Special Series for Signal Processing Magazine
in celebration of the 50th IEEE anniversary of the IEEE Signal Processing

Past, Present, and Future of Audio in the Signal Processing Society

Technical Committee on Audio and Electroacoustics
Edited by Mark Kahrs

1 Preface

It is fitting that the newest technical committee review, published in the last *Signal Processing* magazine be followed by the oldest: the Technical Committee on Audio and Electroacoustics. As many readers know, the Signal Processing Society has its roots in the Audio and Electroacoustics committee. In the past 50 years, there have been incredible changes in audio and electroacoustics directly due to the invention and use of the transistor and integrated circuit technology. In the following sections, the authors (all members of the Technical Committee) will review some of the progress in signal processing technology since the creation of the Audio and Electroacoustics Committee almost 50 years ago.

Appropriately, the article begins with Gary Elko's description of the development of electroacoustic devices. Starting with a complete history of electroacoustic transducer design, Elko shows how judicious use of array signal processing technology such as beamforming can enhance the reception of signals. DSP technology can also be used to compensate for the very real nonlinear performance of electroacoustic transducers.

Active Noise Control (discussed by Stephen Elliot) exemplifies the application of signal processing methods to acoustic problems with feedback. Active Noise Control can be used to cancel offensive sound fields as well as compensate for room acoustics.

Shoji Makino next reviews Acoustic Echo cancellation (as opposed to cancellation of echoes due to line echo). Acoustic echo cancellation is increasingly important with the development of portable communication devices used in acoustically reverberant environments like automobiles as well as the use of speakerphones for teleconferencing. Adaptive filtering technology is used to account for the changing acoustic environment.

Another direct application of signal processing technology is the use of signal processing to Hearing Aids (written by Jim Kates). New algorithms that accommodate the properties of human hearing are discussed along with the fundamental limitations of hearing aids themselves.

Another example of psychoacoustic modelling can be found in the use of models in audio coding. These methods are discussed by Marina Bosi and show how a wide range of different coding methods are derived from fundamental studies of human auditory perception.

The Audio and Electroacoustics committee is interested in *wideband* audio – not just speech but also signals such as music. Julius Smith reviews recent work in developing physical models of musical instruments for synthesis.

Finally, Mark Kahrs discusses the incredible progress of consumer and professional audio gear due to the the implementation of sophisticated signal processing. In particular, he concentrates on the development of digital technology.

2 Electroacoustic Transducers

Gary W. Elko
Acoustics Research Department
Bell Labs,
Lucent Technologies,
Murray Hill, NJ 07974
gwe@research.bell-labs.com

2.1 Introduction

The word *transducer* is generally used to denote a device that converts between one form of energy to another. The more descriptive term, *electroacoustic transducer*, connotes a transducer the converts acoustic energy to electrical and vice-versa. The history of electroacoustics actually predates that of living organisms on Earth. You might initially think that this conjecture is heretical, but one only needs to think about the physics of a common thunderstorm with its bright flashes of lighting (electrical) and commensurate thunder (acoustical). More recently, the history of electrical communication is inexorably linked to electroacoustics. The most early forms of electrical communication via telephony hinged on the successful discovery and invention of a working electromagnetic transducer in 1877 by Alexander Graham Bell [9].

The present field of electroacoustics is marked by a wide array of applications not limited to buy including: ultrasonic imaging and cleaning, non destructive testing, condition monitoring, audio, telephony, acoustic agglomeration, acoustic levitation and manipulation, acoustic refrigeration, sonar, active noise control, acoustic ranging and direction finding, material physics, nonlinear acoustics, atmospheric and ocean acoustics.

As you can see from the wide and varied list above, there are many applications of electroacoustic transducers and each application can have many unique transducer designs that are effective. One way of narrowing down the categories of transducers is apply different criteria for the transducer operation. One typical breakdown in classification is to divide the transducers on the basis of *linearity*, *reversibility*, and *passivity*.

Transducers are linear if the predominant variables that describe the output or performance of the device, such as sound pressure or velocity, are predominantly linear functions of the transducer input. Typical transducers that fall into this category are dynamic and condenser microphones, and loudspeakers

of various types. The qualifying term "predominantly" allows for the inevitable small nonlinear operation found in almost all real transducer systems. The term passivity implies that all energy that is transduced is obtained from the acoustic input (or electric) source ... no other external energy sources are required for the transducer to operate. Finally, the term reversibility, implies that the transducer can convert between acoustic and electrical energy irrespective of the direction of the input/output direction(acoustic or electrical). The term is sometimes more strictly applied to indicate that the conversion sensitivity is also independent of the input/output direction.

2.2 Electroacoustic transducer history

The rich history of electroacoustic transducers and the recent marriage (in the past few decades) of this field to digital signal processing makes for an exciting story. This section will review a few of the milestones in the electroacoustic transducer field [1] and speculate on applications that might change the way we work and communicate in the future.

The first electromagnetic electroacoustic transducer appears to be a classroom demonstration device created by Joseph Henry in 1831. The demonstration device utilized and electromagnet that would attract and repulse a magnetic armature. By switching the direction of the applied current to one electromagnet, (through the distance about 1 mile of wires strung along the walls of a lecture hall) the magnetic armature would radiate impulsive sound as the armature impacted the driving electromagnet. Magnetostriction was discovered shortly after when it was observed that a soft iron bar wrapped with a conductor would radiate a tone when the current in the conductor was interrupted [136]. The frequency of the tone was observed to be equal to the fundamental longitudinal vibration mode of the bar. Joule [96] made very precise measurements of the strain in the bar as a function of applied current and is generally credited with the discovery of magnetostriction.

2.2.1 The Telephone

But probably the biggest catalyst to the further investigation of acoustic transducers was the invention of the telephone by A. G. Bell in 1876. It is well known that Bell was experimenting with the "harmonic telegraph". What is not commonly known is that even though Bell had a notion and a desire to make a device that would transmit speech, his discovery of the moving armature reversible transducer was actually by luck. In his experimentation with tuned vibrating reeds for the telegraph, his assistant Thomas Watson accidentally over-adjusted a contact screw so that current to the electromagnet was not interrupted by the motion of the reed. When Watson plucked the magnetized steel over the electromagnet while adjusting the contact screw, Bell noticed an

[1] For the reader who would like a more complete and engaging historical review, see the classic book *Electroacoustics* by Hunt [81]. Another very useful reference for the history telephony is, *A History of Engineering and Science in the Bell System* [65]

interesting new sound coming from his end of the telegraph. Even though this setup never produced intelligible sound, it can be claimed that this device was the first to transmit voice by electricity. [This serendipitous event, which led to a one of the most important inventions in modern history (in dollar amounts as well), should be more well known in engineering and business circles. Research breakthroughs can not be planned; they occur when people with the required skills and desire, work with insight and persistence in their pursuit of a vision.]

Figure 1 shows a figure from of Bell's first telephone patent that shows the essential invention of the moving armature transducer. The DC batteries shown in Figure 1 are used to form a magnetic field. The vibrating armature varies the reluctance of the magnetic circuit set up by the electromagnet. The variation

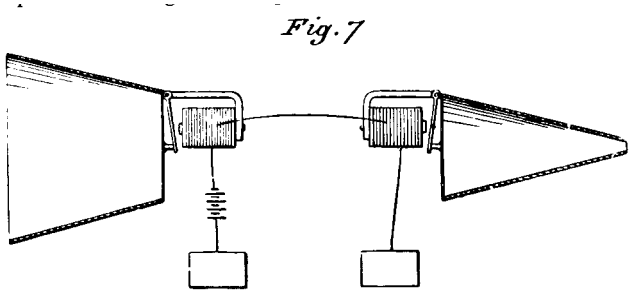


Figure 1: Figure 7. from Bell's first patent of the telephone [9].

of the reluctance causes an oscillating current proportional to the displacement of the armature. Since this transducer is reversible, Bell used the same transducer for both sending and receiving. Although Bell used another "transmitter" (telephone jargon for microphone), the variable-resistance liquid transmitter for early demonstrations of his telephone, he quickly abandoned this type of nonreversible transducer. The next major advance in transducer design came with the invention of the contact-pressure carbon microphone that is credited to Thomas Edison, although others also proposed essentially the same design around the same time in the late 1870's. The contact-pressure carbon microphone works on the principle that the resistance of packed carbon granules varies with the applied pressure. By applying an external bias current, the voltage generated by this transducer is proportional to the applied acoustic pressure. This transducer is both non-passive and non-reversible. The gain attained by the carbon microphone was as much as 30 dB over previous microphone and as a result, quickly replaced Bell's electromagnetic transmitter. (It should be pointed out here that the electronic amplifier did not become available until 1914 ... about 35 years later!).

2.2.2 Moving Coil Transducers

Another major advance in transducer design was the invention of the moving coil loudspeaker in 1877 by E. W. Siemens [172]. Although Siemens had patented a motor invention that used a circular coil in a radial magnetic field in 1874 (two years before Bell's patent), he never pursued this device for telephony. Siemen's 1877 patent of the moving coil receiver (loudspeaker) is essentially the motor design used in modern loudspeakers and telephone receivers (telephone jargon for loudspeaker). The famous 1925 paper by Rice and Kellogg [159] essentially summarized the refinements of made by many designers in the field of moving coil loudspeakers. The Rice and Kellogg paper was the first publication that described the basic design and physics of moving coil loudspeakers. For the readers who are unfamiliar with the basic radiation physics of a loudspeaker: the acoustic output power of a loudspeaker can be calculated by taking the square of the volume velocity (average normal velocity over the loudspeaker surface) and multiplying this be the real-part of the radiation impedance. For a flat rigid circular piston it can be shown that the radiation resistance increases by the square of frequency [14]. If we now have a transducer whose velocity decreases linearly with frequency, then the power output will be constant. The modern moving coil loudspeaker accomplishes this by tuning the resonance of the diaphragm to the lowest desired frequency. Above resonance the loudspeaker operates in the mass-controlled region and the velocity falls linearly with frequency ... just what is desired for flat operation. The fundamental principles described in the Rice and Kellogg paper is the basis upon which all moving coil loudspeakers are designed today.

One enhancement that is commonly found in modern loudspeakers is the use of a bass reflex design. The design and fundamental explanation of the bass-reflex was due to A. L. Thuras [196] who patented the design in 1930. The bass reflex essentially extends the low frequency response of a baffled moving coil loudspeaker and also reduces low frequency distortion by minimizing diaphragm displacement at the box resonant frequency. A good engineering review of the design of vented-box and closed-box loudspeaker systems can be found in papers by Thiele and Small (reprinted in AES loudspeaker anthology [38]).

At around the same time as the bass reflex enclosure, Wente and Thuras [208] also were responsible for the first high-quality dynamic moving coil microphone and telephone receivers. Western Electric commercialized this microphone as the 618-A moving coil microphone. The microphone was operated as a resistance controlled system so that the output voltage on the moving coil was proportional to the velocity of the cone. Since the voice coil impedance was rather low ($\approx$ 30 ohms) the moving coil microphone was able to drive long wires without loss in level or frequency response. As a result, the dynamic microphone became very popular as a high-quality easy to use general purpose microphone and was used extensively in sound recording and radio broadcasts. As an aside here, a variant of the 618-A microphone was the Western Electric 630 non directional microphone that became available in the late 1930's [137]. The physical structure of the WE-630 microphone was a sphere with a thin resistive

screen disk mounted over the moving coil microphone. A cross sectional view of this microphone is now the official IEEE symbol for a microphone.

The next advance in microphone design occurred with the invention of the unidirectional ribbon microphone by Olson at RCA in 1931. The ribbon microphone us constructed by placing a light and highly compliant conducting ribbon in a magnetic field. If the ribbon has a mechanical impedance close to that of the medium (air), it moves sympathetically with the acoustic particle velocity. The motion of a conductor in a magnetic field induces a current in the ribbon proportional to the motion of the ribbon. Further developments at RCA led to the first unidirectional microphone (RCA model 77). The RCA 77 unidirectional microphone was constructed by combining two ribbon transducers: one pressure sensing and another acoustic particle velocity sensing (directional response of $\cos(\theta)$ where θ is the angle relative to the normal of the ribbon surface). By properly combining these two sensors (by equal amounts in this case), a unidirectional cardioid microphone with directional pattern $1/2[1+\cos(\theta)]$ can be constructed. The directional gain from this microphone is 4.8 dB. Due to the directional gain unidirectional microphones became very popular since they offered some relief from reverberation and feedback. Western Electric followed in 1939 [137] with a variation on this theme by combining ribbon microphone with a moving coil pressure microphone. The unique contribution here, was the inclusion of a six-position switch that allowed the user to select a desired beampattern between omnidirectional and dipole. Also, this microphone was the first to offer both hypercardioid (maximum directional gain) and supercardioid (maximum front to rear power) patterns for this type of microphone design. Finally, in 1938 B.B. Bauer [6] noted some directional pattern irregularities for acoustic velocity sensing microphones when he used different acoustic screens on either side of the diaphragm. This discovery led to an equivalent circuit analysis which showed that the irregularities could be attributed to phase-shifts between the two sides of the diaphragm. By designing an acoustic R-C phase shifting circuit, Bauer was able to make a unidirectional microphone with only one single element. This simple and elegant design is the basis of almost all unidirectional microphones being manufactured today.

2.2.3 Electrostatic Transducers

The appearance of electrostatic transducers which is inherently a reversible transducer, came with two independent developments of electrostatic transmitters and receivers for telephony by Dolbear [47] and Edison in the late 1870's. The fact that they could even make a working device with these inherently high impedance devices is a testament to the ingenuity of these inventors. With the advent of the vacuum-tube amplifier the condenser (capacitor) microphone became a very useful device for high precision measurement. Edward C. Wente is credited with the invention of the modern condenser microphone [207] in 1917 and the most high quality microphones found today are based on this principle. The condenser microphone has the advantageous quality of having an extremely flat and smooth frequency response and very low distortion. The Wente con-

denser microphone worked in the stiffness controlled region (below the 10 kHz resonance), and was very flat to frequencies up to 8 kHz. Modern laboratory quality condenser microphones are constructed in a similar fashion to Wente's microphone, but have resonance frequencies typically in excess of 20 kHz.

One basic problem with the condenser microphone design that limited its applicability in consumer products is the need for an external high-voltage DC bias. This limitation was overcome by the invention of a stable "electret" microphone by Sessler and West in 1962 [170]. The electret microphone embeds a static negative charge into a stable dielectric (teflon) and thereby removes the requirement for and external bias. The electret microphone is the most common microphone found today and greater than 100,000,000 are manufactured annually! More recently there has been an interest in making microphones using standard microelectronics techniques [169]. At present, these transducers are still in the research stage but show great promise by allowing the transducer(s) and digital signal processing to be combined in a single easily fabricated device.

2.2.4 Piezoelectric transducers

Interest in reversible piezoelectric transducers began in the early 1900's with efforts directed to finding icebergs and submarines. P. Langevin [124] working on receivers and transmitter for underwater applications experimented with quartz that was well known to be a piezoelectric material (the common erroneous consensus of the day was that the piezoelectric effect was too weak to be useful). Langevin experimented with many inventive sandwich transducers made with layers of quartz and steel. His research led to underwater transducers that had output powers in excess of 20 dB more than previous electrostatic receivers used for underwater sound generation. Research and development of piezoelectric material continued with the investigation of other crystalline structures such as Rochelle Salt. In the mid 1950s sintered ferroelectric ceramics became available. The piezoelectric ceramic PZT was the most common of these new materials. PZT is a combination of lead (Pb), zirconium (Zr), and titanium (Ti) oxides or salts and after sintering, consists of small ferroelectric domains. The ceramics are "poled" by heating to a temperature close the Curie temperature and applying a strong DC electric field ($\approx$ 20 kV/cm) while slowly lowering the temperature. Newer materials and composite structures for piezoceramic transducers is still an active area of research. Some of the applications where piezoelectric devices are widely used are: hydrophones (underwater microphones) and accelerometers.

2.3 Signal processing and transducers

The historical descriptions of the electroacoustic transducers in the previous sections is by no means exhaustive. Even with this incomplete list, it can be seen that this field is quite broad and has had many inventive and creative contributors.

The transducer is essentially the "front-end" of a system that utilizes acoustic energy to transmit or receive information. With the advent of more sophisticated electronics and the desire for more directional and controllable directional patterns, modern transducer designers typically work with arrays of transducers. The processing of the multiple input/output signals now allows the research and design engineer to incorporate sophisticated multidimensional signal processing algorithms to attain some desired performance goal.

The basic goals of array signal processing are: to enhance the signal-to-noise ratio (SNR) of a desired signal (relative to a single sensor), estimate parameters of the sound field or medium (number of sources, source position, physical properties of the medium etc.), and to track the source or sources as they move around in the medium. The field of array signal processing is now a very active research area and only a few examples will be discussed here. (There are many excellent references that cover varying aspects of the problem [49], [94]).

2.3.1 Delay-Sum Beamforming

Delay and sum beamforming is one of the oldest and most straightforward approaches to spatial filtering (beamforming). The basic idea is to delay the output of the multiple sensors so that a signal arriving from a desired direction adds-up constructively and commensurately, uncorrelated noise and signals arriving from other directions add incoherently. If the noise is uncorrelated from sensor to sensor, the gain in SNR is equal the square-root of the number of sensors. Figure 3 shows the general structure for the delay-sum beamformer. Simple amplitude weighting of the elements adds another design degree of freedom and is referred to as array taper or array shading. Another feature of array beamforming is the ability to electrically steer the desired receiving direction, or to form many unique beams simultaneously. Although the reader might consider such a simple approach to be ineffective, the delay-sum beamformer is still in active use today. It is interesting to note that it was the original analysis of the a uniformly spaced discrete array that laid the foundations of modern digital signal processing FIR filter design and analysis [165]. A few of the applications for delay-sum beamforming include: underwater active and passive sonar, medical ultrasound and ultrasonic imaging, and in self-steering microphone arrays for teleconferencing [67].

2.3.2 Filter-Sum Beamforming

A more general structure of the beamformer is to place an selectable filter behind each sensor input. Figure 4 shows a schematic representation of this type of array structure. Clearly the delay-sum beamformer is a subset of this topology. The advantage of using this more general structure is that it allows for the most general spatio-temporal filtering necessary for broadband applications. In audio teleconferencing applications, it is desirable to keep the directional pattern of the beamformer constant over the many octaves of speech and music bandwidth. Work on this problem has been an active area of research over the past decade

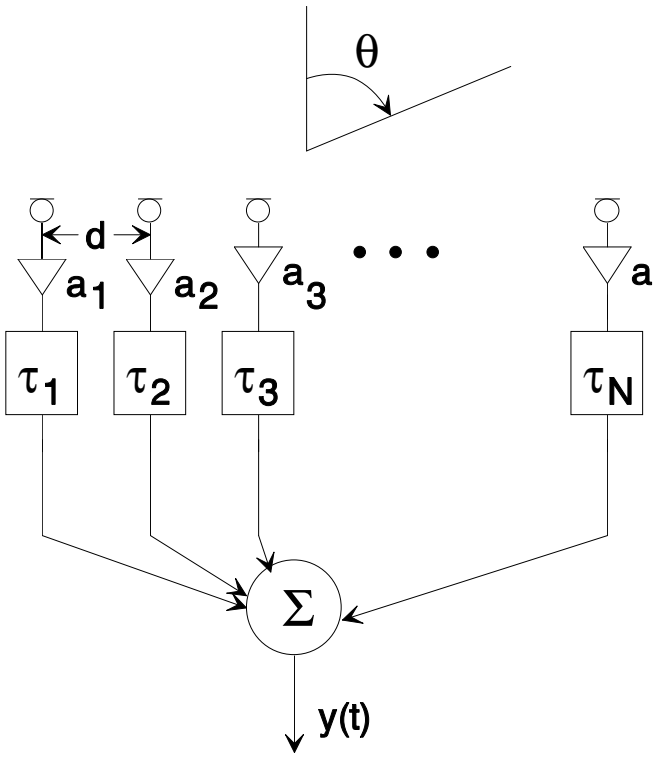


Figure 2: Schematic diagram of delay and sum beamformer. The elements are spaced by d, the a_n variables are amplitude scaling and τ_n are the delay elements.

([56], [205]).

2.3.3 Adaptive Beamforming

The techniques of classical delay and sum beamforming as well as filter-sum beamforming rely on the designer to adjust the shading or filter coefficients depending on some model of the acoustic environment in which the array will be used. Rarely (actually, almost never) do real-world sound fields fit nice simple mathematical models and typically the sound fields are nonstationary (constantly changing). To motivate the reader to better appreciate this problem, think of standing on a street corner as cars and trucks drive by. In order to deal with imperfect (or just plain wrong) assumptions of the sound field and to handle the time-varying nature of sound fields, beamformers that adjust their parameters based on the received data have been developed. Generically the beamformers have been classified as adaptive beamformers. Adaptive beamforming

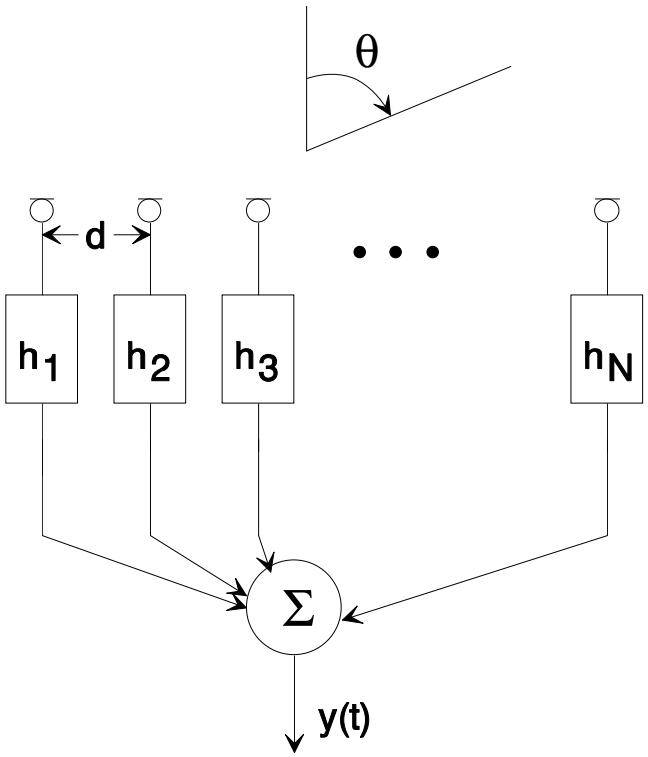


Figure 3: Schematic diagram of the filter and sum beamformer. The variables h_n represent general filters that follow each microphone output.

began with Howells [80] who was working on interference nulling. Applebaum [5] contributed an adaptive algorithm for the Howells adaptive nulling scheme that maximizes a generalized signal-to-noise ratio (SNR). Concurrently, Widrow and others applied adaptive control techniques to arrays [211]. In the following decades literally hundreds of papers and many books have been written on the subject (see [142] for instance). In the area of audio, work on adaptive arrays has concentrated on the problem of teleconferencing [106],[187],[56]. The problem of dynamically adjusting the gain and frequency response of close-talking microphone that is moving relative to the talker was also attacked using an adaptive technique by West et. al. [58].

2.3.4 3D sound-field microphones

With the recent interest in surround sound and virtual sound field processing, there have been some interesting new transducers arrays that have been sug-

gested. The earliest is probably due to Christensen [157] at RCA labs in the early 1970's. Christensen used a group of four closely-spaced microphones that were combined in a way as to transmit the acoustic pressure and two components of the pressure difference. By appropriately combining these signals, the user could synthesize a surround sound field (with 4 or more sources). Craven and Gerzon [41] added another pressure difference dimension so that the encoding would now allow a true three-dimensional sound playback configuration (speakers were not restricted to be placed in a plane). Elko [57] has also recently proposed using a very small rigid sphere of six pressure elements and digital signal processing to produce a 3-D surround output. The advantages of this type of array is that the beampatterns can be controlled to higher frequencies and the use of DSP processing allows for the use of inexpensive elements and accurate calibration.

2.3.5 Loudspeaker arrays

Loudspeaker arrays have been in use in public reinforcement systems now for more than 60 years. Surprisingly, loudspeaker array beamforming using digital signal processing techniques has been a relatively recent development. (Sonar systems were using digital beamforming techniques in the late 1960's). A digital beamforming loudspeaker array system developed by Goodwin and Elko [72] was used for experiments in frequency independent beamforming (constant beamwidth over 3 octaves) and feedback mitigation. Another digital system was developed at IRCAM by Derogis, Causse, and Warfusel [44] for music and room acoustics research. The array that was constructed for this research consisted of 12 loudspeakers located on the pentagonal facets of a dodecahedron.

2.3.6 Loudspeaker distortion compensation

Another relatively new promising application of signal processing to electroacoustics is the work by Klippel [116] on the reduction of nonlinear distortion in loudspeakers. The scheme proposed by Klippel is an open-loop feedforward design. The desired (undistorted) signal is fed into a "mirror" filter: a filter that predistorts the loudspeaker input signal so that the sound pressure signal produced by the loudspeaker will be essentially free of distortion.

2.4 Closing remarks

As one can see, the field of electroacoustics has an interesting history and a bright future. As people continue to move towards "anytime-anywhere" communication, there will be new requirements for transducers that enable this untethered mode of communication. It is not hard to speculate that complete high-fidelity audio communication and entertainment systems will mount either in or near the ears. By mounting the transducers near or in the ear canal, the required power is low, the user does not interfere with others and has privacy. A transducer system mounted in the ear canal will enable wide-band active

sound control and thereby allow the user to dial in a desired external noise sensitivity or desired spectral equalization and compression. The user will also be able to monitor bodily variables such as temperature, blood pressure etc.,(with acoustic and non-acoustic transducers). Binaural communication will enable many useful new features: spatially distinct information channels, audio driven directional guidance (imagine the combination of GPS with binaural audio presentation for navigation), true full duplex communication even in the harshest environments, and many more synergies yet to be discovered and invented. It is therefore not difficult to see that the combination of digital signal processing and novel electroacoustic transducers will play a very big role in creating and enabling the exciting new modalities of human-to-human and human-to-machine communications.

3 Active Noise Control

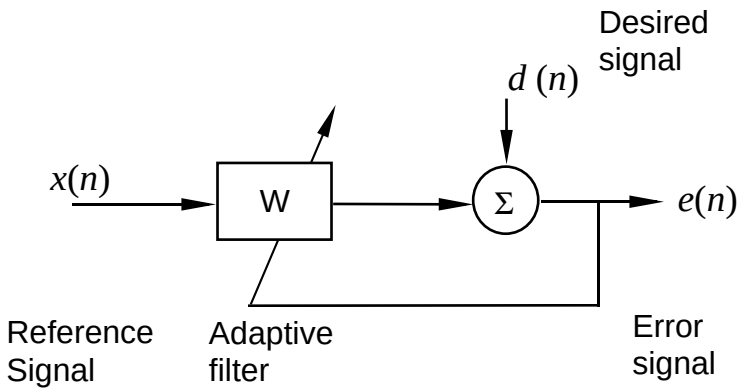
Stephen J. Elliot
Signal Processing & Control Group
Institute of Sound & Vibration Research
University of Southampton
Highfield, Southampton UK SO17 1BJ
sje@isvr.soton.ac.uk

3.1 Introduction

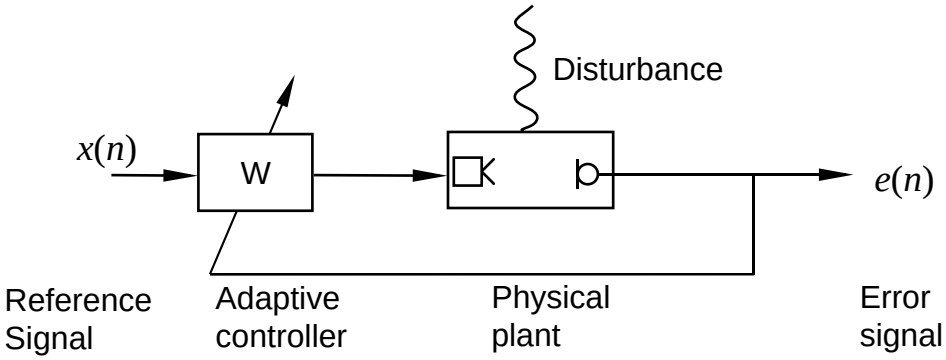
The active control of acoustic noise was first proposed in the 1930's as an electronic way of reducing sound propagating in air ducts or in the free field. Despite several important developments in the 1950's, it has only been in the last decade that developments in DSP have allowed practical control systems to be implemented commercially. This has spurred a rapid advance in our understanding of the physical limitations of this method of controlling sound, particularly in enclosures such as aircraft and cars, and has prompted parallel developments in the active control of vibration and fluid flow, and in the techniques used to equalise sound reproduction systems.

Active noise control (ANC) involves the use of a controllable,"secondary" , source of sound to generate an acoustic field which destructively interferes with the original,"primary", sound field. In order to maintain the delicate balance necessary for good cancellation, it is generally important that the secondary source is adaptively adjusted to compensate for changes in the primary source. When an active noise controller is implemented as an adaptive digital system, it has some similarity to an adaptive noise cancellation system(Widrow et al, 1975), in which the cancellation occurs purely in an electrical signal, rather than in the physical pressure which is the case for active noise control. The same techniques have been used to adjust the digital filter in adaptive noise cancellation, and the digital controller used in many feedforward active noise control systems, and similar block diagrams can be drawn for both systems, as

shown in Figure 4.



(a)



(b)

Figure 4: Block diagrams for an adaptive noise cancellation system, (a), which minimises a purely electrical error signal, and that for a feedforward active noise control system, (b), in which the object is to minimise the pressure at the microphone position, and the adaptive controller can only affect the error signal through the response of the physical plant.

The object of the adaptive noise canceller, in Figure 4(a), is to cancel the components of the electrical desired signal $d(n)$ which are linearly correlated with the reference signal $x(n)$ by minimising the mean square value of the error signal $e(n)$, and this is commonly achieved using the LMS algorithm to adapt the filter, W. The object of the active noise control system is to reduce the pressure at a given microphone signal caused by an external disturbance, and this is achieved by adjusting the coefficients of the digital filter driving the loudspeaker acting as the secondary source to minimise the mean square value of

the electrical output from the microphone as shown in Figure 4(b). The physical "plant" shown in Figure 4(b) contains the electroacoustic response between the secondary loudspeaker and error microphone, together with the response of the data converters and analogue antialiasing and reconstruction filters in a digital implementation. The presence of the plant response has an important influence on the adaptation of the control filter, and a modified form of the normal LMS algorithm, called the filtered-x LMS, must be used to ensure reliable convergence (Widrow and Stearns, 1985; Elliott and Nelson, 1993). The fact that one is trying to influence physical variables, and the fact that these variables can only be influenced through a dynamic and potentially time-varying "plant" response means that in some ways active noise control is closer to being an automatic control problem than it is to a signal processing one. The prevalence of adaptive feedforward techniques, however, has meant that the traditions of adaptive signal processing have carried over into active control and most work in the field is rooted in this tradition.

The reason why feedforward techniques are so prevalent in active noise control has to do with the way the sound is originally generated in many important problems. If the sound wave is propagating along a duct, for example, a measurement of its waveform can be made some distance upstream of the secondary source and used to provide a time-advanced reference signal for a feedforward control system such as that shown in Figure 4(b). This application was one of the original motivations for investigating active noise control and was originally described by Paul Lueg in a patent published in the U.S. in 1936.

Another broad class of noise problem to which feedforward active control systems have been applied, is the reduction of low frequency periodic sound. This is because many rotating or reciprocating machines generate sound of this form and in this case a reference signal having the same fundamental frequency is often directly available from the machine, as a tachometer output for example. This reference signal again contains information about the basic waveform of the sound to be controlled and can be used as the basis for a feedforward control system. Such systems were originally proposed for the active control of transformer noise and it is in this field that the need for adaptation was first clearly identified, to compensate for changes in transformer load and atmospheric conditions. In a seminal paper in 1956, William Conover described this need for adaptation and the difficulty of building an automatic system to accomplish this with the analogue electronic technology available to him at that time.

The developments which have taken place over the past decade in real-time signal processing devices has made possible the practical application of such adaptive feedforward control systems. This in turn has led to the development of more complicated multichannel systems and a great deal of research into establishing the fundamental physical limits of active sound control [145, 121, 76]. Current commercial applications include the active control of noise in ventilation ducts [63], and the active control of low frequency propeller noise in propeller aircraft [60].

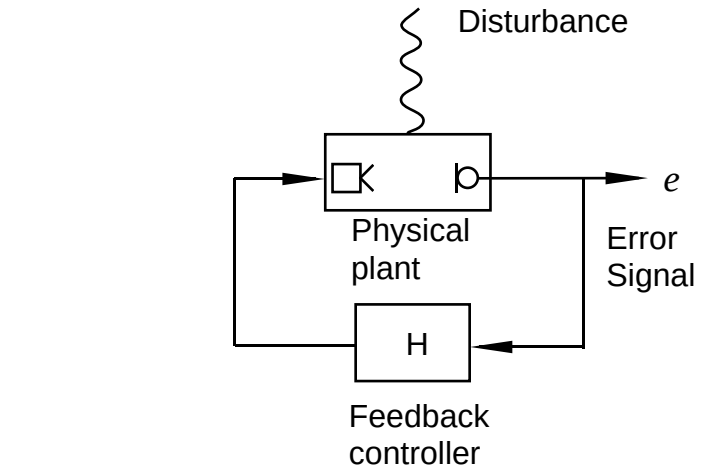
Active noise control conveniently complements more traditional, passive,

methods of controlling noise. This is because passive methods tend to work better at higher frequencies, for which the acoustic wavelength is not too large compared with the thickness of a typical absorber, whereas active control works most effectively at lower frequencies for which the acoustic wavelength is long. The acoustic wavelength is 3.4 m at 100 Hz under normal conditions, for example, and if the separation between the microphones in a multichannel active control system is much smaller than this, then the sound levels between the microphones will tend to be reduced as well as the levels at the microphone locations and"global" reductions are possible.

At higher frequencies, the reductions in pressure caused by an active noise control system are concentrated near the microphone location and become more "local". In a diffuse sound field, 10 dB reductions in pressure can only be obtained within a sphere of diameter about one tenth of an acoustic wavelength round the microphone (Elliott et al, 1988). By placing the microphone close to the loudspeaker, the delays in the response of the "plant" are reduced and stable feedback control systems can then be implemented, as shown in Figure 5(a).

Such a system is described in Olsen and May's important 1953 paper, which also discusses the application of such a local controller to the back of a seat so that a"zone of quiet" is generated round the head of a passenger in an aircraft or automobile. The bandwidth over which significant control can be achieved with a feedback system is inversely proportional to the delay in the control loop, which includes the delay in both the plant and controller. The most successful development of feedback systems in active sound control is in active headsets, where the secondary loudspeaker is positioned very close to the error microphone. Active headsets are now widely used in military vehicles [209] and can give sound reductions of more than 10 dB up to about 500 Hz. In order to reduce the delays in the feedback loop of such a system, the feedback controllers in active headsets have generally used analogue circuits. Although an analogue controller does not suffer from the delays associated with the data converters and antialiasing and reconstruction filters necessary in a digital controller, it is also much more difficult to adapt their response, and so active headsets tend to have a fixed controller, designed to work best with a given disturbance spectrum. The trade-offs in the performance of any feedback control system can be illustrated by transforming the conventional block diagram shown in Figure 5(a), into an equivalent feedforward system using a feedback controller architecture known as Internal Model Control [143], as shown in Figure 5(b). In this arrangement the effect of the secondary source has been removed from the output of the microphone using an electrical model of the plant, and if this model accurately represents the plant response, the resultant signal, x, is equal to the disturbance signal at the microphone. The control system then has an equivalent block diagram which is entirely feedforward [59], in which the reference signal is equal to the disturbance, and if the plant has any delay, reductions in the mean square error can only be achieved if the disturbance signal is predictable over this timescale, as in an adaptive line enhancer.

As well as providing a convenient reformulation of the feedback controller for



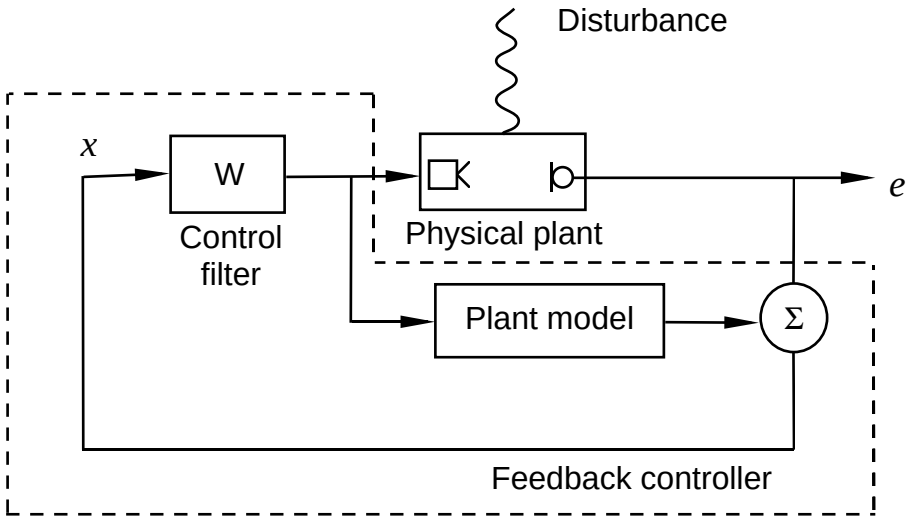


Figure 5: Block diagram of a feedback system used for active control, (a), and an architecture of controller, (b),which allows the feedback system to be cast in an equivalent feedforward form.

performance analysis, the arrangement shown in Figure 5(b) could also be used as the basis for an adaptive feedback controller, which could adjust the response of the control filter to compensate for changes in the spectrum of the disturbance using the algorithms developed for adaptive feedforward control. Such a system would need to be implemented digitally, however, and would thus introduce its own delay into the feedback loop and thus limit the bandwidth over which the disturbance could be attenuated, as discussed above. In applications where the disturbance was particularly non-stationary, the advantages of adaptation may outweigh the loss of control bandwidth, particularly if an inner analogue feedback loop is used for broadband control, and the adaptive digital outer loop is used to further reduce more tonal components.

Although active noise control has its historic roots in the 1930's and 1950's, it is only relatively recently that developments in real-time hardware have made the techniques of digital signal processing available within this application. The success of these systems has prompted an intensive investigation into the performance of these algorithms under realistic conditions, and also into the physical limits of performance. More recent work has begun to look at feedback control systems from a signal processing perspective and has focused on some of the traditional concerns of the control engineer, such as the robustness of the system's stability to uncertainties in the plant response, and the prevention of out of band disturbance enhancement [158].

Because of the physical limitations of active noise control, there has also been considerable interest recently in using active vibration control to attenuate the motion of the original source of sound, and thus control the noise at source [29]. Similar active control techniques are also being investigated for the control of fluid flow, including the use of LMS controllers [123]. Another development of these adaptive techniques, which is more relevant to the original aims of the Audio and Electroacoustics Committee, is the automatic equalisation of sound reproduction systems. This has been successfully demonstrated for the single channel equalisation of the audio response inside a car [64] and has more recently been extended to multichannel systems, which are able to significantly improve the acoustic imaging in stereo systems [146].

4 Acoustic Echo cancellation

Shoji Makino
NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan
makino@splab.hil.ntt.jp

4.1 Introduction

The echo canceller was first studied by Sondhi [185] in 1967 for controlling long-distance telephone line echo. In the 1970s, compact and economically feasible echo canceller hardware became possible with advances in LSI and digital signal

processing technologies. These technologies are good at handling complex and highly accurate calculations. The first echo canceller using LSIs was developed by Duttweiller [50] at AT&T Bell Laboratories. Nowadays, echo cancellers are recommended in ITU-T recommendations G.165 and G.167 [84], [71] and are widely used in many applications [74].

The most modern application of echo cancellers is against acoustic echo, *e.g.*, for teleconferencing and hands-free telecommunication systems. Therefore, the rest of this session focuses on acoustic echo cancellation.

4.2 Echo cancellation and adaptive algorithms

An echo canceller adaptively identifies impulse response $\mathbf{h}(k)$ between the loudspeaker and the microphone. Echo replica $\hat{y}(k)$ is created by convolving $\hat{\mathbf{h}}(k)$ with received input vector $\mathbf{x}(k)$; then $\hat{y}(k)$ is subtracted from actual echo $y(k)$ to give an error $e(k)$. An adaptive FIR filter $\hat{\mathbf{h}}(k)$ is adjusted to decrease the error power at each sampling interval. The adaptive algorithm should provide fast convergence and high echo return loss enhancement.

The LMS and NLMS algorithms are widely applied to echo cancellation since they are simple and robust. However, when the received input signal is a colored signal, their speed of convergence must be increased. To speed up the convergence of the LMS and NLMS algorithms, time-varying stepsize control has been studied [28]. Using individual stepsize for each coefficient is also effective. One method of doing this is time-varying [156] and another uses the *a priori* knowledge of the exponentially decaying characteristics of the room impulse response [162], [131].

The projection algorithm, or affine projection algorithm [148] whitens the received input signal. This process can be explained as whitening by the Gram-Schmidt process and/or whitening by the linear prediction process. Therefore, convergence can be improved for a colored input signal, such as speech, which has a highly non-white spectrum. The second-order ES projection algorithm [131] converges four times faster [75] while its computational power is almost the same as that of the conventional NLMS. The fast affine projection algorithm [194]-[48] is attractive; its computational load is $2L + 20p$ multiply-add operations, where L is the number of taps and p is the projection order.

The affine projection algorithm lies between the NLMS and RLS algorithms [130], [141]. The first-order projection algorithm corresponds to the NLMS; the infinity-order one is equivalent to the RLS algorithm. In the RLS algorithm, the received input is fully whitened, so convergence is independent of the input signal, resulting in fast convergence for all input signals. Fast RLS algorithms [11], [175] have the good convergence of the RLS algorithm at a computational complexity of $8L$.

4.3 Subband echo canceller

Subband echo cancellers divide signals into smaller frequency subbands and independently cancel echoes in each subband [113]. Since the narrower frequency

subbands have a smaller eigenvalue spread than the fullband for speech input, the convergence speed is improved. Since downsampling expands the sampling interval and reduces the number of taps needed for the adaptive filter, the subband echo canceller is computationally efficient.

In the conventional subband echo canceller, the received input in each subband is bandlimited and not fully whitened, which results in slow asymptotic convergence [144].

To overcome this problem, one suggestion is to increase the bandwidth of the analysis filter relative to the synthesis filter [125]. On the other hand, the projection algorithm was shown to be effective at whitening the received input in each subband, which speeds up convergence [132]. The major drawback of the subband structure is the delay that is introduced by the filter banks.

4.4 Stereo echo canceller

Stereo echo cancellation is a hot topic and of interest from both the theoretical and implementation points of view.

A stereo teleconferencing system provides more realistic presence than a monaural system. It helps listeners distinguish who is talking at the other end by means of spatial information. A stereo (two-channel) telecommunication system is shown in Fig. 6.

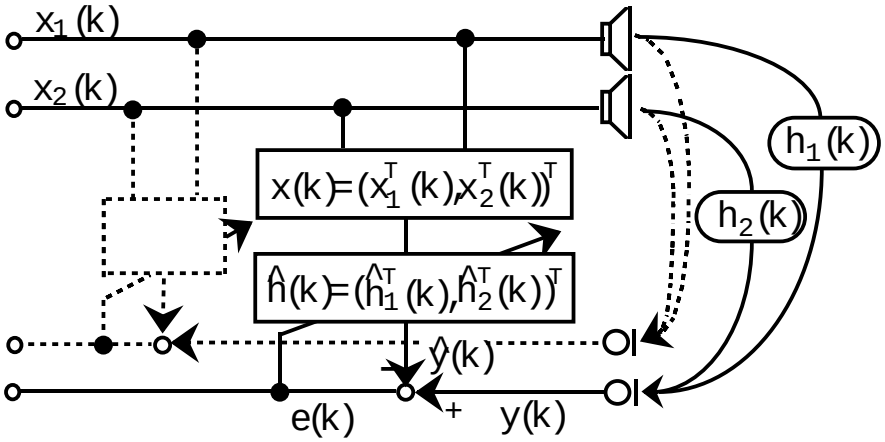


Figure 6: Stereo telecommunication system

Input signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ and filter coefficients $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$ are combined as $\mathbf{x}(k) = [\mathbf{x}_1^T(k), \mathbf{x}_2^T(k)]^T$ and $\hat{\mathbf{h}}(k) = [\hat{\mathbf{h}}_1^T(k), \hat{\mathbf{h}}_2^T(k)]^T$. Thus, stereo echo cancellation is achieved by linearly combining two monaural echo cancellers.

Unlike monaural echo cancellation, stereo echo cancellation has the specific problem of non-uniqueness. The problem is that the adaptive filter often misconverges or, even when it converges, its convergence speed is very slow because of crosscorrelation between the stereo signals [186]. As a result, the conventional

stereo echo canceller suffers from variation in both the near-end echo path and the far-end transmission path. Consequently, talker movement or changes in the transmission room are considered as variations in the echo path in the receiving room. Accordingly, the performance of the stereo echo canceller degrades at the instant of abrupt changes in the environment in the transmission room.

A clue to solving the non-uniqueness problem can be found in practical teleconferencing situations.

1. The stereo signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ contain independent noise.

2. The length of the adaptive filter $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$ is shorter than that of the impulse response in the transmission room. These truncated components act as independent noise.

3. The crosscorrelation between the stereo signals $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ varies slightly even when the talker does not move his body or head while speaking.

Several methods for overcoming this problem have been proposed.

Regarding items (1) and (2), several functions were successfully applied to create an independent component in the stereo input signals [171], [13]. The important point is that the noise generated should not be audible and should not degrade stereo perception.

Regarding item (3), one might think that the change in the variation causes another misconvergence, hence, it would not suppress the non-uniqueness problem. Fortunately, however, a "new" convergence process starts from the "old" misconverged solution. Thus, after many variations in the crosscorrelation, the stereo echo canceller can converge to the "true" solution [171].

To emphasize the slightly varying crosscorrelation between stereo signals in actual teleconferencing situations, the stereo projection algorithm [171], [12] and subband processing [133] were successfully introduced, and thus convergence speed to the true echo path impulse response was significantly improved.

4.5 Implementation

Echo cancellers are now commercial products and play a very important role in teleconferencing systems. To implement them in a single chip DSP at low cost, the subband processing mentioned above was successfully used [33], [45].

The most important technique in implementing echo cancellers in an actual teleconferencing system is double-talk control. An echo canceller has to distinguish between double-talk and an echo path change to determine whether to freeze or update the coefficients. If the echo canceller fails to realize effective double-talk control, the adaptive filter will be misadjusted, resulting in annoying echo and howling. The echo canceller must detect the double-talk situation before the adaptive filter becomes misadjusted. To solve this difficult problem,

two unique solutions have been successfully implemented. Both methods have the same approach of using one continually running adaptive filter to detect double-talk and distinguish it from an echo path change.

One solution is the so called "dual echo path model" which uses an adaptive filter and a fixed filter [147]. The fixed filter reduces the echo, and the adaptive filter estimates the room impulse response. The coefficients for the fixed filter are transferred from the adaptive filter when the adaptive filter converges. During double-talk, the coefficients of the adaptive filter may become misadjusted, increasing the error in the adaptive filter. If this occurs, the filter coefficients of the adaptive filter are not transferred to the fixed filter. The filter coefficients of the fixed filter are, therefore, not updated during double-talk, and the echo canceling level before double-talk is maintained. This method was brushed up and implemented in a commercial echo canceller and its good performance has been demonstrated [75].

The other solution is to use an "auxiliary" adaptive filter which adapts continually, but whose coefficients are not copied to any other filter [45].

While adaptive filters can achieve full-duplex communication, they are limited by their convergence speed and the echo canceling level. Therefore, other echo-suppression methods, such as variable loss insertions and center clipping, are used to enhance echo canceling.

The convergence of the continually working adaptive filter ensures single-talk. Hence, the loss controller can measure the acoustic coupling level, which is very important information of the "unknown system". Thus, the howling margin level is determined from the acoustic coupling level, and the necessary and minimum loss level can be decided.

4.6 Concluding remarks

Echo cancellers will become more and more important in digital networks, where the delay becomes longer and longer. The usage of echo cancellers will become more widespread in the future. To achieve comfortable speech communication, the total system should be studied considering noise reduction, the microphone array, de-reverberation, etc.

5 Signal Processing for Hearing Aids

James M. Kates
AudioLogic
4870 Sterling Drive
Boulder, CO 80301
jim@audiologic.com

5.1 Introduction

Approximately 7.5 percent of the population has some degree of hearing loss, and about 1.0 percent has a loss that is moderately severe or greater [151]. The majority of the hearing-impaired population would benefit significantly from improved methods of acoustic amplification. Hearing aids, however, are not as widely used as they might be; even within the population of hearing-aid users, there is widespread discontent with the quality of hearing-aid amplification [118]. One of the most common complaints is that speech is especially difficult to understand in noisy environments. In general, the signal-to-noise ratio (SNR) needed by a hearing-impaired person to give speech intelligibility in noise comparable to that for speech in quiet is substantially greater than the corresponding SNR required by a normal-hearing person [151].

While most commercial hearing aids are still based on analog signal processing strategies, much research involves digital signal processing. This research is motivated by the desire for improved algorithms, especially for dealing with the problem of understanding speech in noise. Cosmetic considerations, however, limit what can be actually implemented in a practical hearing aid. Most users of hearing aids want a device that is invisible to bystanders and thus does not advertise their impairment. As a result, the strongest pressure on manufacturers is to put simple processing into the smallest possible package, rather than develop sophisticated algorithms that require a larger package. Thus practical signal processing, as opposed to research systems, is constrained by the space available for the circuitry and the power available from a single small battery.20

5.2 Linear Amplification

The basic hearing-aid circuit is a linear amplifier, and the simplest hearing aid consists of a microphone, amplifier, and receiver (output transducer). In addition to being commonly prescribed on its own, the linear hearing aid also forms the fundamental building block for more advanced designs. Thus many of the problems associated with linear amplification will also affect other processing approaches when implemented in practical devices. Conversely, improvements in linear instruments will lead to improvements in all hearing aids.20

The signal processing in a linear hearing aid consists of frequency-response shaping and amplification. The limitations in typical hearing aids include restricted dynamic range, distortion, and restricted bandwidth. The dynamic range in a hearing aid is bounded by noise at low input signal levels and by amplifier saturation at high signal levels, with an available dynamic range of about 55 dB in a typical instrument. Thus the typical hearing aid has about half the dynamic range of a normal human ear. Distortion in hearing aids normally takes the form of symmetric peak clipping, and the restricted dynamic range can result in distortion for many everyday sounds, including monitoring the user's own voice. The high-frequency response of a typical hearing aid tends to fall off rapidly above about 5 kHz, which is less than the 8 kHz that is needed for optimal speech intelligibility [150] or music appreciation [115]. Increased

hearing-aid bandwidth, however, increases the possibility of distortion due to increased amplifier demand and can also increase the levels of mechanical and acoustic feedback in the hearing aid.

5.3 Feedback Cancellation

Mechanical and acoustic feedback limits the maximum gain that can be achieved in most hearing aids and also degrades the system frequency response. Most instruments have an acoustic feedback path that limits the maximum possible gain to about 40 dB or even less [107], depending on how the hearing aid is fit in the ear. Acoustic feedback problems are most severe at high frequencies since this is where a typical hearing aid has the highest gain.

The traditional procedure for increasing the stability of the hearing aid is to reduce the gain at high frequencies. Phase shifting and notch filters have also been tried [54], but have not proven to be very effective. A more effective technique is feedback cancellation, in which the feedback signal is estimated and subtracted from the microphone input. Simulations and digital prototypes of adaptive feedback cancellation systems [27, 108, 52, 61] indicate that increases in gain of between 6 and 17 dB can be achieved before the onset of instability with no loss of high-frequency response. The majority of feedback cancellation systems explored to date have used an injected noise probe signal with LMS adaptation [210] to adjust the filter modeling the feedback path. Field trials of a practical adaptive feedback-cancellation system built into a hearing aid have shown increases of 8-10 dB in the gain used by the subjects [15].

5.4 Compression Amplification

Compression amplification is used in hearing aids to prevent amplifier saturation and to match the dynamic range of the amplified signal to the reduced dynamic range of the impaired ear. Commercial products having up to three frequency bands have been designed using analog electronics, and digital research systems have been used to implement multi-band compression algorithms. The choice of optimum compression parameters to maximize speech intelligibility or speech quality is still open to contention. Rapid attack time constants (less than 5 ms) are used to prevent transients from saturating the circuitry. Arguments for fast release times (less than 20 ms), also termed syllabic compression, are based on considerations of the syllabic variations in speech and the desire to amplify soft speech sounds on a nearly instantaneous basis [204]. Arguments for long release times are based on the desire to preserve the envelope structure of speech [152].

Several multi-channel digital compression systems have been investigated. One objective of such systems is to place as much of the speech signal as possible within the residual hearing region of the impaired ear. Approaches reflecting this objective include modifications of the relative amplitudes of the principle components of the short-time spectrum [26] or the coefficients of a polynomial series fit to the spectrum [126]. A second objective is to match the loudness in

the impaired ear to that in a normal ear, and this approach has been investigated by [46] and [119]. The general result of these approaches has been some improvement in speech quality, but little improvement in speech intelligibility, especially in noise.

5.5 Noise Suppression

Improving speech intelligibility in noise has long been an objective in hearing-aid design. The single-microphone noise-suppression techniques that have been developed are based on the assumption that an improvement in SNR will yield a corresponding improvement in intelligibility, but this has not been found to be true in practice. Spectral subtraction attempts to estimate and remove the noise power; spectral subtraction in the frequency domain [17] and the pseudo-cepstral domain [206] have both been tried. Spectral enhancement attempts to emphasize the salient features of the speech spectrum and suppress noise. Approaches include comb filtering[127], Wiener filtering of the power spectrum [139], and maximum-likelihood envelope estimation [62]. While improvements in the measured SNR of up to 20 dB have been reported [200], no improvements in speech intelligibility have been observed.

Instead of trying to remove the noise, one can try instead to enhance the speech. The general approach that has been used is to increase the spectral contrast of the signal short-time spectrum by preserving or increasing the amplitude of frequency regions containing spectral peaks while reducing the amplitude of regions containing valleys. Techniques include squaring and then normalizing the spectral magnitude [16], increasing the spectral magnitude in pre-selected spectral regions while reducing it in others [25], filtering the spectral envelope to increase the higher rates of fluctuation [173, 191], and using sinusoidal modeling of the speech to remove the less-intense spectral components while preserving the peaks[110]. In general, spectral enhancement has not yielded any substantive improvement in speech intelligibility.

5.6 Microphone arrays

In many situations, the desired signal comes from a single well-defined source, such as a person seated across a table, while the noise is generated by a large number of sources located throughout the area, such as other diners in a restaurant. Under these conditions the speech and the noise tend to have similar spectral distributions, but the spatial distributions differ. The spatial separation of the speech and the noise can be exploited to reduce the noise level without any deleterious effects on the speech. Furthermore, unlike the situation for single-microphone noise-suppression techniques, the improvements in SNR measured with directional microphones and microphone arrays give corresponding improvements in speech intelligibility.

A directional microphone will improve the SNR by maintaining high gain in the direction of the desired source and reduced gain for sources coming

from other directions. Greater improvements in the SNR and speech intelligibility require arrays that combine the outputs of several microphones. The simplest multi-microphone processing approach is delay-and-sum beamforming. The benefit of delay-and-sum microphone arrays of the sort that can be built into an eyeglass frame, for example, is an improvement of 5-10 dB in SNR, with the greatest improvement at higher frequencies [183, 184]. The performance of delay-and-sum beamforming can be bettered by using superdirective array processing [39] to give the optimum improvement in SNR for a stationary noise field. Simulation studies for a spherically isotropic noise field [188, 109], show that a superdirective array will give a SNR about 5 dB better than that obtained for delay-and-sum beamforming using the same set of microphones, and this relative performance benefit is also found in measurements in reverberant rooms [111].

Adaptive array systems have the potential of giving even greater improvements in SNR than those resulting from an array using fixed weights. Microphone arrays using constrained LMS adaptation [40] have proven effective in simulations [73, 79]. A processing comparison of fixed and adaptive arrays [111] showed that the adaptive system was somewhat more effective than the superdirective system when the noise power exceeded the speech power, and somewhat less effective than the superdirective system when the noise level in the room was below that of the speech.

5.7 Conclusions

Different hearing-aid signal-processing strategies have been developed based on different assumptions as to what is most important in dealing with the hearing loss. Linear amplification is intended to overcome the loss in auditory sensitivity for normal conversational speech levels. Wide dynamic-range compression amplification is intended to compensate for the reduced dynamic range in the impaired ear. Noise suppression algorithms are intended to compensate for the loss of frequency and temporal resolution in the impaired ear. The complexities of the auditory system and the nature of auditory impairment may well mean that total compensation for hearing loss is not possible in a conventional hearing aid. Improved signal processing for hearing aids is thus a deceptively difficult engineering problem.

6 Overview of Perceptual Audio Coding

Marina Bosi
Digital Theater Systems (DTS)
Los Angeles, California, U.S.A.
mab@dtstech.com

6.1 Introduction

The challenge of audio coding today is to minimize costs while providing an audio signal with no audible differences from a signal encoded according to the compact disc (CD) audio standard. According to the CD audio specifications, the audio signal is a two-channel signal, sampled at 44.1 kHz with each sample PCM-coded with 16 bits. Since the advent of the CD technology, the expectation for any good audio quality system has become that it provides a CD quality signal with 20 kHz (or higher) of bandwidth and a dynamic range equal or above 90 dB. Furthermore, widespread exposure to CD quality audio is leading to consumer demand that most consumer products and broadcast applications provide CD-like audio quality.

Typically a higher coding efficiency implies an increased complexity and delay of the overall system. Being able to deliver CD-quality audio signals at a reduced data rate can translate into cost savings if the increase in complexity and delay don't add too much to the costs. The appropriate balance will depend on the application. Applications like digital broadcasting, audio transmission via ISDN lines, and audio transmission via Internet, require high coding efficiency. For point to multi-point applications it is desirable to keep the decoders costs, in terms of computational complexity and memory storage, low. As low cost RAM and processing power become more and more available, decoder costs will become less of an issue in the design of audio coding systems. Issues like error resilience, editability of the encode bit stream, however, remain important issues to be addressed in design of audio coding systems.

Applications with more audio channels than the two-channel CD specification, are becoming more and more popular. The most typical configuration, the 5.1 channel configuration, includes five full-bandwidth channels plus a low frequency channel (¡ 200 Hz) for additional low frequencies support[88]. The film industry has been using multichannel audio formats for many years, switching from analog to digital formats over past five years or so. Similarly, high definition television (HDTV) in North America[35] and the newly defined digital versatile disc standard, DVD-Video[51] embraced the 5.1-channel configuration in their audio specifications. In the case of the DVD-Audio requirements higher sampling frequencies (e.g. 96 kHz) and increased sample resolution (e.g. 24 bits) are also being considered.

While speech and video coding have been well-established research topics for many decades, high-quality audio coding is a relatively new subject. First applications of audio coding, including CD-I, made use of ADPCM techniques. Typical compression ratios for this technology applied to music signals is about 4:1. To reach high compression ratios (above 4:1) for general audio signals while keeping high fidelity, perceptual coding principles are typically employed in audio coding. One of the first applications of psychoacoustics principles to audio coding can be found in[120]. In this publication, a sub-band coder based on a tree-structure quadrature mirror filter (QMF) and a fixed bit allocation is described. It should be noted that previous work in speech coding also took advantage of psychoacoustics principles and adaptive transforms[166, 212, 93]

and laid the ground for the research work in perceptual audio coding.

With the standardization efforts of ISO/IEC JTC1/ SC 29/WG 11 (MPEG), audio coding has become widely spread in a variety of applications. CD-like quality at data rates of 128 kb/s per channel (6:1 compression ratio) has been reached for a number of coding schemes including MPEG-1 Layers II and III, and AC-2 [102, 104, 103, 85]. A new standard of high quality audio coding, MPEG-2 Advanced Audio Coding, is capable of providing indistinguishable audio quality at data rates of 320 kb/s per 5.1 channels [100], or equivalently a compression ratio of 12:1.

6.2 Audio Coding Standards

The standardization body ISO/IEC JTC1/SC29/WG11, also known as the Moving Pictures Expert Group (MPEG), was established in 1988 to specify digital video and audio coding schemes at low data rates. MPEG completed its first phase of specifications (MPEG-1) in November 1992 [2]. The MPEG-1 audio coding system, specified in ISO/IEC 11172-3 [see also 19] operates in single channel or two-channel stereo modes at sampling frequencies of 32 kHz, 44.1 kHz, and 48 kHz. Three Layers are specified. MPEG-1 Layers I and II make use of 32 subbands, 511-tap polyphase quadrature mirror filter, PQMF (see also section 6.3). A psychoacoustics model determines the adaptive bit allocation which controls the quantization stage. Block floating point is used and the coded data are segmented into fixed frames. Longer frames (24 ms compared to 8 ms in Layer I) are used in Layer II, which also provides additional coding of bit allocation, scale factors, and samples. MPEG-1 Layer I provides high quality at data rates of 192 kb/s per channel, while Layer II provides high quality at data rates of 128 kb/s per channel[102, 104, 103]. MPEG-1 Layer III combines some of the features of Layer II with aspects of ASPEC[21]. A hybrid filter bank which cascades the 32-band PQMF with an 18-point modified cosine transform [154] (see also 6.3), for a total of 576 frequency lines, is employed in Layer III. In addition to the high frequency resolution filter bank, Layer III utilizes entropy coding in conjunction with non-uniform quantization and analysis-by- synthesis control of the quantization error levels. Targeted data rates for Layer III are below 128 kb/s per channel.

In its second phase of development, MPEG's goals were to define a multichannel extension to MPEG-1 audio that would be backwards compatible with existing MPEG-1 systems (MPEG-2 BC) and to define and an audio coding standard at lower sampling frequencies (16 kHz, 22.5 kHz, 24 kHz) than MPEG-1, MPEG-2 LSF. Both MPEG-2 BC and MPEG-2 LSF were completed in November 1994[82]. Started in 1994, another effort of the MPEG-2 audio standardization committee aimed to define a higher quality multichannel standard than achievable while requiring MPEG-1 backwards compatibility, the so called MPEG-2 non-backwards compatible audio standard, later renamed MPEG-2 Advanced Audio Coding (MPEG-2 AAC)[83]. The MPEG-2 AAC standard was finalized in April 1997. While MPEG-2 BC provides good audio quality at data rates of 640-896 kb/s[99] for five full-bandwidth channels,

MPEG-2 AAC provides very good audio quality at less than half that data rate. Tests carried out in the fall of 1996 at BBC, UK, and NHK, Japan, showed that MPEG-2 AAC satisfies the ITU-R quality requirements[87] at 320 kb/s per five full-bandwidth channels (or lower according to the NHK data)[100]. AAC combines the coding efficiency of a high resolution filter bank (a 1024-point modified cosine transform), temporal noise shaping (see also section 6.3), prediction techniques, intensity and Mid/Side stereo coding, and noiseless coding to achieve very high quality audio at very low data rates (see also Figure 7). It operates at sampling frequencies between 8 to 96 kHz and supports up to 48 audio channels.

Insert `mab:fig5` here!!!

Figure 7: Example of the Encoder Process: MPEG-2 AAC Encoder

MPEG-2 AAC will constitute the kernel of the forthcoming MPEG-4 audio standard at data rates at or above 16 kb/s per channel. The MPEG-4 Audio standard will be finalized in January 1999[101]. In MPEG-4 audio, three different coder types are integrated in a unified framework: coders based on time to frequency mapping, T/F coders like AAC, CELP coders, and parametric coders, PARA coders. In the PARA coders scheme, parameters that describe the input samples sinusoidal and noise components are extracted. The coding of these parameters is controlled by a perceptual model in order to reduce the data rate. The motivation for the integration of different coding schemes is that for different types of signals and at different bit rates one type or a combination of coders performs better than others. For example, CELP coders perform better for speech while PARA coders perform better for music signals for data rates starting at 2 kb/s per channel. T/F coders are employed to achieve very high quality audio at higher data rates. In addition, the advantages of sophisticated source models of CELP coders can be combined with the advantages of the T/F scheme and its psychoacoustics model. A new set of functionalities is introduced in MPEG-4 audio. These functionalities include content and complexity scalability and pitch/time manipulation. Additional flexibility and interactivity is possible if synthetic audio signals are also taken into consideration. The MPEG-4 synthetic/natural hybrid coding, SNHC, activities contain two major contributions concerning audio. The first is a text to speech, TST, system which specifies a syntax and a decoder for the combination of text, auxiliary information, and control commands. This allows the integration of TTS synthesizers for different languages and interfaces with other modules like face animation etc. The second contribution is "structured audio", which exploits methods for defining structural similarity in audio signals, e.g. repetition of sound sequences, timbre, etc., to achieve ultra-low data rates by using model-based assumptions to represent the signals.

6.3 Background

If we consider the basic audio coding chain (see Figure 8), the input signal is coded then transmitted and/or stored. After the decoding process, the last stage in this process is the reception of the audio signal by the human ear. The basic building blocks of an audio encode/decoder (see Figure 9), contain two crucial pieces: one is transforming the signal into a representation which concisely models the source and the other is the psychoacoustics model which provides an approximation of the perception mechanisms of the human ear. These two pieces will determine the amount of reduction in the data rate that can be achieved by the coding process.

Insert `mab:fig1` here!!!

Figure 8: Basic Audio Coding Chain

Insert `mab:fig2` here!!!

Figure 9: Basic Audio Encoder (a) and Decoder (b)

In the signal representation stage, the removal of signal redundancies takes place. In speech coding, a physical model of the vocal tract is employed to define speech parameters. These parameters together with residual information are then encoded. While this technique allows for very high compression ratios, it is not very successful with music signals since it is very difficult to accurately model all possible music sources. In audio coding, typically the time representation of the signal is mapped to a time-frequency representation via a filter bank. In this case, the frequency domain filter bank output provides the primary signal representation. The general assumption is that audio signals are quasi stationary, therefore a mapping to the frequency domain results in a signal representation which is more efficient than straight PCM. The longer the analysis window the better founded is this assumption. In general, there is a tradeoff between the high coding efficiency granted by a high-resolution filter bank employed in the signal representation stage and memory cost/delay of the overall coding system. Investigations on spectral resolution showed that a good choice is a frequency resolution around 20 Hz corresponding to a time resolution of around 25 ms [98]. Time-variant filter banks are often used in order

to avoid the spreading of quantization noise in time in the reconstructed signal [53, 192, 42]. These methods allow the filter bank to adapt to a higher time resolution, typically equal or less than 5 ms, in the presence of a transient.

In addition to the removal of redundancies, the perceptual irrelevant portion of the signal can also be removed. In other words, it is assumed that quantization error can be hidden below a signal-dependent masking curve which is based on the ability of the human ear to perceive various signals (see Figure 10).

Insert `mab:fig3` here!!!

Figure 10: Frequency Masking

The quantization levels for each spectral component of the signal are determined by the targeted data rate and the masking thresholds derived for the current signal. Masking of a softer signal (maskee) by a louder one (masker) can occur when masker and maskee are presented simultaneously (frequency masking) and/or before (pre-masking) or after (post-masking) the presentation of the masker (temporal masking). A simple frequency masking model for a sinusoidal masker of level Lm is shown in Figure 11.

Insert `mab:fig4` here!!!

Figure 11: Simple Masking Model

Note that in the horizontal axis the frequency is expressed in Bark units as per[213]. A Bark scale represents a non-linear frequency scale. A Bark unit corresponds to a constant distance on the cochlea and represents a frequency interval within which the combination of the maskees is masked by the masker. The quantization error level localized in the gray area of Figure 11 is assumed not to be audible. The simple masking model can be described by three parameters:

- The difference between the level of the masker and the masked threshold, Δ.
- The masking curve slope, $slope1$, toward lower frequencies. According to [213] a good estimate of slope1 is about 27 dB.

- The masking curve slope, $slope2$, towards higher frequencies as a function of the masker level, Lm

$$slope2(Lm) \approx -27dB/Bark + max(Lm - 40, 0) * 0.37$$

In addition to the masking effects due to another masking signal, one should also take into consideration the masking threshold in quiet or threshold of hearing (see Figure 10).

6.4 Basic Audio Coding Tools

The basic audio encoding process can be described as follows (see Figure 7). First, a filter bank is used to decompose the input signal into subsampled spectral components (time-frequency domain samples). The filter bank can be implemented via a PQMF[161] . A 32-band PQMF is employed In MPEG-1 and MPEG-2 Layers I and II. For higher frequency resolution, a critically-sampled filter based on time-domain aliasing cancellation TDAC[154] can be used. This technique allows for a critically sampled filter bank and introduces a phase modification to a series of discrete cosine transforms (oddly stacked TDAC) or a series of alternate discrete cosine and sine transforms (evenly stacked TDAC). MPEG-2 Advanced Audio Coding [19] and PAC[95] use a modified cosine transform, MDCT, with 1024 frequency bands, while AC-3[66] employs a 256 frequency lines MDCT. A hybrid filter in which the PQMF stage is cascaded with an MDCT stage is found in MPEG-1 and MPEG-2 Layer III[23], and ATRAC[197]. Wavelet-based filter banks[203, 174] can also be found in literature and are employed in the enhanced PAC coder. While the PQMF filter utilized in MPEG Layers I and II is time-invariant, all the other coding schemes mentioned above make use of adaptive filter banks.

Based on the input signal, an estimate of the current masking curve is computed using rules known from psychoacoustics. Either the output of the main data path filter bank (see for example PAC[95], AC-3[66]) or a side chain analysis filter, typically an FFT as in the MPEG psychoacoustic models[2], is used for this estimate. Often, in addition to the masker level and its frequency, the noise-like versus harmonic characteristics are taken into consideration, keeping in mind that noise-like signals are *better maskers* than tones. A *signal to mask ratio*, i.e. an assessment of how much quantization noise can be masked by the input signal, is derived from the masking curve. This information is utilized in the quantization stage to minimize the audible distortion of the quantized signal at any given data rate.

Sometimes to further reduce the redundancy in the audio signal an intra-channel, time-domain backward or forward adaptive prediction tool is employed in order to take advantage of correlation between subsampled spectral components of subsequent frames. In the case of MPEG-2 AAC, a second order, backwards adaptive set of predictors is applied to frequency coefficients up to 16 kHz[70]. Another method to increase the coding efficiency is to normalize the signal spectrum by using an LPC analysis stage after the filter bank.

This method is used in the Twin-VQ coding scheme[89] currently proposed for MPEG-4 audio coding.

A novel concept in audio coding is to apply predictive coding to the spectral data in order to enhance the overall time resolution of the system[78]. The temporal noise shaping, TNS, tool performs an in-place filtering operation on the spectral values, i.e. replaces the target spectral coefficients with the prediction residual. The TNS technique permits the encoder to control the temporal structure of the quantization noise even within a filter-bank window.

The main data reduction process takes place in the quantization and coding stage of the encoder. With the exception of Layer III and MPEG-2 AAC which employ non-uniform quantization with a power law, the majority of the audio coding systems apply uniform quantization. Vector quantization has also been applied in audio coding[89]. A bit allocation procedure in conjunction with block floating point coding is sometimes used [66]. The bit allocation procedure, driven either by data statistics or by the perceptual model output, assigns the number of bits for each spectral components to be used in the quantization stage [66]. In some audio coding schemes [19, 95, 23] a noise allocation procedure is adopted instead. In this case no explicit bit allocation is performed, rather the amount of quantization noise allowed by the estimated masking curve is used to adapt the quantization step size for each frequency band. This process is also known as analysis-by-synthesis quantization, and is generally controlled by nested iteration loops. The iteration loops typically include entropy coding of the quantized spectral values. A count of how many bits per spectral component can be done only upon the completion of the iteration loops. In addition, in order to satisfy the signal demands on a frame-by-frame basis, a mechanism called *bit reservoir* which allows for a locally variable data rate is sometimes adopted [19, 95, 23].

For multi-channel signals, joint stereo coding is applied in order to remove stereo redundancy and stereo irrelevance in multichannel audio signal. M/S (Mid/Side) stereo coding and intensity stereo coding are the techniques most commonly found in multichannel audio coding schemes. Instead of transmitting the left and right signal, the normalized sum (M as in Mid) and difference signals (S as in Side) are transmitted in M/S stereo coding. This matrixing operation can be done in the time or frequency domain. This technique allows for redundancy removal for mono-like signals. Stereo irrelevancy effects can also be exploited by using M/S stereo coding. In intensity stereo coding only the energy envelope of the signal is transmitted. Intensity stereo coding allows for a reduction in the signal spatial information transmitted. A coupling channel is sometimes used as the equivalent of an n-channel intensity stereo coding system. Stereo coding is a powerful method generally used at very low data rates. To enable a smooth transition from two- channel stereo to multichannel, many of the audio coding systems provide downmix matrix capability from 5.1 channels to two channels.

The final step in the encoding process is the coded bit stream packing. A bit stream formatter is used to assemble the bit stream, which consists of the quantized and coded spectral coefficients and control parameters.

6.5 Codecs Assessment

The goal of audio coding is to reduce the data rate of the overall system while maximizing the quality at a given data rate and keeping the complexity as low as possible. The audio quality of an audio system is classically measured in terms of signal-to-noise ratio (SNR), total harmonic distortion (THD), and total harmonic distortion plus noise (THD+N). These measurement techniques are obsolete when we try to apply them to the measurement of quality for perceptual audio coding systems. A clear example of the inadequacy of these techniques is the *13 dB miracle* paradox developed by Johnston and Brandenburg in 1990. This demonstration showed the dramatic difference that can be achieved by carefully crafting where a given quantity of noise (corresponding to a signal to noise ratio of 13 dB) appears in the original signal spectrum.

A number of objective perceptual measurement systems [22, 34, 8, 149] have been trying to address the need of objective measurement of audio coding schemes. These systems provide an estimate of the difference in the perceptual domain between the reference signal and the output of the codec under test. Some correlation can be established between the results of listening tests and objective perceptual measurement of audio quality[97]. The development of these techniques, however, is not at the stage were they can be considered a substitute to formal listening tests.

The basic assumption made when assessing audio coders via subjective listening tests is that the audio system generates a signal with only small impairments compared to the original (or reference) signal. The work of ITU-R TG-10/3 resulted in the specifications of a formal test method for the subjective assessment of small impairments in audio systems including multichannel systems, ITU-R recommendation BS.1116 [86]. Only expert, reliable listeners, i.e. pre and post screenings of the listeners are performed in order to qualify their ability to make correct identifications, are selected to participate in the tests. The double- blind, triple-stimulus with hidden reference method is adopted. The ITU-R 5-point impairment scale or the equivalently defined difference in grade[86] (see Table 1) is also adopted. Critical material which stresses the systems under test is selected by an expert panel. Room size, shape, proportions, reverberation time, background noise levels as well as the reproduction system characteristics are specified. Carefully designed listening tests are typically reliable and stable. Parametric analysis like ANOVA methods (analysis of variances) are commonly applied to difference grades, i.e. the difference between the system under test grade and the hidden reference grade. The outcome of the data analysis provide a measure of the average performance of the systems under test and the reliability of the differences among the systems under tests. The confidence interval gives an indication of the resolution achieved by the test; a confidence interval of 95% is often used.

6.6 Future and Trends

State-of-the-art audio codecs are capable nowadays of providing CD-like audio quality at data rates of approximately 64 kb/s per channel (12:1 reduction in data rate) which was thought unreachable in "real life" systems only a few years ago. Has the development of audio coding reached its limits in terms of maximum compression or is there still room for improvement? Certainly, from the theoretical point of view, considering the resolution of the human auditory perception and the amount of audio information that we can transmit per second, there is hope that lower data rates at very high quality can be achieved. Improvements in the audio signal modeling and representation, improvements in the auditory system perceptual models, and a better understanding of the interaction between these two fundamental stages in audio coding may lead us to the next breakthrough in this relatively young research area. Results from research centers around the world seem to bolster this idea. Work on filter banks [203, 134], including signal adaptive filterbanks [155, 174], oversampled filter banks, and low delay filter banks [168] is very promising in terms of audio signal representation. Results from MPEG-4 work seem also very encouraging in the use of physical modeling and structural modeling for audio representation[101]. Regarding perceptual modeling, models for additive masking [7], models for neural processing, and in general improved auditory system insight and depiction can be brought to fruition in audio coding.

Some may argue that, with increasing media capacity, no compression at all or the limited compression available using lossless techniques is the answer for up-coming consumer technology including DVD. Although there have been dramatic increases in media capacity, there is little doubt that new applications and content will emerge that continue to push the available limits, again requiring efficient use of media real estate to fit all of the content demanded. Most of us only need look back at our own experiences with PC hard drive space to feel very certain that what currently feels like unimaginably large capacity may soon feel constraining.

7 Sound Synthesis Based on Physical Models

Julius O. Smith III
Center for Computer Research in Music and Acoustics (CCRMA)
Department of Music
Stanford University
Stanford, California 94305
jos@ccrma.stanford.edu

7.1 Sound Synthesis Based on Physical Models

In this section, we review some activities related to sound synthesis based on physical models of musical instruments. To summarize, much past effort has

been devoted to the problem of *speech synthesis* using a variety of signal models including some based on physical descriptions, and this provided a fertile background for later research in *musical* sound synthesis. At the present time, advanced music synthesizers are appearing which are based on computational *physical models of musical instruments*, and "physical-modeling synthesis" is just beginning to supplement FM and wavetable synthesis in MIDI synthesizers on certain multimedia sound cards and even Pentium-based software synthesizers. Also, high quality *voice synthesis* from a physical model has been demonstrated in the lab for "generic" voices (i.e., not sounding like a particular speaker). For the future, sound synthesis technology has been proposed for inclusion in the MPEG-4 *audio coding* standard, and physical-modeling synthesis is one of the supported methods.

7.1.1 The Past

It has been recognized for decades that model-based compression stands to provide the highest possible compression ratios in speech coding [167, 68]. For example, very low bit-rate speech coders have been based on "articulatory speech synthesis" in which a model of the voice-producing apparatus is "remote controlled" by a bit stream specifying vocal-tract shape and articulator motion [112, 43].

Like articulatory models of speech, physical models of musical instruments have historically been too expensive computationally to be suitable as a basis for sound coding or even sound synthesis in a commercial music synthesizer. For a representative example of computational models in musical acoustics, see [31].

More recently, highly efficient "digital waveguide" models [181, 176] of musical instruments have been developed which are closely related to the acoustic-tube model for speech production [114, 135, 36, 37, 178]. These models achieve high efficiency gains over more standard "finite-difference" models by applying principles of linear systems theory such as *commutativity* of linear, time-invariant elements in order to greatly reduce computational requirements at a given audio quality level. The "lumping" of losses into a single "per-period" filter yields orders of magnitude efficiency gains over finite difference methods; as a specific example, the SynthBuilder [153] virtual electric guitar patch implements 6 "steel" strings, a stereo flanger, amplifier distortion and feedback, and on-chip vibrato, in real time at a 22 kHz sampling rate, on a single Motorola DSP56001 DSP chip, clocked at 25 MHz, with 8 kWords of zero-wait-state SRAM. For an overview of recent research related to digital waveguide modeling, see [182].

To provide a specific example, the digital waveguide clarinet is shown in Fig. 7.1.1 [177]. If the bore is cylindrical, as it is in the clarinet, it can be modeled quite simply using a bidirectional delay line. If the bore is conical, as in a saxophone, it can still be modeled as a bidirectional delay line, but interfacing to it is slightly more complex, especially at the mouthpiece [164]. Because the main control variable for the instrument is air pressure in the mouth at the reed, it is convenient to choose pressure wave variables. Hence, the delay-lines

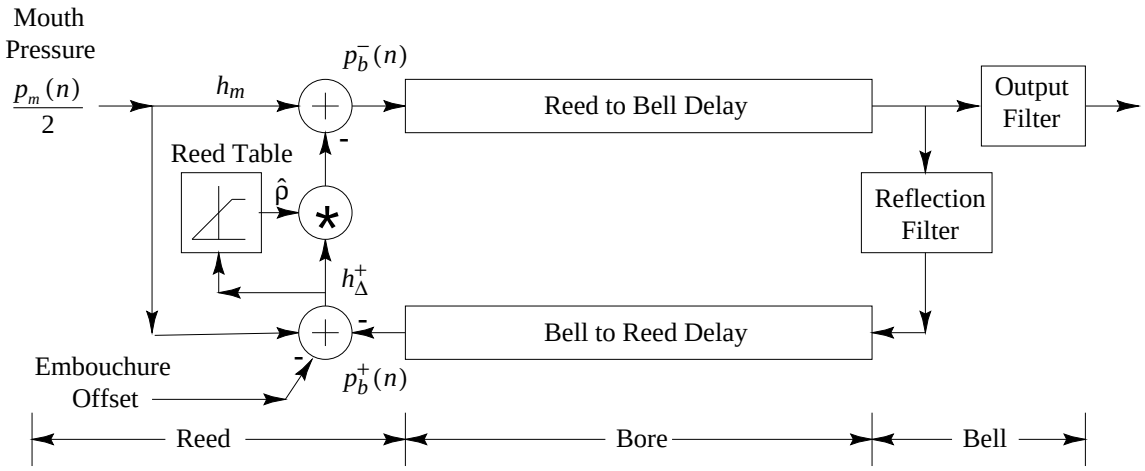


Figure 12: Waveguide model of a single-reed, cylindrical-bore woodwind, such as a clarinet.

carry left-going and right-going *pressure* samples p_b^+ and p_b^- (respectively) which sample the traveling pressure-wave components within the bore.

The reflection filter at the right implements the bell or tone-hole losses as well as the round-trip attenuation losses from traveling back and forth in the bore. To first order, the bell passes high frequencies and reflects low frequencies, where "high" and "low" frequencies are divided by the wavelength which equals the bell's diameter. For a clarinet bore, the nominal "cross-over frequency" is around 1500 Hz [10]. Thus, the bell output filter is highpass, and power complementary with respect to the bell-reflection lowpass filter.

Tone holes can also be treated as simple cross-over networks. However, it is more accurate to utilize measurements of tone-hole acoustics in the musical acoustics literature and convert the "transmission matrix" description (often used in acoustics) to the traveling-wave formulation by a simple linear transformation. For typical fingerings, the first few open tone holes jointly provide a bore termination [10]. Either the individual tone holes can be modeled as (interpolated) scattering junctions, or the whole ensemble of terminating tone holes can be modeled in aggregate using a single reflection and transmission filter, like the bell model. Digital waveguide models of tone holes are elaborated further in [164, 198]. For simple practical implementations, the bell model can be used unchanged for all tunings, as if the bore were being cut to a new length for each note and the same bell were attached.

The reed mouthpiece is controlled primarily by *mouth pressure* p_m. The reed is modeled as a signal- and embouchure-dependent *nonlinear reflection coefficient* terminating the bore. Such a model is possible because the reed mass is neglected [140]. The player's embouchure controls damping of the reed, reed aperture width, and other parameters, and these can be implemented as param-

eters on the contents of the lookup table or nonlinear function, thus modifying the reed's *reflection-coefficient* $\rho(h_\Delta^+)$, where $h_\Delta^+ \triangleq p_b^-/2 - p_b^+$. A simple choice of embouchure control is an offset in the reed-table address. Since the main feature of the reed table is the pressure-drop where the reed begins to open, a simple embouchure offset can implement the effect of biting harder or softer on the reed, or changing the reed stiffness. A qualitatively chosen reed table is shown in Fig. 7.1.1. Brighter tones are obtained by increasing the curvature of the function as the reed begins to open.

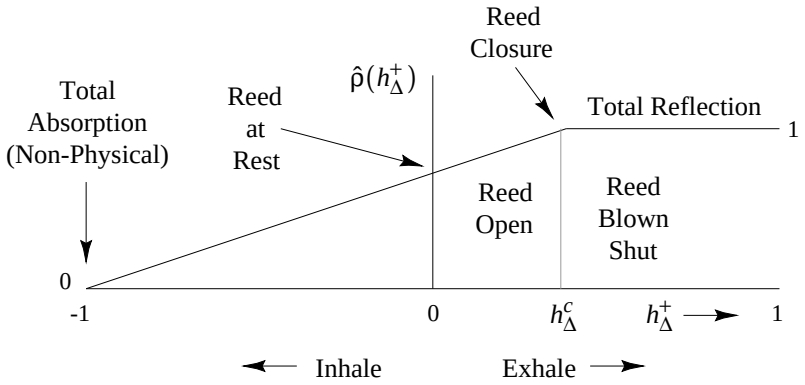


Figure 13: Simple, qualitatively chosen reed table for the digital waveguide clarinet.

The input mouth pressure is usually summed with a small amount of white noise, corresponding to turbulence. For example, 0.1% is generally used as a minimum, and larger amounts are appropriate during the attack of a note. The turbulence level may also be computed automatically as a function of pressure drop p_Δ and reed opening geometry [69, 202].

Referring again to Fig. 7.1.1, the reflection filter can be as simple as a one-pole lowpass filter with transfer function

$$H(z) = \frac{1 + a_1}{1 + a_1 z^{-1}} \tag{1}$$

where, for example, $a_1 = -0.642$. Further loop filtering occurs as a result of using low-order delay-line interpolation. Often simple linear interpolation is used, but there are better choices [122]. (There need be only one delay line in a practical implementation since the lower delay line of Fig. 7.1.1 can be commuted with the reflection filter and combined with the upper delay line, ignoring the path to the output filter since a pure delay of less than a period in the final output sound is inconsequential.) The output filtering can be accomplished by simply adding together the input and output of the (inverting) lowpass filter. However, the output filtering is often left out altogether, since it attenuates the bass response, and such an unfiltered output corresponds loosely to an "internal

microphone" inside the bore at the bell. Finally, "legato" note transitions may be managed using two delay-line taps and cross-fading from one to the other during a transition [92, 182].

7.1.2 The Present

Parametric sound synthesis based on physical models of musical instruments is presently an emerging technology in commercial music synthesizers. The Yamaha VL1, introduced in 1994, appears to be the first commercial synthesizer based on digital waveguide models. Korg introduced a related product in 1995. Since then, simplified software-only implementations have appeared, and there is work under way at a number of companies aimed at delivering this kind of sound synthesis to desktop PCs in the near future. Performing electronic musicians are very interested in these methods because they can provide a complete set of intuitive parameters that can be "performed" in real time to give compelling and expressive musical performances.

Model-based audio rendering is also one of several synthesis methods in the *structured audio* proposal for MPEG-4 digital audio compression [201]. Physical models can be very good at representing specific classes of sound, but they tend to lack generality across all sounds. As a result, when used for audio compression, specific models need to be combined with at least one fully general audio coding method, such as the perceptually based transform coders used in MPEG-2 [24, 20].

7.1.3 The Future

The kinds of models needed for model-based multimedia audio compression in the future include voice, musical instruments, and sound effects (particularly for video games). In addition to models based on physics, there are more abstract *signal models*, such as the source-filter paradigm [30] and psychoacoustically motivated parametrizations [163, 189, 201]. In fact, every sound synthesis technique can be considered to define a kind of signal model. In all cases, the goal is to be able to generate a wide variety of useful sounds from a small set of control parameters. For a review of different kinds of synthesis techniques and their characteristics, see [160, 201, 180, 18].

Since model-based compression requires the ability to transmit the models themselves in advance of the encoded bit stream, and since no single model can efficiently handle all sounds, future audio compression standards such as MPEG-4 must support *arbitrary synthesis models*, without giving up *decoding efficiency*, and without compromising *software security*. An excellent solution to these conflicting desires is provided by the *unit-generator* paradigm for sound synthesis [138, 180]. Unit generators can be thought of as subroutine calls with names like "oscillator", "envelope", "noise", "filter", "mixer", and so on, in a standardized signal-processing library. The unit generators must (1) execute efficiently on the local host (decoder), (2) be combinable to implement any sound synthesis algorithm, and (3) not be configurable to have side effects outside of their signal

outputs (for software security). If there is a secure, locally compilable language, such as Java, it will also be possible to support new unit-generator definitions in the coded bit stream without loss of security.

7.2 Conclusions

"Digital waveguide synthesis" was briefly reviewed as an example of emerging practice in sound synthesis based on physical models. It can be considered a descendent of speech models using a sampled traveling-wave representation of the vocal tract [114] as well as time-domain models from the musical acoustics literature [140]. At the present time, some very fine synthesizer voices exist based on this new approach [182]. In the future, we can expect further refinements of the models, producing better and better "virtual acoustic" instruments, including the human voice, as well as completely new instruments such as "bowed pipes" (already supported in the VL synthesizers). As these models proliferate and improve, they will provide both excellent "virtual instruments" for the performing electronic musician, as well as compact sound-generation specifications for use in compressed audio formats for multimedia.

8 Professional and Consumer Gear: Hardware & Software

Mark Kahrs
Multimedia Group
CAIP Center
Rutgers University
08855-1390
kahrs@caip.rutgers.edu

The story of professional and consumer gear for audio is both long and glorious; in this short review, we can only skim the highlights of the past 50 years with an eye towards the use of DSP technolgy in the future improvements. Recent developments in signal processing have been made possible by developments in digital technology. In considering the advance of computer technology, one must remember it is not only the advance in computing but also advances in storage that make our present audio equipment possible.

8.1 The Past

In 1947, vinyl recordings were destructively readout and amplified by high voltage electronic devices encased in glass. The recent development of oxidized tape as a recording medium would eventually revolutionize professional recording. Furthermore, the exploitation of the newly invented transistor would dramatically reduce the size of amplifiers and analog signal processing hardware.

The 1950s saw the increasing use of the transistor in both professional and consumer audio electronics. The reduction in power consumption as well as heat production made the first portable transistor radio possible. Les Paul experimented with multiple track recording in the mid-1950s. 1957 saw the introduction of commercial stereo recordings (although Blumlein proposed them in 1931). Stereo FM broadcasting was introduced in 1961[55].

In the early 1960s, Thiele and later, his student Small, building on the large body of work on loudspeaker modelling finally put enclosure design on a firm theoretical foundation[38]. This effectively eliminated the "cut-and-try" method of loudspeaker enclosure design. The latter 1960s also saw the modern introduction of the Fast Fourier Transform and other DSP techniques including the design and implementation of digital filtering. 1965 also marked the introduction of "Dolby A" noise reduction followed by the introduction of the compact cassette. Robert Moog also introduced his first analog music synthesizer.

But it remained for the computer revolution to bring DSP to audio. The first digital synthesizers were constructed from discrete SSI TTL level components. Beginning in the early 1970s, analog/digital and digital/analog converters were finally reaching toward 16 bits of dynamic range. DSP could be done, albeit not often in real time. The approaching MSI era would make real-time DSP processors possible, but only if you had the right amount of power and air conditioning. However, this was not an obstacle to audio processing: in the mid-1970s, Stockham showed how DSP could be used to process historical recordings of Caruso[190].

A review of DSP architecture circa 1975 was given by Allen[4]. Further development of DSP processors was critically dependent on the technology of multiplier implementation. The introduction of the TRW 16 by 16 multiplier (MPY16) in 1979 was critical to the development of many real-time DSP machines.

In addition, the increasing miniaturization of electronics resulted in the now famous Sony "Walkman", which brought two track tape technology together with the integrated circuit to revolutionize the notion of portable personal electronic audio devices.

The joint Philips-Sony Compact Disk standard (circa 1981) produced another boom for the audio industry as consumers relegated their vinyl disks and players to back rooms and attics and professionals retooled to provide digital recordings. Initally, all CDs were mastered onto video tape.

In 1981, the introduction of the Texas Instruments TMS320C10 demonstrated that a DSP could be commercially produced. (the earlier DSPs from Western Electric were for a captive telecommunications market). Motorola introduced the 56000[117] in 1985. The 56000 has been used quite succesfully in many digital audio projects because the 24 bits of data provides room for scaling, higher dynamic range, extra security against limit cycles in recursive filters, better filter coefficient quantization and also additional room for dithering. Other positive aspects of the 56000 include memory moves in parallel with arithmetic operations and modular addressing modes. The 56000 was used by the NeXT MusicKit[91] very effectively. AT&T introduced the first widely avail-

able commercially floating point DSP in 1986[90] which was also used in some audio products.

Custom VLSI for audio DSP was also developed by many manufacturers for their commercial and professional products. Besides large volume cost reductions, custom VLSI has the additional advantage that it's more difficult to reverse engineer and therefore preserves trade secrets. Custom LSI was used by Yamaha in the very popular DX-7 Synthesizer. This synthesizer used the FM synthesis technique developed by Chowning[32] in 1968.

Improvements in Digital/Analog and Analog/Digital converter designs (using oversampling delta-sigma architectures [199][77][3]) produced improved audio quality at lower cost than the previous generation of successive approximation converters.

8.2 The Present

At present, the Compact Disk player is a common feature in many homes worldwide. The introduction of recordable CDs makes it is possible to produce a CD master in a home or garage studio.

The new Digital Video Disk (DVD) standard[51] features multichannel audio (Dolby AC-3[66]) as well as higher sample rates and widths. The MPEG audio standards[24, 20] have found their way into Digital Audio Broadcast[195] (DAB) in Europe. Similar psychoacoustic coders[197] were used by Sony in the MiniDisk[129] system. Phillips used a different system in the Digital Compact Cassette (dcc) system[128].

New "multimedia" VLSI processors[1] are now available offering video and sound processing in a single chip. Multichannel AC-3 decoders are also available in single chips. New A/D and D/A converters are offering 24 bit samples and 96 KHz sampling rates. A current review of digital audio system architecture is given by Kahrs[105].

8.3 The infinite future

Improvements in networking technology, transistor integration and storage media promise the delivery of extremely high quality multichannel audio over the networks like the Internet directly to the consumer. Already, audio coders have been available for use on the Internet.

Prototype completely solid state audio recorders[193] have been built using flash memory as the storage medium and the MPEG coder to save memory bits. It's easy to imagine downloading the most recent multichannel recordings (or live recording) from a favorite artist to your home or portable and saving the recording on your completely electronic recorder.

The combination of fast networking with powerful, portable, low power computation and storage promises new audio features for the consumer and the professional. If the past 50 years are any indication, we are in for an exciting time!

9 Closing remarks and Acknowledgements

The past 50 years have seen absolutely incredible advances in signal processing technology applied to problems in audio and electroacoustics. These solutions have enriched our lives by providing access to the artistry of performers and composers as well as enhanced our acoustic environment.

The editor would like to thank each of the authors for their part in demonstrating the exciting possibilities for audio signal processing.

References

[1] Special issue on media processing. *IEEE Micro*, 16(4), Aug. 1996.

[2] ISO/IEC 11172-3. Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3: Audio. Technical report, ITU, 1992.

[3] R. W. Adams. Design and Implemenation of an Audio 18-bit Analog-to-Digital Converter Using Oversampling Techniques. *J. Audio Eng. Society*, 34(3):143–166, March 1986.

[4] J. Allen. Computer Architecture for Signal Processing. *Proc. IEEE*, 63(4):624–633, April 1975.

[5] S. P. Applebaum. Adaptive arrays. Technical report, Syracuse University Research Corporation, 1966. Rep. SPL TR66-1.

[6] B. B. Bauer. Uniphase unidirectional microphones. *J. Acoust. Soc. Amer.*, 13:41–45, 1941.

[7] F. Baumgarte, C. Ferekidis, and H. Fuchs. A Non-Linear Psychoacoustic Model Applied to ISO MPEG Layer III Coder. In *99th AES-Convention*, page preprint 4087, 1995.

[8] J. Beerends and J. Stemerdink. A perceptual audio quality measure based on a psychoacoustic sound representation. *J. Audio Eng. Soc.*, 40:963–978, December 1992.

[9] Alexander Graham Bell. U.S. Patent No. 174,465, filed Feb 14, 1876, issued March 7, 1876.

[10] A. H. Benade. *Fundamentals of Musical Acoustics.* Dover, New York, 1990.

[11] A. Benallal and A. Gilloire. A new method to stabilize fast RLS algorithms based on a first-order model of the propagation of numerical errors. *Proc. ICASSP88*, pages 1373–1376, Apr 1988.

[12] J. Benesty, P. Duhamel, and Y. Grenier. A multichannel affine projection algorithm with applications to multichannel acoustic echo cancellation. *IEEE Signal Processing letters*, 3(2):35–37, Feb 1996.

[13] J. Benesty, D. Morgan, and M. Sondhi. A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation. *Proc. ICASSP97*, pages 303–306, Apr 1997.

[14] L. L. Beranek. *Acoustics*. American Institute of Physics, for the Acoustical Society of America, (516)349-7800 x 481, 1986. 1st ed. 1954.

[15] N. Bisgaard. Digital feedback suppression: Clinical experiences with profoundly hearing impaired. In J. Beilin and G.R. Jensen, editors, *Recent Developments in Hearing Instrument Technology: 15th Danavox Symposium*, pages 370–384. 1993.

[16] P.M. Boers. Formant enhancement of speech for listeners with sensorineural hearing loss. Technical report, Inst. voor Perceptie Onderzoek, 1980.

[17] S.F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech and Sig. Proc.*, 27:113–120, 1979.

[18] G. Borin, G. De Poli, , and A. Sarti. Musical signal synthesis. In C. Roads, S. T. Pope, A. Piccialli, and G. De Poli, editors, *Musical Signal Processing*, pages 5–30. Swets and Zietlinger, 1997.

[19] M. Bosi, K. Brandenburg, S. Quackenbush, K. Akagiri, H. Fuchs, J.Herre, L. Fielder, M.Dietz, Y. Oikawa, and G. Davidson. ISO/IEC MPEG-2 Advanced Audio Coding. *J. AES*, 51:780 – 792, October 1997.

[20] Marina Bosi, Karlheinz Brandenburg, Schuyler Quackenbush, Louis Fielder, Kenzo Akagin, Hendrick Fuchs, Martin Dietz, Jeurgen Herre, Grant Davidson, and Yoshiaki Oikawa. ISO / IEC MPEG-2 advanced audio coding. *Audio Eng. Soc. Convention*, Preprint 4382, Nov 1996. 36 pages. See also ISO/IEC International Standard IS 13818-7 entitled "MPEG-2 Advanced Audio Coding," April, 1997.

[21] K. Brandenburg, J. Herre, J. D. Johnston, Y. Mahieux, and E. F. Schroeder. ASPEC: Adaptive spectral perceptual entropy of high quality music signals. 1991.

[22] K. Brandenburg and T. Sporer. NMR and Masking Flag: Evaluation of quality using perceptual criteria. In *Proc. of AES 11th Intl. Conf.*, May 1992.

[23] K. Brandenburg and G. Stoll. The ISO/MPEG-1 audio codec: A generic standard for coding of high quality digital audio. *J. AES*, 42:780 – 792, October 1994.

[24] Karlheinz Brandenburg. Perceptual coding of high quality digital audio. In Mark Kahrs and Karlheinz Brandenburg, editors, *Applications of Signal Processing to Audio and Acoustics*. Kluwer, 1997 (to appear).

[25] H.T. Bunnell. On enhancement of spectral contrast in speech for hearing-impaired listeners. *J. Acoust. Soc. Am.*, 88:2546–2556, 1990.

[26] D.K. Bustamante and L.D. Braida. Principal-component amplitude compression for the hearing impaired. *J. Acoust. Soc. Am.*, 82:1227–1242, 1987.

[27] D.K. Bustamante, T.L. Worrell, and M.J. Williamson. Measurement of adaptive suppression of acoustic feedback in hearing aids. *Proc. 1989 Int. Conf. Acoust. Speech and Sig. Proc.*, pages 2017–2020, 1989.

[28] J. Grunwald C. Antweiler and H. Quack. Approximation of optimal step size control for acoustic echo cancellation. *Proc. ICASSP97*, pages 295–298, Apr 1997.

[29] S. J. Elliott C. R. Fuller and P. A. Nelson. *Active Control of Vibration*. Academic Press, 1996.

[30] J. P. Campbell Jr., T. E. Tremain, and V. C. Welch. The proposed federal standard 1016 4800 bps voice coder: Celp. *Speech Technology Magazine*, pages 58–64, April-May 1990.

[31] Antoine Chaigne and Anders Askenfelt. Numerical simulations of piano strings. i. a physical model for a struck string using finite difference methods. *J. Acoustical Soc. of America*, 95(2):1112–1118, Feb 1994.

[32] John M. Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7):526–534, 1973. Reprinted in Roads and Strawn.

[33] P. Chu. Weaver SSB subband acoustic echo canceller. In *Proc. International Workshop on Acoustic Echo Control*, Sep 1993.

[34] C. Colomes, M. Lever, J.B. Rault, and Y.F. Dehery. A perceptual model applied to audio bit rate reduction, 1993.

[35] United States Advanced Television Systems Committee. Digital Audio Compression (AC-3) Standard, December 1995.

[36] Perry R. Cook. *Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing*. PhD thesis, Elec. Eng. Dept., Stanford University, Dec. 1990.

[37] Perry R. Cook. Singing voice synthesis: History, current work, and future directions. *Computer Music J.*, 20(3), Fall 1996.

[38] R. E. Cooke, editor. *Loudspeakers*, volume 1. Audio Engineering Society, 60 East 42nd St. New York, NY 10165, 1980.

[39] H. Cox, R.M. Zeskind, and T. Kooij. Practical supergain. *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-34:393–398, 1986.

[40] H. Cox, R.M. Zeskind, and M.M. Owen. Robust adaptive beamforming. *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-35:1365–1376, 1987.

[41] P. G. Craven and M. A. Gerzon. Coincident microphone simulation covering three dimensional space and yielding various directional responses. U.S. Patent No. 4,042,7779 granted Aug 16, 1997.

[42] G. A. Davidson and M. Bosi. AC-2: High Quality Audio Coding for Broadcasting and Storage. In *46th Annual Broadcast Engineering Conference*, pages 98–105, April 1992.

[43] John R. Deller Jr., John G. Proakis, and John H. Hansen. *Discrete-Time Processing of Speech Signals.* Macmillan, New York, 1993.

[44] P. Derogis, R. Causse, and O. Warfusel. On the reproduction of directivity patterns using multi-loudspeaker sources. *Proc. of I.S.M.A.*, pages 387–392, 1995.

[45] E. Diethorn. Perceptually optimum adaptive filter tap profiles for subband acoustic echo cancellers. *Proc. IEEE Workshop on Audio and Acoustics*, Oct 1995.

[46] N. Dillier, T. Frölich, M. Kompis, H. Bögli, and W.K. Lai. Digital signal processing (DSP) applications for multiband loudness correction digital hearing aids and cochlear implants. *J. Rehab. Res. and Devel.*, 30:95–109, 1993.

[47] A. E. Dolbear. A new system of telephony. *Scientific American*, 44, June 1881.

[48] S. Douglas. Efficient approximate implementations of the fast affine projection algorithm using orthogonal transforms. *Proc. ICASSP96*, pages 1656–1659, May 1996.

[49] D. E. Dudgeon and R. M. Merserau. *Multidimensional Digital Signal Processing.* Prentice Hall, Englewood Cliffs, NJ, 1984.

[50] D. Duttweiller and Y. Chen. A single chip VLSI echo canceller. *Bell Syst. Tech. J.*, 59(2):149–160, 1980.

[51] DVD-Video, Book B, Version 1.1, November 1996.

[52] O. Dyrlund and N. Bisgaard. Acoustic feedback margin improvements in hearing instruments using a prototype dfs (digital feedback suppression) system. *Scand. Audiol.*, 20:49–53, 1991.

[53] B. Edler. Coding of audio signals with overlapping transform and adaptive window shape. *Frequenz*, 43(9):252–256, Sept. 1989.

[54] D.P. Egolf. Review of the acoustic feedback literature from a control theory point of view. In *The Vanderbilt Hearing-Aid Report*, Monographs in Contemporary Audiology, pages 94–103. 1982.

[55] C. G. Eilers. Stereophonic FM Broadcasting. *IRE Trans. Broadcast & TV Receivers*, 7(2):73–80, 1961. Reprinted in IEEE Trans. Consumer Electronics, vol. CE-28, no. 1, Feb. 1982, pp. 5–12.

[56] G. W. Elko. Microphone array systems for hands-free telecommunication. *Speech Communication*, 20:229–240, 1996.

[57] G. W. Elko. A steerable and variable first-order differential microphone array. In *IEEE ICASSP '97*, volume 1, pages 223–226, 1997.

[58] G. W. Elko, R. A. Kubli, D. R. Morgan, and J. E. West. Adjustable filter for differential microphones. U.S. Patent No. 5,586,191, granted Dec 17, 1996.

[59] S. J. Elliott and T. J. Sutton. Performance of feedforward and feedback system s for active control. *IEEE Trans. Speech Audio Processing*, 4:214–223, 1996.

[60] K. Emborg and C. F. Ross. Active control in the SAAB 340. 1993.

[61] A.M. Engebretson and M. French-St.George. Properties of an adaptive feedback equalization algorithm. *J. Rehab. Res. and Devel.*, 30:8–16, 1993.

[62] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-32:1109–1121, 1984.

[63] L. J. Eriksson et al. The selection and application of an IIR adaptive filter for use in active sound attenuation. *IEEE Trans. on ASSP*, ASSP-35:433–437, 1987.

[64] S. J. Elliott et al. Practical implementation of low frequency equalisation using adaptive digital filters. *Journal of the Audio Engineering Society*, 42(12):988–998, Dec 1994.

[65] M. D. Fagen, editor. *A History of Engineering and Science in the Bell System: The Early Years (1875-1925).* Bell Telephone Laboratories, 1975.

[66] L. D. Fielder, M. Bosi, G. A. Davidson, M. Davis, C. Todd, , and S. Vernon. AC-2 and AC-3: Low complexity transform-based audio coding. In *Collected Papers on Digital Audio Bit-Rate Reduction*, pages 54–72. AES, 1996.

[67] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West, and M. M. Sondhi. Autodirective microphone systems. *Acustica*, 73:58–71, 1991.

[68] J. L. Flanagan, K. Ishizaka, and K. L. Shipley. Signal models for low bit-rate coding of speech. *J. Acoustical Soc. of America*, 68(3):780–791, 1980.

[69] Jim L. Flanagan and K. Ishizaka. Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer. *IEEE Trans. Acoustics, Speech, Signal Processing*, 24(2):163–170, April 1976.

[70] H. Fuchs. Improving MPEG audio coding by backward adaptive linear stereo prediction. In *99th AES-Convention*, page preprint 4086, 1995.

[71] A. Gilloire. Performance evaluation of acoustic echo control: required values and measurement procedures. *Ann. Telecommun.*, 49:368–372, 1994.

[72] M. M. Goodwin and G. W. Elko. Constant beamwidth beamforming. In *IEEE ICASSP '93*, volume 1, pages 169–172, 1993.

[73] J.E. Greenberg and P.M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *J. Acoust. Soc. Am.*, 91:1662–1676, 1992.

[74] E. Haensler. The hands-free telephone problem: an annotated bibliography update. *Ann. Telecommun.*, 49:360–367, 1994.

[75] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi. Implementation and evaluation of an acoustic echo canceller using duo-filter control system. *Proc. EUSIPCO96*, pages 1115–1118, Sep 1996.

[76] C. H. Hansen and S. D. Snyder. Active control of noise and vibration, 1996.

[77] Max Hauser. Principles of Oversampling A/D Conversion. *J. Audio Eng. Society*, 39(1/2):3–26, January/February 1991.

[78] J. Herre and J. D. Johnston. Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS). In *101st AES Convention*, 1996.

[79] M.W. Hoffman, T.D. Trine, K.M. Buckley, and D.J. Van Tasell. Robust adaptive microphone array processing for hearing aids: Realistic speech enhancement. *J. Acoust. Soc. Am.*, 96:759–770, 1994.

[80] P. W. Howells. Explorations in fixed and adaptive resolution at ge and surc. *IEEE Trans. Antennas Propag.*, AP-24(5):575–584, 1976.

[81] F. V. Hunt. *Electracoustics*. American Institute of Physics for the Acoustical Society of America, 1954, 1982.

[82] ISO. Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 3: Audio. Technical report, 1994-1997.

[83] ISO. Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 7: Advanced Audio Coding (AAC). Technical report, 1997.

[84] ITU. ITU-T recommendations G.165 and G.167.

[85] ITU. Low bitrate audio coding. Technical report, 1994.

[86] ITU. Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems. Technical report, 1994.

[87] ITU-R. Basic Audio Quality Requirements for Digital Audio Bit-Rate Reduction Systems for Broadcast Emission and Primary Distribution. Technical report, October 1991.

[88] ITU-R. Multichannel stereophonic sound system with and without accompanying picture. Technical report, 1992-1994.

[89] N. Iwakami, T. Moriya, and S. Miki. High quality audio coding at less than 64 kb/s by using transform-domain interleaved vector quantization (Twin-VQ). In *Proc. of ICASSP 1995*, pages 3095–3098. IEEE, 1995.

[90] J. R. Boddie, et. al. The Architecture, Instruction Set and Development Support for the WE DSP-32 Digital Signal Processor. In *Proc. IEEE ICASSP*, pages 421–424, April 1986.

[91] D. Jaffe J. Smith and L. Boyton. Music System Architecture on the NeXT Computer. In Ken Pohlmann, editor, *Audio in Digital Times*, pages 301–312. Audio Engineering Society, May 1989.

[92] David A. Jaffe and Julius O. Smith. Performance expression in commuted waveguide synthesis of bowed strings. In *Proc. 1995 Int. Computer Music Conf., Banff*, pages 343–346. Computer Music Association, 1995.

[93] N. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video.* Prentice-Hall, 1982.

[94] Don H. Johnson and Dan E. Dungeon. *Array Signal Processing: Concepts and Techniques.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[95] J. D. Johnston, D. Sinha, S. Downward, and S. R. Quackenbush. AT& Perceptual Audio Coding (PAC). In N. Geilchrist and C. Grewin, editors, *Collected Papers on Digital Audio Bit-Rate Reduction*, pages 73–82. AES, 1996.

[96] J. P. Joule. On the effects of magnetism upon the dimensions of iron and steel bars. *Phil. Mag.*, 30(3):76–87, 1847.

[97] ISO/IEC JTC1/SC29/WG11. Chairman's Report on the Work of the Audio Ad Hoc Group on Objective Measurements. Technical report, July 1995.

[98] ISO/IEC JTC1/SC29/WG11. NBC Time/Frequency Module Subjective Tests: Overall Results. Technical report, July 1995.

[99] ISO/IEC JTC1/SC29/WG11. MPEG-2 Backwards Compatible CODECS Layer II and III: RACE dTTb Listening Test Report. Technical report, March 1996.

[100] ISO/IEC JTC1/SC29/WG11. Overview of the Report on the Formal Subjective Listening Tests of MPEG-2 NBC Multichannel Audio Coding. Technical report, November 1996.

[101] ISO/IEC JTC1/SC29/WG11. MPEG-4 Audio Working Draft version 3.0. Technical report, April 1997.

[102] ISO/IEC JTC1/SC2/WG11. MPEG/AUDIO Test Report. Technical report, October 1990.

[103] ISO/IEC JTC1/SC2/WG11. Report on the MPEG/AUDIO Subjective Listening Tests. Technical report, November 1991.

[104] ISO/IEC JTC1/SC2/WG11. The MPEG/AUDIO Subjective Listening Test. Technical report, April/May 1991.

[105] M. Kahrs. Digital audio system architecture. In M. Kahrs and K. Brandenburg, editors, *Applications of Digital Signal Processing to Audio and Acoustics*. Kluwer, 1997 (to appear).

[106] Y. Kaneda and J. Ohga. Adaptive microphone array system for noise reduction. *IEEE Trans. Acoust. Speech, and Signal Processing*, ASSP-34:1391–1400, 1986.

[107] J.M. Kates. A computer simulation of hearing aid response and the effects of ear canal size. *J. Acoust. Soc. Am.*, 83:1952–1963, 1988.

[108] J.M. Kates. Feedback cancellation in hearing aids: Results from a computer simulation. *IEEE Trans. Sig. Proc.*, Vol.39:553–562, 1991.

[109] J.M. Kates. Superdirective arrays for hearing aids. *J. Acoust. Soc. Am.*, 94:1930–1933, 1993.

[110] J.M. Kates. Speech enhancement based on a sinusoidal model. *J. Speech and Hear. Res.*, 37:449–464, 1994.

[111] J.M. Kates and M.W. Weiss. A comparison of hearing-aid processing techniques. *J. Acoust. Soc. Am.*, 99:3138–3148, 1996.

[112] Eric Keller. *Fundamentals of Speech Synthesis.* John Wiley and Sons, Inc., New York, 1994.

[113] W. Kellermann. Analysis and design of multirate systems for cancellation of acoustical echoes. *Proc. ICASSP88*, pages 2570–2573, Apr 1988.

[114] J. L. Kelly and C. C. Lochbaum. Speech synthesis. *Proc. Fourth Int. Congress on Acoustics, Copenhagen*, pages 1–4, September 1962. Paper G42.

[115] M.C. Killion. Principles of high-fidelity hearing-aid amplification. In R.E. Sandlin, editor, *Handbook of Hearing Aid Amplification*, volume Volume I: Theoretical and Technical Considerations, pages 45–79. College-Hill Press, 1988.

[116] W. Klippel. The mirror filter - a new basis for reducing nonlinear distortion and equalizing response in woofer systems. *J. Audio Eng. Soc.*, 40(9):675–691, 1992.

[117] K. Kloker. Motorola DSP56000 Digital Signal Processor. *IEEE Micro*, 6(6):29–48, December 1986.

[118] S. Kochkin. Marketrak iii identifies key factors in determining consumer satisfaction. *Hearing J.*, 45:39–44, 1992.

[119] B. Kollmeier, J. Peisseig, and V. Hohnmann. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *J. Rehab. Res. and Devel.*, 30:82–94, 1993.

[120] M. A. Krasner. Digital encoding of speech and audio signals based on the perceptual requirements of the auditory system. Technical Report Technical Report 535, MIT Lincoln Laboratory, 1979.

[121] S. M. Kuo and D. R. Morgan. *Active Noise Control Systems.* John Wiley, 1996.

[122] Timo I. Laakso, Vesa Välimäki, Matti Karjalainen, and Unto K. Laine. Splitting the Unit Delay—Tools for Fractional Delay Filter Design. *IEEE Signal Processing Magazine*, 13(1):30–60, January 1996.

[123] D. M. Ladd and E. W. Hendricks. Active control of 2-d instability waves on an axisymmetric body. *Experiments in Fluids*, 6:69, 1988.

[124] Paul Lagevin, Sept. 1918. French Patent No. 505,703.

[125] P. Leon and D. Etter. Experimental results with increased bandwidth analysis filters in oversampled, subband acoustic echo cancellers. *IEEE Signal Processing letters*, 2(1):1–3, Jan 1995.

[126] H. Levitt and A. Neuman. Evaluation of orthogonal polynomial compression. *J. Acoust. Soc. Am.*, 90:241–252, 1991.

[127] J.S. Lim, A.V. Oppenheim, and L.D. Braida. Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition. *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-26:354–358, 1978.

[128] G. C. P. Lokhoff. dcc – Digital Compact Cassette. *IEEE Trans. Consumer Electronics*, CE-37(3):702–706, Aug. 1991.

[129] Y. Maeda. Minidisk system. *J. Acoustical Soc. Japan*, 49(4), Apr. 1993.

[130] S. Makino. Relationship between the "ES family" algorithms and conventional adaptive algorithms. *Proc. International Workshop on Acoustic Echo and Noise Control*, June 1995.

[131] S. Makino and Y. Kaneda. Exponentially weighted stepsize projection algorithm for acoustic echo cancellers. *Trans. IEICE Japan*, E75-A(11):1500–1508, Nov. 1992.

[132] S. Makino, J. Noebauer, Y. Haneda, and A. Nakagawa. SSB subband echo canceller using low-order projection algorithm. *Proc. ICASSP96*, pages 945–948, May 1996.

[133] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa. Subband stereo echo canceller using the projection algorithm with fast convergence to the true echo path. *Proc. ICASSP97*, pages 299–302, Apr 1997.

[134] H. S. Malvar. *Signal Processing with Lapped Transforms.* Artech House, 1992.

[135] J. D. Markel and A. H. Gray. *Linear Prediction of Speech.* Springer Verlag, New York, 1976.

[136] J. P. Marrian. On sonorous phenomena in electro-magnets. *Phil. Mag*, 25:382–384, 1844.

[137] R. N. Marshall and W. R. Harry. A new microphone providing uniform directivity over an extended frequency range. *J. Acoust. Soc. Amer.*, 12:481–498, 1941.

[138] Max V. Mathews. *The Technology of Computer Music.* MIT Press, Cambridge, MA, 1969.

[139] R.J. McAulay and M.L. Malpass. Speech enhancement using a soft-decision noise-suppression filter. *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-28:137–145, 1980.

[140] Michael E. McIntyre, Robert T. Schumacher, and James Woodhouse. On the oscillations of musical instruments. *J. Acoustical Soc. of America*, 74(5):1325–1345, Nov. 1983.

[141] M. Montazeri and P. Duhamel. A set of algorithms linking NLMS and Block RLS algorithms. *IEEE Trans. Signal Processing*, 43(2):444–453, Feb 1995.

[142] R. A. Monzingo and T. W. Miller. *Introduction to Adaptive Arrays.* John Wiley and Sons, New York NY, 1980.

[143] M. Morari and E. Zafiriou. *Robust Process Control.* Prentice Hall, 1989.

[144] D. Morgan. Slow asymptotic convergence of LMS acoustic echo cancellers. *IEEE Trans. Speech and Audio*, 3(2):126–136, Mar 1995.

[145] P. A. Nelson and S. J. Elliott. *Active Sound Control.* Academic Press, 1992.

[146] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada. Inverse filter design and equalization zones in multichannel sound reproduction. *IEEE Trans. on Speech and Audio Processing*, 3(3):185–192, May 1995.

[147] K. Ochiai, T. Araseki, and T. Ogihara. Echo canceller with two echo path models. *IEEE Trans. Commun.*, COM-25:589–595, 1977.

[148] K. Ozeki and T. Umeda. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Trans. IEICE Japan*, J67-A:126–132, Feb 1984.

[149] B. Paillard, P. Mabilleu, S. Morissette, and J. Soumagne. PERCEVAL: Perceptual evaluation of the quality of audio signals. *J. Audio Eng. Soc.*, 40:21–31, January/February 1992.

[150] C.V. Pavlovic. Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J. Acoust. Soc. Am.*, 82:413–422, 1987.

[151] R. Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J. Acoust. Soc. Am.*, 63:533–549, 1978.

[152] R. Plomp. The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *J. Acoust. Soc. Am.*, 83:2322–2327, 1988.

[153] Nick Porcaro, Pat Scandalis, Julius O. Smith, David A. Jaffe, and Tim Stilson. SynthBuilder—a graphical real-time synthesis, processing and performance system. In *Proc. 1995 Int. Computer Music Conf., Banff*, pages 61–62. Computer Music Association, 1995. See http://www-leland.stanford.edu/group/OTL/SynthBuilder.html for information on how to obtain and run SynthBuilder.
See also http://www-ccrma.stanford.edu for related information.

[154] J. Princen, A. Johnson, and A. Bradley. Subband/transform coding using filter bank designs based on time domain aliasing cancellation. In *Proc. of the ICASSP 1987*, pages 2161–2164, 1987.

[155] J. Princen and J. D. Johnston. Audio coding with signal adaptive filterbanks. In *Proc. of ICASSP 1995*, pages 3071 – 3074. IEEE, 1995.

[156] D. Chabries R. Harris and F. Bishop. A variable step (VS) adaptive filter algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-34(2):309–316, Apr 1986.

[157] A. L. Limberg R. M. Christensen, J. J. Gibson. Omnidirectional sound field reproducing system. U.S. Patent No. 3,824,342, granted July 16, 1974.

[158] B. Rafaely and S. J. Elliott. $H2/H_\infty$ active control of sound in a headrest; design and implementation. *IEEE Trans. Control Systems Technology*, 1997. To be published.

[159] C. W. Rice and E. W. Kellogg. Notes on the development of a new type of hornless loud speaker. *Trans. Am. Inst. Elec. Engrs.*, 44:475–480, April 1925.

[160] Curtis Roads. *The Computer Music Tutorial.* MIT Press, Cambridge, MA, 1996.

[161] J. H. Rothweiler. Polyphase quadrature filters - a new subband coding technique. In *Proc. ICASSP 1983*, pages 1280–1283, 1983.

[162] Y. Kaneda S. Makino and N. Koizumi. Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response. *IEEE Trans. Speech and Audio*, 1(1):101–108, Jan 1993.

[163] G. J. Sandell and W. L. Martens. Perceptual evaluation of principal-component-based synthesis of musical timbres. *J. Audio Eng. Soc.*, 43:1013–1028, December 1995.

[164] Gary Paul Scavone. *An Acoustic Analysis of Single-Reed Woodwind Instruments with an Emphasis on Design and Performance Issues and Digital Waveguide Modeling Techniques.* PhD thesis, Music Dept., Stanford University, March 1997. Available as CCRMA Technical Report No. STAN–M–100 or from ftp://ccrma-ftp.stanford.edu/pub/Publications/Theses/GaryScavone-Thesis/.

[165] S. A. Schelkunoff. A mathematica theory of linear arrays. *Bell Syst. Tech. J.*, 22:80–107, 1943.

[166] M. Schroeder, B. Atal, and J. Hall. Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoustical Soc. America*, 66(6):1647–1652, 1979.

[167] J. Schroeter and M. M. Sondhi. Techniques for estimating vocal-tract shapes from the speech signal. *IEEE Trans. Speech and Audio Processing*, 2:133–150, Jan. 1994.

[168] G. Schuller. A low delay filter bank for audio coding with reduced pre-echos. page preprint 4088, 1995.

[169] G. M. Sessler, editor. *Electrets*, volume 33, chapter 8, pages 383–420. Springer-Verlag, 1987.

[170] G. M. Sessler and J. E. West. Self-biased condenser microphone with high capacitance. *J. Acoust. Soc. Amer.*, 34:1787–1788, 1962.

[171] S. Shimauchi and S. Makino. Stereo projection echo canceller with true echo path estimation. *Proc. ICASSP95*, pages 3059–3062, May 1995.

[172] E. W. Siemens, Jan. 1874. German Patent No. 2355.

[173] A.M. Simpson, B.C.J. Moore, and B.R. Glasberg. Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Otolaryngol. Suppl. 469*, pages 101–107, 1990.

[174] D. Sinha and A. H. Tewfik. Low bit-rate transparent audio compression using adapted wavelets. *IEEE Trans. Acoust., Speech, and Signal Processing*, 41(12):3463–3479, 1993.

[175] D. Slock and T. Kailath. Numerically stable fast transversal filters for recursive least-squares adaptive filtering. *IEEE Trans. Signal Processing*, 39(1):92–114, Jan 1991.

[176] J. O. Smith. Acoustic modeling using digital waveguides. In C. Roads, S. T. Pope, A. Piccialli, and G. De Poli, editors, *Musical Signal Processing*, pages 221–263. Swets and Zietlinger, 1997.

[177] Julius O. Smith. Efficient simulation of the reed-bore and bow-string mechanisms. In *Proc. 1986 Int. Computer Music Conf., The Hague*, pages 275–280. Computer Music Association, 1986. Also available in [179].

[178] Julius O. Smith. Elimination of limit cycles and overflow oscillations in time-varying lattice and ladder digital filters. In *Proc. IEEE Conf. Circuits and Systems, San Jose*, pages 197–299, May 1986. Short conference version. Full version available in [179].

[179] Julius O. Smith. Music applications of digital waveguides. Technical Report STAN–M–39, CCRMA, Music Dept., Stanford University, 1987. A compendium containing four related papers and presentation overheads on digital waveguide reverberation, synthesis, and filtering. CCRMA technical reports can be ordered by calling (415)723-4971 or by sending an email request to info@ccrma.stanford.edu.

[180] Julius O. Smith. Viewpoints on the history of digital synthesis. In *Proc. 1991 Int. Computer Music Conf., Montreal*, pages 1–10. Computer Music Association, 1991. Keynote paper.

[181] Julius O. Smith. Physical modeling using digital waveguides. *Computer Music J.*, 16(4):74–91, Winter 1992. Special issue: Physical Modeling of Musical Instruments, Part I.

[182] Julius O. Smith. Physical modeling synthesis update. *Computer Music J.*, 20(2):44–56, Summer 1996. Available online at http://www-ccrma.stanford.edu/~jos/.

[183] W. Soede, A.J. Berkhout, and F.A. Bilsen. Development of a directional hearing instrument based on array technology. *J. Acoust. Soc. Am.*, 94:785–798, 1993.

[184] W. Soede, F.A. Bilsen, and A.J. Berkhout. Assessment of a directional microphone array for hearing-impaired listeners. *J. Acoust. Soc. Am.*, 94:799–808, 1993.

[185] M. Sondhi. An adaptive echo canceller. *Bell Syst. Tech. J.*, 46(3):497–511, 1967.

[186] M. Sondhi, D. Morgan, and J. Hall. Stereophonic acoustic echo cancellation - an overview of the fundamental problem. *IEEE Signal Processing letters*, 2(8):148–151, Aug 1995.

[187] M. M. Sondhi and G. W. Elko. Adaptive optimization of microphone arrays under a nonlinear constraint. In *IEEE Proc. ICASSP*, pages 981–984, 1986.

[188] R.W. Stadler and W.M. Rabinowitz. On the potential of fixed arrays for hearing aids. *J. Acoust. Soc. Am.*, 94:1332–1342, 1993.

[189] J. C. Stapleton and Stephen C. Bass. Synthesis of musical tones based on the karhunen-loeve transform. *IEEE Trans. Acoustics, Speech, Signal Processing*, ASSP-36(3):305–319, March 1988.

[190] T. G. Stockham, T. M. Cannon, and R. B. Ingebretsen. Blind deconvolution through digital signal processing. *Proc. IEEE*, 63(4):678–692, April 1975.

[191] M.A. Stone and B.C.J. Moore. Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality. *J. Rehab. Res. and Devel.*, 29:39–56, 1992.

[192] A. Sugiyama, F. Hazu, and M. Iwadare. Adaptive Transform Coding with an Adaptive Block Size (ATCABS). In *Proc. of the ICASSP 1990*, pages 1093–1096, 1990.

[193] A. Sugiyama, M. Iwadare, T. Manake, and N. Ohdate. A new implementation of the Silicon Audio Player based on a MPEG/Audio decoder LSI. *IEEE Trans. Consumer Electronics*, CE-43(2):207–215, May 1997.

[194] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima. Fast projection algorithm and its step size control. *Proc. ICASSP95*, pages 945–948, May 1995.

[195] K. Taura, M. Tsuijshita, M. Takeda, and H. Kato. A DAB Receiver. *IEEE Trans. Consumer Electronics*, CE-42(3):322–327, Aug. 1996.

[196] A. L. Thuras, July 1932. U.S. Patent 1,869,178.

[197] K. Tsutsui, H. Suzuki, O. Shimoyoshi, M. Sonohara, K. Akagiri, and R.M. Heddle. ATRAC: Adaptive Transform Acoustic Coding for MiniDisc. page preprint 3456, 1992.

[198] Vesa Välimäki. *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters.* PhD thesis, Report no. 37, Helsinki University of Technology, Faculty of Elec. Eng., Lab. of Acoustic and Audio Signal Processing, Espoo, Finland, Dec. 1995.

[199] Rudy van de Plassche. *Integrated Analog-to-Digital and Digital-to-Analog Converters.* Kluwer Academic Publishers, 1994.

[200] P. Vary. On the enhancement of noisy speech. *Signal Processing II: Theories and Applications*, pages 327–330, 1983.

[201] B. L. Vercoe, W. G. Gardner, and E. D. Scheirer. Structured audio: Creation, transmission, and rendering of parametric sound representations. *Proc. IEEE (to appear)*, Feb. 1998.

[202] M. P. Verge. *Aeroacoustics of Confined Jets with Applications to the Physical Modeling of Recorder-Like Instruments.* PhD thesis, Eindhoven University, 1995.

[203] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice Hall, 1995.

[204] E. Villchur. Signal processing to improve speech intelligibility in perceptive deafness. *J. Acoust. Soc. Am.*, 53:1646–1657, 1973.

[205] D. B. Ward, R. A. Kennedy, and R. C. Williamson. Theory and design of broadband sensor arrays with frequency invariant farfield beam patterns. *J. Acoust. Soc. Amer.*, 97(2):1023–1034, 1995.

[206] M.R. Weiss, E. Aschkenasy, and T.W. Parsons. Study and development of the INTEL technique for improving speech intelligibility. Technical report, Rome Air Devel. Ctr., 1975.

[207] E. C. Wente. The sensitivity and precision of the electrostatic transmitter for measuring sound intensities. *Phys. Rev.*, 19, 1932.

[208] E. C. Wente and A. L. Thuras. Moving-coil telephone receivers and microphones. *J. Acoust. Soc. Amer.*, 3:44–55, 1931.

[209] P. D. Wheeler. *Voice communications in a cockpit noise environment - the role of active noise reduction.* PhD thesis, University of Southampton, 1986.

[210] B. Widrow, J.R. Jr. Glover, J.M. McCool, C.S. Williams, R.H. Hearn, J.R. Ziedler, E. Jr. Dong, and R.C. Goodlin. Adaptive noise canceling: Principles and applications. *Proc. IEEE*, 63:1692–1716, 1975.

[211] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode. Adaptive antenna systems. *Proc. IEEE*, 55, 1967.

[212] R. Zelinski and P. Noll. Adaptive transform coding of speech signals. *IEEE Trans. Acoust., Speech, and Signal Processing*, 25:299 – 309, 1979.

[213] E. Zwicker and H. Fastl. *Psychoacoustics, facts, and models.* Springer-Verlag, 1990.