs angle off-axis, when the main beam is focused at +30 deg.

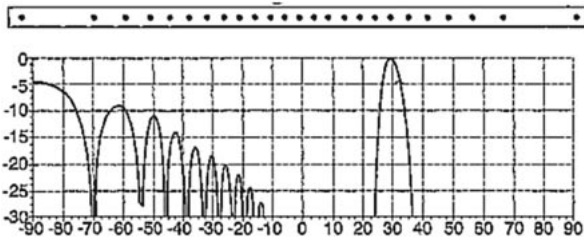


Fig. 7 A *non-uniform* 1D array (top) and the greatly reduced alias beam (bottom) produced above its Nyquist frequency. The lower plot shows beam intensity [dB] vs angle off-axis, when the main beam is focused at +30 deg. This array has the same total length and number of transducers as that in Fig. 6 and is operating at the same frequency.

count, in-band alias beams seem to be inevitable. However, a method of greatly reducing the strength of alias beams is to *non-uniformly* space the array transducers, in such a way that while their emissions all add in-phase at the desired beam focal point, there is *no other* location where this occurs. It would take a whole paper to describe even the most useful 1D and 2D non-uniform spatial layouts. It is however worth pointing out that NUA reduction of alias-beam strength does come at some cost (as with appodisation), and in this case the depth of nulls between beam and alias directions is reduced, and the amplitude of residual sidelobes (even after appodisation) is increased (see Fig. 7).

As with most such engineering tradeoffs, very careful selection of specific NUA layouts is needed to gain a useful *subjective* advantage, this after all being what a surround-sound system is for!

A special form of NUA is a planar array with a “hole” in it. Such an array might usefully be formed by arraying a number of transducers around e.g. the perimeter of a flat-screen TV. The primary problem with such gapped structures is that the gap is equivalent to an aperture illumination function with central zeroes and these generally give rise to significant unwanted lobes which are a combination of *sinc* sidelobes and alias beams. However,

we have found by analysis, modelling and experiment that excellent subjective performance may be achieved with such *screen-surround* arrays by distributing the transducers along certain complex 2D curves. These as well as many useful 1D and 2D non-uniform array structures may be found in the patent literature.

7. NON-PLANAR ARRAYS

Finally we consider arrays where not all (and perhaps no more than two) of the array transducers lie in a plane, instead being distributed on some 3D surface (for practical *transmitting* arrays this is usually a non-planar 2D surface, such as that of a cylinder or sphere, but for a *receiving* array a true 3D distribution of (small) microphones is possible and useful).

For reasons of space and clarity we will here consider only a 2D spherical DDA. Assuming the sphere’s surface is uniformly triangularly arrayed with transducers, then in the far-field the array looks much like a planar circular array and indeed behaves much like that, when appropriate beam steering and appodisation parameters are applied. However, the unique advantage of such a device is that it is capable of steering a beam in any direction at all within all 4π steradians. Importantly, it is capable of steering *any number* of such beams simultaneously each in a different direction and with different focal length (or beam divergence), and each carrying different programme material if desired.

The essential difference between any non-planar array and the planar arrays we have examined more closely, is that for certain beam directions, some of the transducers will be “shadowed” by the non-planar structure and other transducers (i.e. there will be no straight-line path between the shadowed transducers and the focal point). At very low frequencies this is less important as diffraction allows shadowed transducers to participate effectively in beams in most if not all directions. At high frequencies however the Tx shadowing is dominant and the beam forming processor is then required to select only an appropriate transducer subset (e.g. no shadowed Tx) to be used for each beam direction, on a case by case basis and allowance also needs to be made for any transducer-focus distance modified by diffraction. In practice the situation is more complex because just as shading interferes with certain Tx contributing effectively to the *constructive* interference at the desired focal point, so it also interferes with the process of *destructive* interference which is what produces the low intensity levels everywhere else in planar arrays. Special measures have to be introduced to eliminate spurious beams due to this effect.

8. FINALLY

There are numerous subtle refinements needed to make a *good* Digital Delay Array, which there is not space

enough here to describe. Nonetheless the above should serve to provide a solid grounding in all major aspects of DDA engineering.

REFERENCES

- [1] E. J. Jordan, "Multiple-array loudspeaker system," *Wireless World*, Mar., pp. 132–134 (1971).
- [2] H. Yanagawa, K. Saito and S. Hagiwara, "Speaker system and method of controlling directivity thereof," *United States Patent*, filed Feb. 4 (1992).
- [3] A. Hooley, P. T. Troughton, A. G. Goudie, I. A. Bienek and P. R. Windle, "Method and apparatus to direct sound using an array of output transducers," *International Patent Application*, WO0123104, filed Sep. 29 (2000).
- [4] H. Haas, "On the influence of a single echo on the intelligibility of speech," *Acustica*, **1**, 49–58 (1951), reprinted *JAES*, **20** (1972).
- [5] H. Wallach, E. B. Newman and M. R. Rosenzweig, "The precedence effect in sound localization," *Am. J. Psychol.*, **62**, 315–336 (1949).
- [6] E. Meyer and C. R. Schodder, *Goett. Nachr. Math. Phys. Kl.*, **11a** (1962) [in German]; also in L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics*, Vol. 1, Trans-

lated by T. J. Schultz (Applied Science Publishers, London, 1982).

- [7] J. P. A. Lochner and J. F. Burger, "The subjective masking of short time delayed echoes, their primary sounds, and their contributions to the intelligibility of speech," *Acustica*, **8**, 1–10 (1958).



Tony Hooley gained a 1st in Electronics from London University in 1971, and then gained a PhD in Radio Astronomy at the Cavendish Laboratory, Cambridge. He then moved to the Institute of Astronomy Cambridge as part of the design team of the Automatic Plate Measuring machine, and as IBM Research Fellow. Hooley founded his first company, Eicon Research Ltd, in Cambridge in 1978. In 1986 he moved to Western Australia to work on image processing and systems software, and later to Sydney where he ran the University of Technology's commercial company Insearch Ltd. After returning to Cambridge in 1993, Hooley founded *I... Limited* in 1995, on the basis of his first digital loudspeaker patent based on unary coding. With the invention of the Helimorph actuator and the Digital Sound Projector, Hooley's company grew to ~40 people by 2001. Hooley remains as Chief Scientific Officer of *I... Limited*.