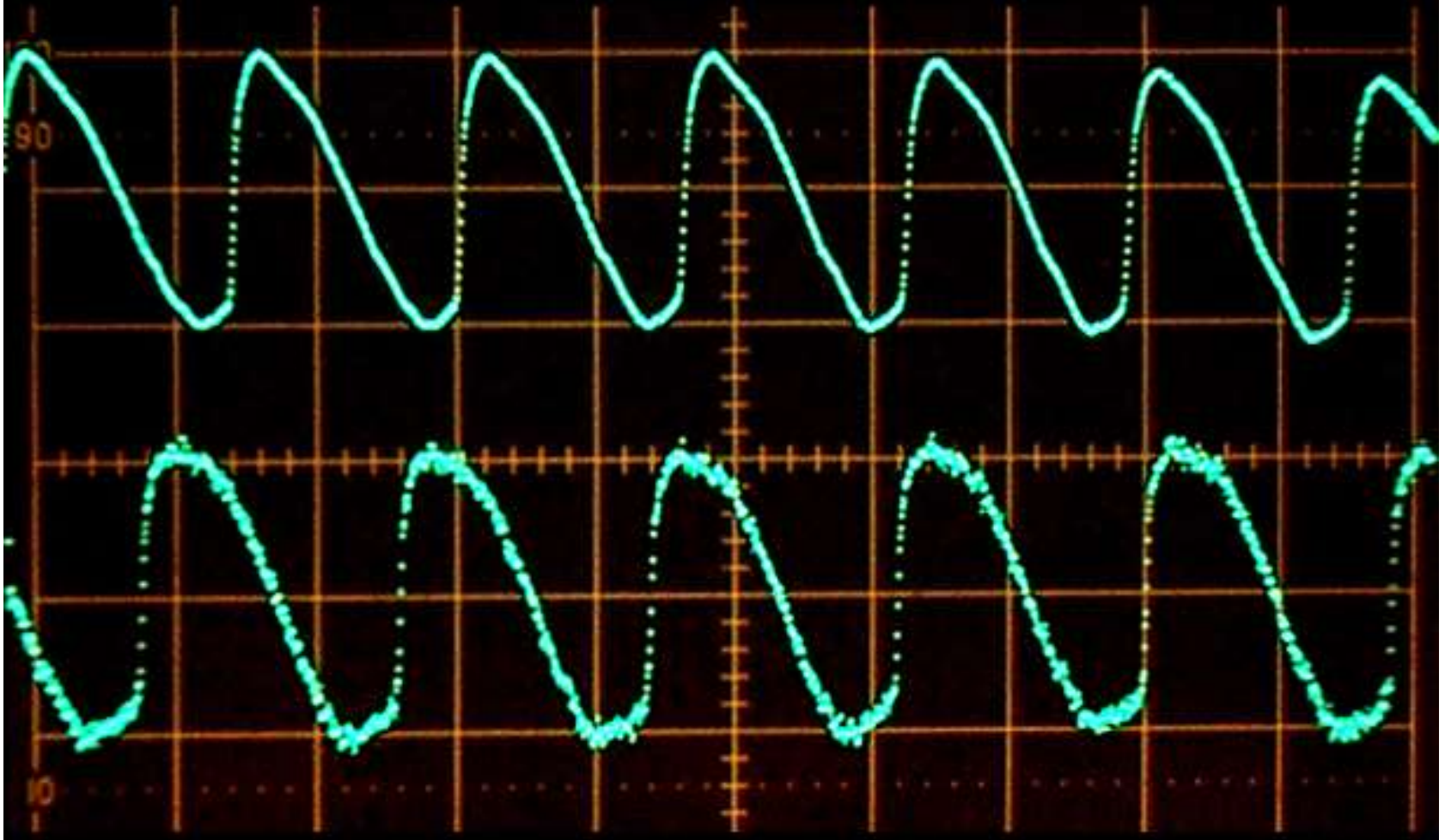
Audio Engineering and Voice Technology



Abagail Morales

Nya Dover

First Edition, 2012

ISBN 978-81-323-1469-1

Published by:
College Publishing House
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Introduction to Audio Engineering

Audio engineering is a skilled trade that deals with the use of machinery and equipment for the recording, mixing and reproduction of sounds. The field draws on many artistic and vocational areas, including electronics, acoustics, psychoacoustics, and music. An audio engineer is proficient with different types of recording media, such as analog tape, digital multitrack recorders and workstations, and computer knowledge. With the advent of the digital age, it is becoming more and more important for the audio engineer to be versed in the understanding of software and hardware integration from synchronization to analog to digital transfers.

Audio engineering concerns the creative and practical aspects of sounds and music, in contrast with the formal engineering discipline known as acoustical engineering. Producer, engineer, mixer Phil Ek has described audio engineering as the "physical recording of any project—the placing of microphones, the turning of pre-amp knobs, the setting of levels—and the producer is the guy who directs that process." Many recording engineers also invented new technology, equipment and techniques, to enhance the process and art.

Lexical dispute

The expressions "audio engineer" and "sound engineer" are ambiguous. Such terms can refer to a person working in sound and music production, as well as to an engineer with a degree who designs professional equipment for these tasks .

Individuals who design acoustical simulations of rooms, shaping algorithms for digital signal processing and computer music problems, perform institutional research on sound, and other advanced fields of audio engineering are most often graduates of an accredited college or university, or have passed a difficult civil qualification test.

Certain jurisdictions specifically prohibit the use of the title engineer to any individual not a registered member of the local professional engineering body, responsible for regulating ethics and the safety of the public with respect to the engineering profession, which often may not include audio engineers. In such situations they are formally referred to as audio technicians.

Other languages, such as German and Italian, have different words to refer to these activities. For instance, in German, the *Tontechniker* (audio technician) is the one who operates the audio equipment and the *Tonmeister* (sound master) is a person who creates recordings or broadcasts of music who is both deeply musically trained (in 'classical' and non-classical genres) and who also has a detailed theoretical and practical knowledge of virtually all aspects of sound, whereas the *Toningenieur* (audio engineer) is the one who designs, builds and repairs it.

Practitioners

An engineer at an audio console

An audio engineer is someone with experience and training in the production and manipulation of sound through mechanical (analog) or digital means. As a professional title, this person is sometimes designated as a sound engineer or recording engineer instead. A person with one of these titles is commonly listed in the credits of many commercial music recordings (as well as in other productions that include sound, such as movies).

Audio engineers are generally familiar with the design, installation, and/or operation of sound recording, sound reinforcement, or sound broadcasting equipment, including large and small format consoles. In the recording studio environment, the audio engineer records, edits, manipulates, mixes, and/or masters sound by technical means in order to realize an artist's or record producer's creative vision. While usually associated with music production, an audio engineer deals with sound for a wide range of applications, including post-production for video and film, live sound reinforcement, advertising, multimedia, and broadcasting. When referring to video games, an audio engineer may also be a computer programmer.

In larger productions, an audio engineer is responsible for the technical aspects of a sound recording or other audio production, and works together with a record producer or director, although the engineer's role may also be integrated with that of the producer. In smaller productions and studios the sound engineer and producer is often one and the same person.

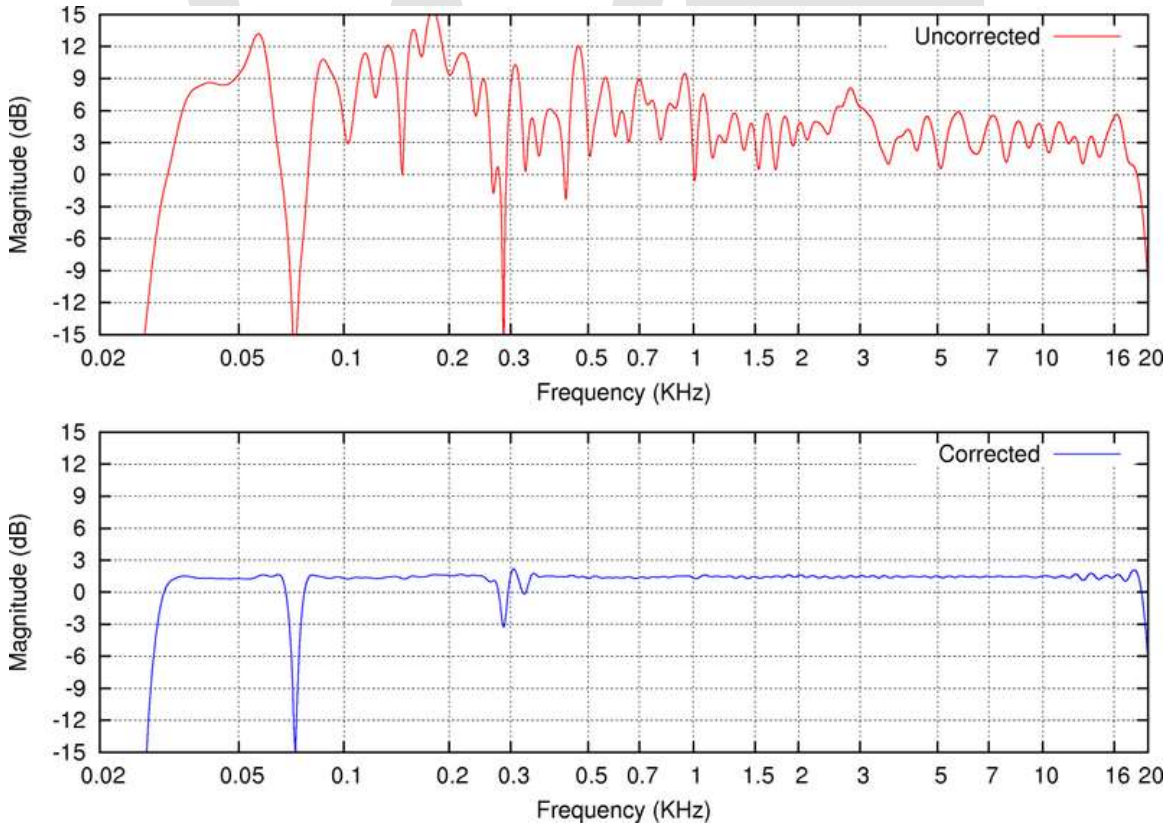
In typical sound reinforcement applications, audio engineers often assume the role of producer, making artistic and technical decisions, and sometimes even scheduling and budget decisions.

Different professional branches

There are four distinct steps to commercial production of a recording: Recording, editing, mixing, and mastering. Typically, each is performed by a sound engineer who specializes only in that part of production.

- Studio engineer – an engineer working within a studio facility, either with a producer or independently
- Recording engineer – engineer who records sound.
- Assistant engineer – often employed in larger studios, allowing them to train to become full-time engineers. They often assist full-time engineers with microphone setups, session breakdowns and in some cases, rough mixes.
- Mixing engineer – a person who creates mixes of multi-track recordings. It is common for a commercial record to be recorded at one studio and later mixed by different engineers in other studios.
- Mastering engineer – typically the person who mixes the final stereo tracks (or sometimes just a few tracks or stems) that the mix engineer produces. The mastering engineer makes any final adjustments to the overall sound of the record in the final step before commercial duplication. Mastering engineers use principles of equalization and compression to affect the coloration of the sound.
- Game audio designer engineer – deals with sound aspects of game development.
- Live sound engineer – a person dealing with live sound reinforcement. This usually includes planning and installation of speakers, cabling and equipment and mixing sound during the show. This may or may not include running the foldback sound. A live/sound reinforcement engineer hears musical material and tries to correlate that sonic experience with system performance.

- Foldback or Monitor engineer – a person running foldback sound during a live event. The term "foldback" is outdated and refers to the practice of folding back audio signals from the FOH (Front of House) mixing console to the stage in order for musicians to hear themselves while performing. Monitor engineers usually have a separate audio system from the FOH engineer and manipulate audio signals independently from what the audience hears, in order to satisfy the requirements of each performer on stage. In-ear systems, digital and analog mixing consoles, and a variety of speaker enclosures are typically used by monitor engineers. In addition most monitor engineers must be familiar with wireless or RF (radio-frequency) equipment and must interface personally with the artist(s) during each performance.
- Systems engineer – responsible for the design setup of modern PA systems which are often very complex. A systems engineer is usually also referred to as a "crew chief" on tour and is responsible for the performance and day-to-day job requirements of the audio crew as a whole along with the FOH audio system.
- Audio post engineer – a person who edits and mixes audio for film and/or television.



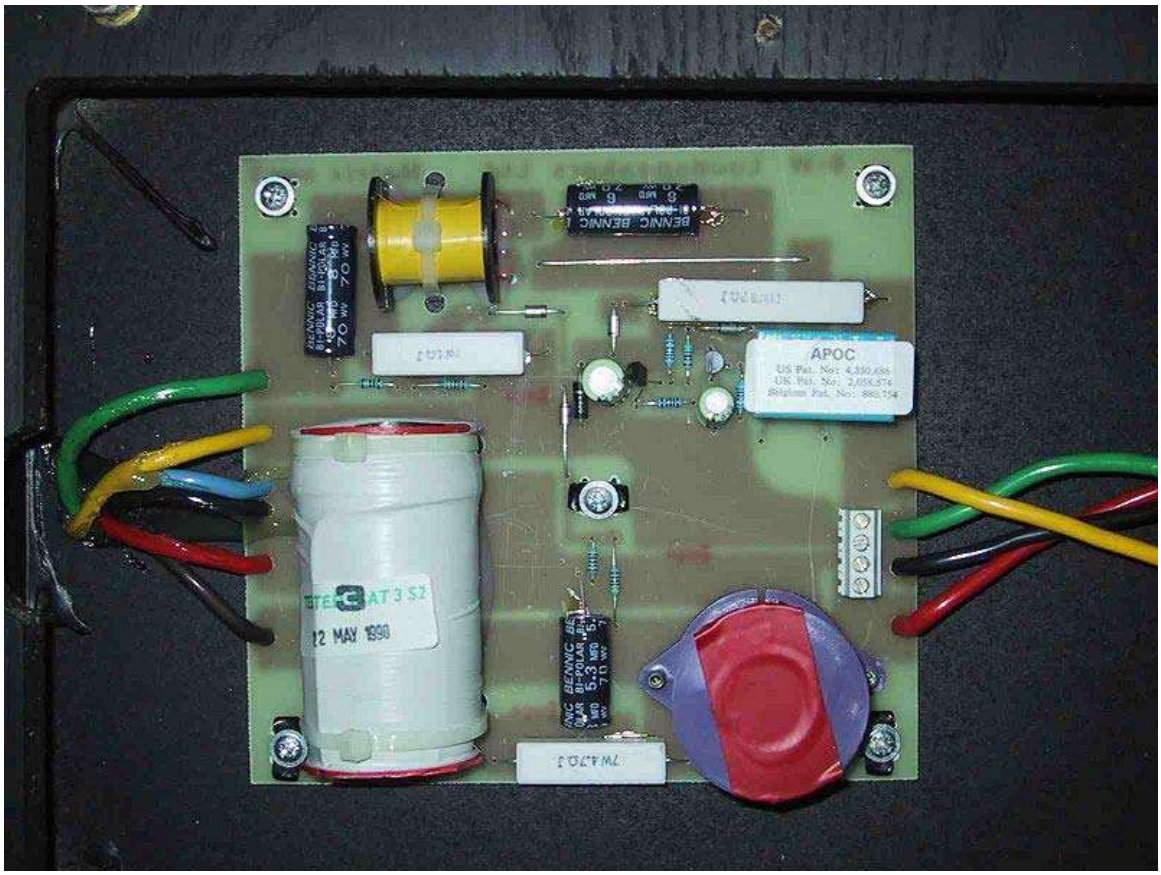
Correcting a room's frequency response

Education

Audio engineers come from backgrounds such as fine arts, broadcasting, music or electronics. Many colleges and accredited institutions around the world offer degrees in audio engineering, such as a BS in audio production. The University of Miami's Frost School of Music was the first university in the United States to offer a four-year Bachelor of Music degree in Music Engineering Technology. In the last 25 years, some contemporary music schools have initiated audio engineering programs, usually awarding a Bachelor of Music degree to graduates. Additionally, a number of audio engineers are autodidacts with no formal training.

Chapter 2

Audio Crossover

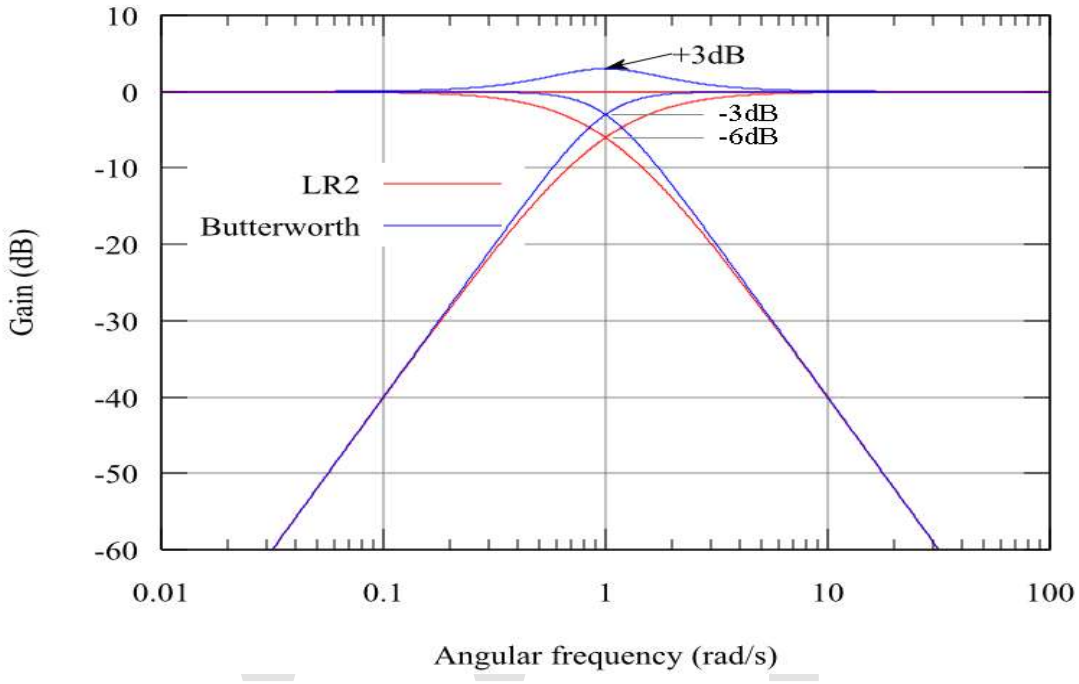


A passive 2-way crossover designed to operate at loudspeaker voltages

Audio crossovers are a class of electronic filter used in audio applications. Most individual loudspeaker drivers are incapable of covering the entire audio spectrum from low frequencies to high frequencies with acceptable relative volume and lack of distortion so most hi-fi speaker systems use a combination of multiple loudspeakers or drivers, each catering to a different frequency band. Crossovers split the audio signal into separate frequency bands that can be separately routed to loudspeakers optimized for those bands.

Crossovers also enable multiband processing and multiple amplification where the audio signal is split into bands that are adjusted (equalized, compressed, echoed, etc.) separately before they are mixed together again. Some examples are: multiband dynamics (compression, limiting, de-essing), multiband distortion, bass enhancement, high frequency exciters, and noise reduction (for example: Dolby A noise reduction).

Overview



Comparison of the magnitude response of 2 pole Butterworth and Linkwitz-Riley crossover filters. The summed output of the Butterworth filters has a +3dB peak at the crossover frequency.

The definition of an ideal audio crossover changes relative to the task at hand. If the separate bands are to be mixed back together again (as in multiband processing), then the ideal audio crossover would split the incoming audio signal into separate bands that do not overlap or interact and which result in an output signal unchanged in frequency, relative levels, and phase response. This ideal performance can only be approximated. How to implement the best approximation is a matter of lively debate. On the other hand, if the audio crossover separates the audio bands in a loudspeaker, there is no requirement for mathematically ideal characteristics within the crossover itself, as the frequency and phase response of the loudspeaker drivers within their mountings will eclipse the results. Satisfactory output of the complete system comprising the audio crossover *and* the loudspeaker drivers in their enclosure(s) is the design goal. Such a goal is often achieved using non-ideal, asymmetric crossover filter characteristics.

Many different crossover types are used in audio, but they generally belong to one of the following classes.

Classification

Classification based on the number of filter sections

In loudspeaker specifications, one often sees a speaker classified as an "N-way" speaker. N is a positive whole number greater than 1, and it indicates the number of filter sections. A 2-way crossover consists of a low-pass and a high-pass filter. A 3-way crossover is constructed as a combination of low-pass, band-pass and high-pass filters (LPF, BPF and HPF respectively). The BPF section is in turn a combination of HPF and LPF sections. 4 (or more) way crossovers are not very common in speaker design, primarily due to the complexity involved, which is not generally justified by better acoustic performance.

An extra HPF section may be present in an "N-way" loudspeaker crossover to protect the lowest-frequency driver from frequencies lower than it can safely handle. Such a crossover would then have a bandpass filter for the lowest-frequency driver. Similarly, the highest-frequency driver may have a protective LPF section to prevent high frequency damage, though this is far less common.

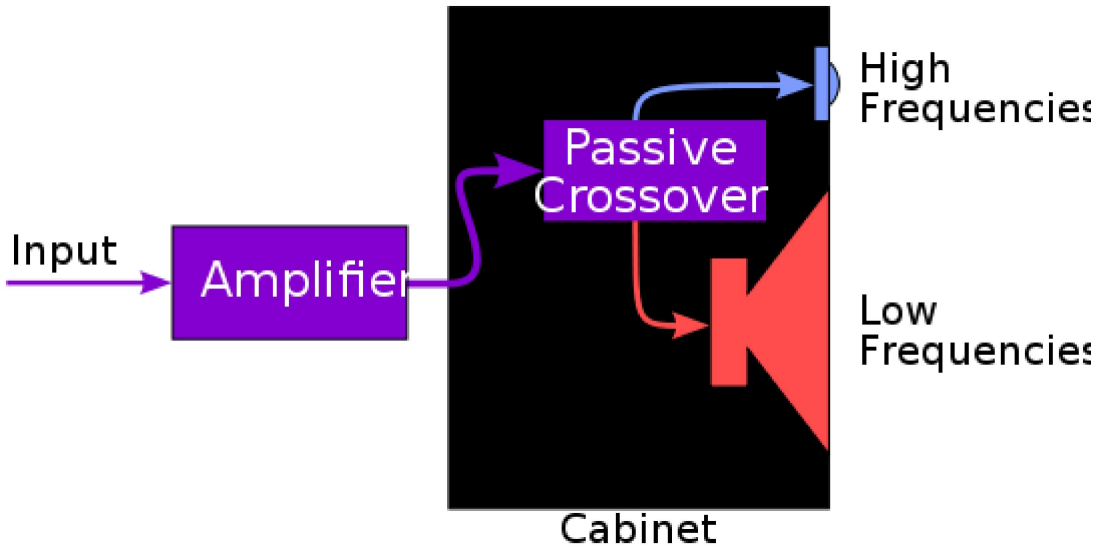
Recently, a number of manufacturers have begun using what is often called "N.5-way" crossover techniques for stereo loudspeaker crossovers. This usually indicates the addition of a second woofer that plays the same bass range as the main woofer but rolls off far before the main woofer does.

Remark: Filter sections mentioned here is not to be confused with the individual 2-pole filter sections that a higher order filter consists of.

Classification based on components

Crossovers can also be classified based on the design approach; by the type of components used.

Passive



A passive crossover

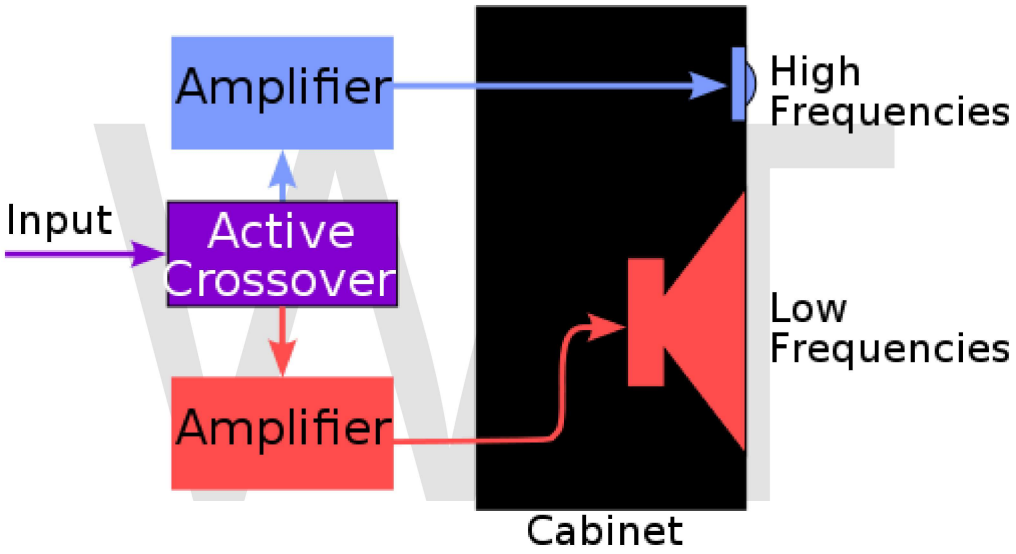
A passive crossover is made entirely of passive components, arranged most commonly in a Cauer topology to achieve a Butterworth filter. Passive filters use non-reactive resistors combined with reactive components such as capacitors and inductors. Very high performance passive crossovers are likely to be more expensive than active crossovers since individual components capable of good performance at the high currents and voltages at which speaker systems are driven are hard to make, and expensive. Polypropylene, metalized polyester foil, and paper-electrolytic capacitors are common. Inductors may have air cores, powdered metal cores, ferrite cores, or laminated silicon steel cores, and most are wound with enamelled copper wire. Some passive networks include devices such as fuses, PTC devices, bulbs or circuit breakers to protect the loudspeaker drivers from accidental overpowering. Modern passive crossovers increasingly incorporate equalization networks (e.g., Zobel networks) that compensate for the changes in impedance with frequency inherent in virtually all loudspeakers. The issue is complex, as part of the change in impedance is due to acoustic loading changes across a driver's passband.

On the negative side, passive networks may be bulky and cause power loss. They are not only frequency specific, but also impedance specific. This prevents interchangeability with speaker systems of different impedances. Ideal crossover filters, including impedance compensation and equalization networks, can be very difficult to design, as the components interact in complex ways. Crossover design expert Siegfried Linkwitz said of them that "the only excuse for passive crossovers is their low cost. Their behavior changes with the signal level dependent dynamics of the drivers. They block the power amplifier from taking maximum control over the voice coil motion. They are a waste of time, if accuracy of reproduction is the goal."

Alternatively, passive components can be utilised to construct filter circuits before the amplifier. This is called passive line-level crossover.

Active

An active crossover contains active components (i.e., those with gain) in its filters. In recent years, the most commonly used active device is an op-amp; active crossovers are operated at levels suited to power amplifier inputs in contrast to passive crossovers which operate after the power amplifier's output, at high current and in some cases high voltage. On the other hand, all circuits with gain introduce noise, and such noise has a more deleterious effect when introduced prior to the signal being amplified by the power amplifiers.



Typical usage of an active crossover, though a passive crossover can be positioned similarly before the amplifiers

Active crossovers always require the use of power amplifiers for each output band. Thus a 2-way active crossover needs two amplifiers—one each for the woofer and tweeter. This means that an active crossover based system will often cost more than a passive crossover based system, although none of the amplifiers needs to provide output as high as for an equivalent sound level full-frequency, power amplifier, which reduces cost. The cost and complication disadvantages of active crossovers are offset by the following gains:

- a frequency response independent of the dynamic changes in a driver's electrical characteristics.
- typically, the possibility of an easy way to vary or fine tune each frequency band to the specific drivers used. Examples would be crossover slope, filter type (e.g., Bessel, Butterworth, etc.), relative levels, ...

- isolation of each driver from signals handled by drivers, thus reducing intermodulation distortion and overdriving
- The power amplifiers are directly connected to the speaker drivers, thereby maximizing amplifier damping control of the speaker voice coil, reducing consequences of dynamic changes in driver electrical characteristics, all of which are likely to improve the transient response of the system
- reduction in power amplifier output requirement. With no energy being lost in passive components, amplifier requirements are reduced considerably (up to 1/2 in some cases), reducing costs, and potentially increasing quality.

Digital

Active crossovers can be implemented digitally using a DSP chip or other microprocessor. They either use digital approximations to traditional analog circuits, known as IIR filters (Bessel, Butterworth, Linkwitz-Riley etc.), or they use Finite impulse response (FIR) filters. IIR filters have many similarities with analog filters and are relatively undemanding of CPU resources; FIR filters on the other hand usually have a higher order and therefore require more resources for similar characteristics. They can be designed and built so that they have a linear phase response, which is thought desirable by many involved in sound reproduction. There are drawbacks though—in order to achieve linear phase response, a longer delay time is incurred than would be necessary with an IIR or minimum phase FIR filters. IIR filters, which are by nature recursive have the drawback that if not carefully designed they may enter limit cycles resulting in non-linear distortion.

Mechanical

This crossover type is mechanical and uses the properties of the materials in a driver diaphragm to achieve the necessary filtering. Such crossovers are commonly found in full-range speakers which are designed to cover as much of the audio band as possible. One such is constructed by coupling the diaphragm of the speaker to the voice coil through a compliant section and directly attaching a small lightweight cone called *whizzer* to the voice coil. The compliant section is intended to ensure that the primary full size diaphragm responds only to lower frequencies. The whizzer is directly coupled to the voice coil and responds to all frequencies, but due to its small size only gives a useful level of output at higher frequencies. This combination results in the main diaphragm having an upper cut-off frequency while the size of the whizzer sets the lower limit to the whizzer's response, thereby implementing a crossover action. The choice/weight of materials used for the diaphragm, whizzer and the speaker's suspension determine the crossover frequency and the effectiveness of the crossover. This sort of crossover is much more complex to design, especially if the highest degree of performance is desired. Extensive trial and error is required. Over several years, the compliance of the joint can change, negatively affecting the frequency response of the speaker.

An alternative is to use the dust cap as a high frequency radiating device, also crossed over by mechanical compliance from the primary diaphragm. High frequency dispersion

is somewhat different for this approach than for whizzer cones. Another possibility is to build the primary cone with such profile, and of such materials, that the neck area remains rigid, radiating all frequencies, while the outer areas of the cone are selectively decoupled, radiating only at lower frequencies.

Speakers which use these mechanical crossovers have some advantages in sound quality despite the difficulties of designing and manufacturing them, and despite the inevitable output limitations. Full-range drivers have a single acoustic center, and can have relatively modest phase change across the audio spectrum. For best performance at low frequencies, these drivers require careful enclosure design. Their small size (typically 165 to 200 mm) requires considerable cone excursion to reproduce bass effectively, but the short voice coils required for reasonable high frequency performance can only move over a limited range. Nevertheless, within these constraints, cost and complications are reduced, as no crossovers are required.

Those who do not prefer the sound of full-range drivers sometimes suggest that a single diaphragm that must produce both low and high frequencies does neither justice, and there are plenty of examples to support this theory. Almost all high fidelity speakers are 2 or 3 way designs, which lends weight to this view.

Classification based on filter order or slope

Just as filters have different orders, so do crossovers, depending on the filter slope they implement. The final acoustic slope may be completely determined by the electrical filter or may be achieved by combining the electrical filter's slope with the natural characteristics of the driver. In the former case, the only requirement is that each driver has a flat response at least to the point where its signal is approximately −10dB down from the passband. In the latter case, the final acoustic slope is usually steeper than that of the electrical filters used. A third- or fourth-order acoustic crossover often has just a second order electrical filter. This requires that speaker drivers be well behaved a considerable way from the nominal crossover frequency, and further that the high frequency driver be able to survive a considerable input in a frequency range below its crossover point. This is difficult in actual practice. In the discussion below, the characteristics of the electrical filter order is discussed, followed by a discussion of crossovers having that acoustic slope and their advantages or disadvantages.

Most audio crossovers use first to fourth order electrical filters. Higher orders are not generally implemented in passive crossovers for loudspeakers, but are sometimes found in electronic equipment under circumstances for which their considerable cost and complexity can be justified.

First order

First-order filters have a 20 dB/decade (or 6 dB/octave) slope. All first-order filters have a Butterworth filter characteristic. First-order filters are considered by many audiophiles to be ideal for crossovers. This is because this filter type is 'transient perfect', meaning it

passes both amplitude and phase unchanged across the range of interest. It also uses the fewest parts and has the lowest insertion loss (if passive). A first-order crossover allows more signals of unwanted frequencies to get through in the LPF and HPF sections than do higher order configurations. While woofers can easily take this (aside from generating distortion at frequencies above those they can properly handle), smaller high frequency drivers (especially tweeters) are more likely to be damaged since they are not capable of handling large power inputs at frequencies below their crossovers.

In practice, speaker systems with true first order acoustic slopes are difficult to design because they require large overlapping driver bandwidth, and the shallow slopes mean that non-coincident drivers interfere over a wide frequency range and cause large response shifts off-axis.

Second order

Second-order filters have a 40 dB/decade (or 12 dB/octave) slope. Second-order filters can have a Bessel, Linkwitz-Riley or Butterworth characteristic depending on design choices and the components used. This order is commonly used in passive crossovers as it offers a reasonable balance between complexity, response, and higher frequency driver protection. When designed with time aligned physical placement, these crossovers have a symmetrical polar response, as do all even order crossovers.

It is commonly thought that there will always be a phase difference of 180° between the outputs of a (second order) low-pass filter and a high-pass filter having the same crossover frequency. And so, in a 2-way system, the high-pass section's output is usually connected to the high frequency driver 'inverted', to correct for this phase problem. For passive systems, the tweeter is wired with opposite polarity to the woofer; for active crossovers the high-pass filter's output is inverted. In 3-way systems the mid-range driver or filter is inverted. However, this is generally only true when the speakers have a wide response overlap and the acoustic centers are physically aligned.

Third order

Third-order filters have a 60 dB/decade (or 18 dB/octave) slope. These crossovers usually have Butterworth filter characteristics; phase response is very good, the level sum being flat and in phase quadrature, similar to a first order crossover. The polar response is asymmetric. In the original D'Appolito MTM arrangement, a symmetrical arrangement of drivers is used to create a symmetrical off-axis response when using third-order crossovers.

Third-order acoustic crossovers are often built from first- or second-order filter circuits.

Fourth order

Fourth-order filters have an 80 dB/decade (or 24 dB/octave) slope. These filters are complex to design in passive form, as the components interact with each other. Steep-

slope passive networks are less tolerant of parts value deviations or tolerances, and more sensitive to mis-termination with reactive driver loads. A 4th order crossover with −6 dB crossover point and flat summing is also known as a Linkwitz-Riley crossover (named after its inventors), and can be constructed in active form by cascading two 2nd order Butterworth filter sections. The output signals of this crossover order are in phase, thus avoiding partial phase inversion if the crossover bandpasses are electrically summed, as they would be within the output stage of a multiband compressor. Crossovers used in loudspeaker design do not require the filter sections to be in phase: smooth output characteristics are often achieved using non-ideal, asymmetric crossover filter characteristics. Bessel, Butterworth and Chebyshev are among the possible crossover topologies.

Such steep-slope filters have greater problems with overshoot and ringing but there are several key advantages, even in their passive form, such as the potential for a lower crossover point and increased power handling for tweeters, together with less overlap between drivers, dramatically reducing lobing, or other unwelcome off-axis effects. With less overlap between adjacent drivers, their location relative to each other becomes less critical and allows more latitude in speaker system cosmetics or (in car audio) practical installation constraints.

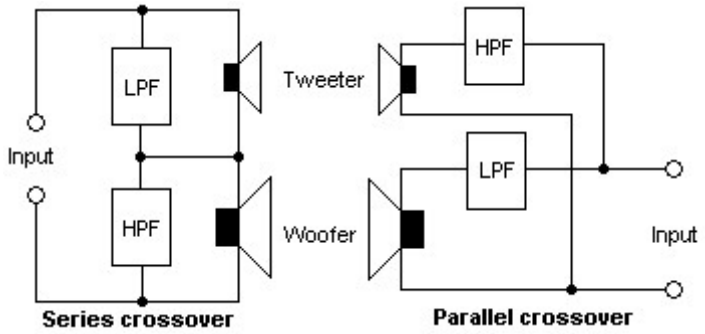
Higher order

Passive crossovers giving acoustic slopes higher than fourth-order are not common because of cost and complexity. Filters of up to 96 dB per octave are available in active crossovers and loudspeaker management systems.

Mixed order

Crossovers can also be constructed with mixed order filters. For example, a second order lowpass combined with a third order highpass. These are generally passive and are used for several reasons, often when the component values are found by computer program optimization. A higher order tweeter crossover can sometimes help compensate for the time offset between the woofer and tweeter, caused by non aligned acoustic centers.

Classification based on circuit topology



Series and parallel crossover topologies. *The HPF and LPF sections for the series crossover are interchanged with respect to the parallel crossover since they appear in shunt with the low & high frequency drivers.*

Parallel

Parallel crossovers are by far the most common. Electrically the filters are in parallel and thus the various filter sections do not interact. This makes two-way crossovers easier to design because the sections can be considered separately, and because component tolerance variations will be isolated. In the years before computer modeling, three-way crossovers were designed using the same value, but the advent of iterative design software has taught that this old technique creates excess gain and a 'haystack' response in the midrange output, together with a lower than anticipated input impedance.

Series

In this topology, the individual filters are connected in series, and a driver or driver combination is connected in parallel with each filter. To understand the signal path in this type of crossover, refer to the "Series Crossover" figure, and consider a high frequency signal that, during a certain moment, has a positive voltage on the upper Input terminal compared to the lower Input terminal. The low pass filter (LPF) presents a high impedance to the signal, and the tweeter presents a low impedance; so the signal passes through the tweeter. The signal continues to the connection point between the woofer and the high pass filter (HPF). There, the HPF presents a low impedance to the signal, so the signal passes through the HPF, and appears at the lower Input terminal. A low frequency signal with a similar instantaneous voltage characteristic first passes through the LPF, then the woofer, and appears at the lower Input terminal.

Derived

Derived crossovers include active crossovers in which one of the crossover responses is derived from the other through the use of a differential amplifier. For example, the difference between the input signal and the output of the high pass section is a low pass response. Thus, when a differential amplifier is used to extract this difference, its output constitutes the low pass filter section. The main advantage of derived filters is that they produce no phase difference between the high pass and low pass sections at any frequency. The disadvantages are either

- (a) that the high pass and low pass sections often have different levels of attenuation in their stop bands, *i.e.* their slopes are asymmetrical, or
- (b) that the response of one or both sections peaks near the crossover frequency,

or both. In case (a), above, the usual situation is that the derived low pass response attenuates at a much slower rate than the fixed response. This requires the speaker to which it is directed to continue to respond to signals deep into the stopband where its physical characteristics may not be ideal. In the case of (b), above, both speakers are

required to operate at higher volume levels as the signal nears the crossover points. This uses more amplifier power and may drive the speaker cones into non-linearity.

Chapter 3

Audio Equipment

A piece of **audio equipment** is any device designed principally to reproduce, record or process sound. This includes: -

Effects unit

A pedalboard allows a performer to create a ready-to-use chain of multiple pedals.

Effects units are electronic devices that alter how a musical instrument or other audio source sounds. Some effects subtly "color" a sound, while others transform it dramatically. Effects can be used during live performances (typically with electric guitar, keyboard, or bass) or in the studio. While most frequently used with electric or electronic instruments, effects can also be used with acoustic instruments and drums. Examples of common effects units include wah-wah pedals, fuzzboxes, and reverb units.

Effects units come in several formats, the most common of which are the "stompbox" and the "rackmount". A stompbox (or "pedal") is a small metal or plastic box placed on the floor in front of the musician and connected to his or her instrument. The box is typically controlled by one or more foot-pedal on-off switches and contains only one or two effects. A rackmount is mounted on a standard 19-inch equipment rack and usually contains several different types of effects.

While there is currently no consensus on how to categorize effects, the following are six common classifications: dynamics, time-based, tone, filter, pitch/frequency and feedback/sustain.

Formats (form factor)

Effects units are available in a variety of formats or "form factors". A musician's choice of form factor is generally determined by the instrument he or she plays, the musical situation (recording or live performance) and what he or she can afford. Stompbox style pedals are usually the smallest, least expensive and most rugged type of effect. Rackmount devices are relatively expensive and offer a wider range of functions. An effects unit can consist of analog or digital circuitry. During a live performance, the effect is plugged into the electrical "signal" path of the instrument. In the studio, the instrument or other sound-source's auxiliary output is patched into the effect. Form factors are part of a studio or musician's outboard gear.

Stompboxes

Stompboxes, or effects pedals, are effects units designed to sit on the floor or a pedalboard and be turned on and off with the user's feet. They typically house a single effect. The simplest stompbox pedals have a single footswitch; one or two potentiometers for controlling the effect, gain, or tone; and a single LED display to indicate whether the effect is on. The most complex stompbox pedals have multiple footswitches, eight to ten knobs, additional switches, and an alphanumeric display screen that indicates the status of the effect with short acronyms (e.g. DIST for "distortion").

An "effects chain" or "signal chain" may be formed by connecting two or more stompboxes. Effect chains are typically created between a preamplifier (“preamp”) and the guitar amplifier. When a pedal is off or inactive, the electrical signal coming in to the pedal is diverted onto a bypass, resulting in a "dry" signal which continues on to other effects down the chain. In this way, the effects within a chain can be combined in a variety of ways without having to reconnect boxes during a performance. A "controller" or "effects management system" allows for multiple effect chain loops to be created, so that one or several effects can be engaged or disengaged by tapping just one switch. The switches are usually organized in a row or a simple grid.

To preserve the clarity of the tone, it is most common to put compression, wah and overdrive pedals at the start of the chain; pitch/frequency pedals (chorus, flanger, phase shifter) in the middle; and time-based units (delay/echo, reverb) at the end. When using

many effects, unwanted noise and hum can be introduced into the sound. Some performers use a noise gate pedal at the end to reduce unwanted noise and hum introduced by overdrive units or vintage gear.

Rackmounts

Rackmounted effects are commonly used in recording studios and "front of house" live sound mixing situations. They are typically controlled by knobs or switches on their front panel, and often by a MIDI digital control interface. Rackmounts are built into a case designed to integrate into a 19-inch rack standard to the telecommunication and computing industries. "Shock mount" racks are designed for musicians who are shipping gear on major tours. Devices that are less than 19 inches wide may use special "ear" adapters that allow them to be mounted on a rack.

Built-in units

A vintage Teisco amplifier with built-in tremolo and echo effects

Effects are often incorporated into amplifiers and even some types of instruments. Electric guitar amplifiers typically have built-in reverb and distortion, while acoustic guitar and keyboard amplifiers tend to only have built-in reverb. Since the 2000s, guitar amplifiers began having built-in multi-effects units or digital modeling effects. Bass

amplifiers are less likely to have built-in effects, although some may have a compressor/limiter or distortion. Instruments with built-in effects include Hammond organs, electronic organs, and electronic pianos. Occasionally, acoustic-electric and electric guitars will have built-in effects.

Multi-effects devices

A multi-effects device (also called a "multi-FX" device) is a single electronics effects pedal or rackmount device that contains many different electronic effects. Multi-FX devices allow users to "preset" combinations of different effects, allowing musicians quick on-stage access to different effects combinations.

Tabletop units

A tabletop unit sits on a desk and is controlled manually. One such example is the Pod guitar amplifier modeler. Digital effects designed for DJs are often sold in tabletop models, so that the units can be placed alongside a mixer, turntables and CD scratching gear.

History

The earliest sound effects were strictly studio productions. In the mid to late 1940s, recording engineers and experimental musicians such as Les Paul began manipulating reel-to-reel recording tape to create echo effects and unusual, futuristic sounds. Microphone placement (“miking”) techniques were used in spaces with specially designed acoustic properties to simulate echo chambers.

Amplifier built-ins were the first effects to be used regularly outside the studio by guitar players. From the late 1940s onward, the Gibson Guitar Corp. began including vibrato circuits in combo amplifiers. The 1950 Ray Butts EchoSonic amp was the first to feature the "slapback” echo sound, which quickly became popular with guitarists such as Chet Atkins, Carl Perkins, Scotty Moore, Luther Perkins, and Roy Orbison. By the 1950s, tremolo, vibrato and reverb were available as built-in effects on many guitar amplifiers. Both Premier and Gibson built tube-powered amps with spring reverb. Fender began manufacturing the tremolo amps Tremolux in 1955 and Vibrolux in 1956.

Distortion was not an effect originally intended by amplifier manufacturers, but could often easily be achieved by “overdriving” the power supply in early tube amplifiers. Guitarists Johnny Burnette and Willie Johnson were among the first to deliberately increase gain beyond its intended levels to achieve "warm" distorted sounds. Dave Davies of The Kinks doctored the speakers of his amp by slitting them with a razor blade to achieve an even grittier guitar sound on the 1964 song "You Really Got Me". In 1965, Marshall Amplification began selling the Marshall 1959, a guitar amplifier capable of producing the warm overtones and distorted "crunch" that rock musicians were starting to covet.

Stand-alone effects of the 1950s and early 60s such as the Gibson GA-VI vibrato unit and the Fender reverb box, were expensive and impractical, requiring bulky transformers and high voltages. The original stand-alone units were not especially in-demand as many effects came built into amplifiers. The first popular stand-alone was the 1958 Watkins Copicat, a relatively portable tape echo effect made famous by the British band, The Shadows.

The electronic transistor finally made it possible to cram the aural creativity of the recording studio into small, highly portable stompbox units. Transistors replaced vacuum tubes, allowing for much more compact formats and greater stability. The first transistorized guitar effect was the 1962 Maestro Fuzz Tone pedal, which became a sensation after its use in the 1965 Rolling Stones hit “(I Can't Get No) Satisfaction”.

Warwick Electronics manufactured the first wah-wah pedal, The Clyde McCoy, in 1967 and that same year Roger Mayer issued the first octave effect, the Octavia. In 1968, Univox began marketing its Uni-Vibe pedal, an effect designed by noted audio engineer Fumio Mieda that mimicked the odd phase shift and chorus effects of the Leslie rotating speakers used in Hammond organs. The pedals soon became favorite effects of guitarists Jimi Hendrix and Robin Trower. Upon first hearing the Octavia, Hendrix allegedly rushed back to the studio and immediately used it to record the guitar solos on "Purple Haze" and "Fire" By the mid-1970s a variety of solid-state effects pedals including flangers, chorus pedals, ring modulators and phase-shifters were available.

In the 1980s, digitized rackmount units began replacing stompboxes as the effects format of choice. Often musicians would record "dry", unaltered tracks in the studio and effects would be added in post-production. The success of Nirvana's 1991 album *Nevermind* helped to re-ignite interest in stompboxes. Throughout the 1990s, musicians committed to a "lo-fi" aesthic such as J Mascis of Dinosaur Jr., Stephen Malkmus of Pavement and Robert Pollard of Guided by Voices continued to use non-digital (analog) effects pedals.

Types

While there is currently no consensus on how to categorize effects, the following are six common classifications: dynamics, time-based, tone, filter, pitch/frequency and feedback/sustain.

Dynamics

A Guyatone VT2 Vintage Tremolo

Clean boost/Volume pedal: A clean boost amplifies the volume of an instrument by increasing some aspect of its electrical signal output. These units are generally used for “boosting” volume during solos and preventing signal loss in long "effects chains". A guitarist switching from rhythm guitar to lead guitar may use a clean boost to increase the volume of his or her solo.
Volume effects: Fender Volume Pedal, Morley Volume Pedal.

Microphone preamplifier: A microphone preamplifier or "mic preamp" is a device that increases a microphone's low voltage output to levels that can be picked up and used by

equipment such as mixing consoles and headphones. Some mic pre-amps also provide additional power (e.g. phantom power) to condenser microphones.

Compressor: A compressor stabilizes volume and smooths a note's "attack" by dampening its onset and amplifying its sustain. Compression is achieved by varying the strength (i.e. "gain") of a signal to ensure volume stays within a specific dynamic range. A compressor can also function as a limiter with extreme settings of its controls.
Compressor effects: Boss CS-3, Keeley Compressor, MXR Dyna Comp.

Tremolo: A tremolo effect produces a slight, rapid variation in the volume of a note or chord. Tremolo effects normally have a "rate" knob which allows a performer to change the speed of the variation. The "tremolo effect" should not be confused with the misleadingly-named "tremolo bar", a device on a guitar bridge which allows the player to create a vibrato or "pitch-bending" effect. The guitar intro in the Rolling Stones' "Gimme Shelter" features a tremolo effect.
Tremelo effects: Fender Tremolux, Roger Mayer Voodoo Vibe.

Tone

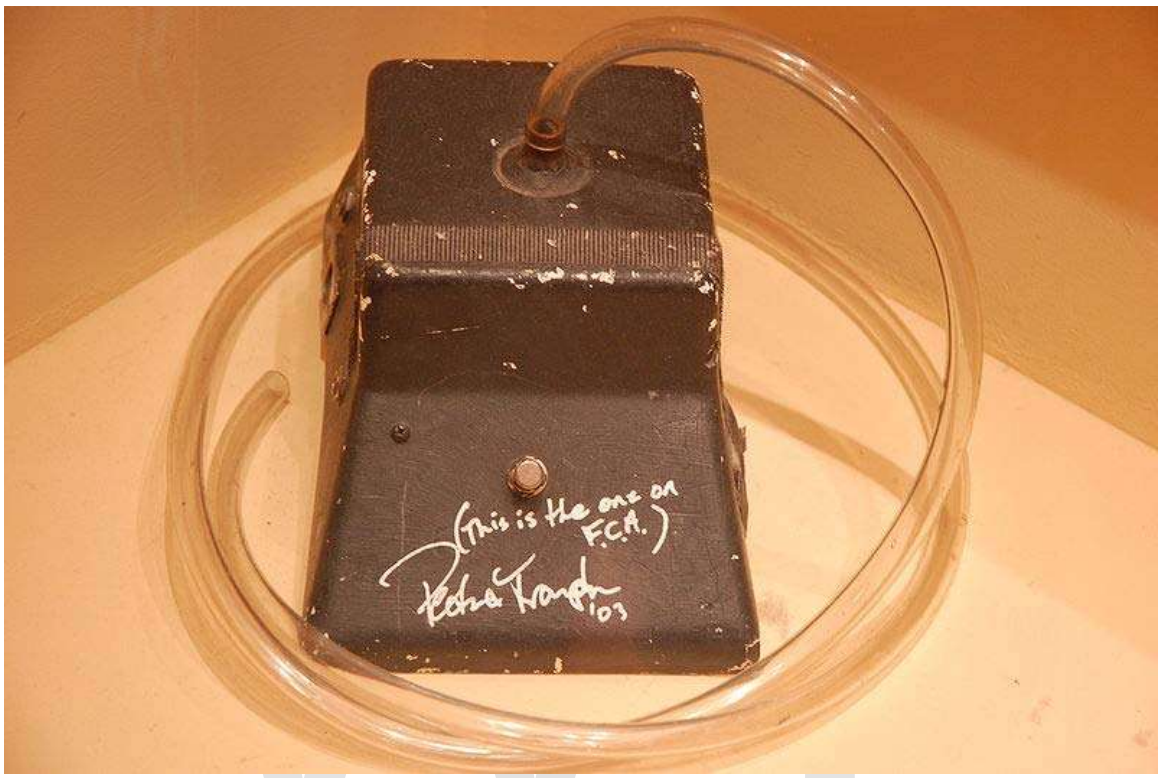
Distortion and Overdrive: Distortion and overdrive units distort the tone of an instrument by adding "overtones", creating a "warm" sound. To create a "dirty" or "gritty" sound, a unit further alters the tone by re-shaping or "clipping" its sound-waves so that they have flat, mesa-like peaks instead of curved ones. In tube amplifiers, distortion is created by compressing the instrument's out-going electrical signal in vacuum tubes or "valves". In digital units, this effect is simulated by transitors or computer chips. Distortion effects differ from overdrive effects in that that the former produces roughly the same amount of distortion at any volume. Overdrive units, on the other hand, produce "clean" sounds at quieter volumes and distorted sounds at louder volumes.
Distortion and overdrive effects: Boss DS-1, Boss MT-2 Metal Zone, Electro-Harmonix LPB-1, Ibanez Tube Screamer, Marshall ShredMaster, MXR Distortion+, MXR Micro Amp, Pro Co RAT.

Fuzz: A fuzz pedal or "fuzzbox" is a type of overdrive pedal that clips a sound-wave until it is nearly a squarewave, resulting in a heavily distorted or "fuzzy" sound. The Rolling Stones' "(I Can't Get No) Satisfaction" greatly popularized the use of fuzz effects.
Fuzz effects: Electro-Harmonix Big Muff, Arbiter Fuzz Face, Maestro Fuzz-Tone, Vox Tone Bender, Univox Super-Fuzz, Z.Vex Fuzz Factory.

Noise gate: Noise gates reduce "hum", "hiss" and "static" by eliminating sounds below a certain gain threshold. This significantly reduces noise as well as any other sounds coming into the unit (the "lo-fi" unit does the exact opposite, adding noise, hiss, and static). If it is used with extreme settings along with reverb, it can create unusual sounds, such as the gated drum effect used in 1980s pop songs, a style popularized by the Phil Collins song "In the Air Tonight".

Lo-fi: Lo-fi effects emulate the hiss, static, and poor tone quality of vintage analog electronic equipment.

Filter



Peter Frampton's Talk box

Equalizer: An equalizer is a set of filters that strengthen ("boost") or weaken ("cut") specific frequency regions. Stereos often have equalizers that adjust bass and treble. Audio engineers use highly sophisticated equalizers to eliminate unwanted sounds, make an instrument or voice more prominent, and enhance particular aspects of an instrument's tone.

Talk box: A talk box directs the sound from a guitar or synthesizer into the mouth of a performer, allowing him or her to shape the sound into vowels and consonants. The modified sound is then picked up by a microphone. In this way the guitar is able to "talk". Some famous uses of the talkbox include Bon Jovi's "Living on a Prayer", Stevie Wonder's "Black Man" and Peter Frampton's "Show Me the Way".
Talk boxes: Dunlop HT1 Heil Talk Box, Rocktron Banshee.

Wah-wah: A wah-wah pedal creates vowel-like sounds by altering the frequency spectrum produced by an instrument—i.e. how loud it is at each separate frequency—in what is known as a spectral glide. The device is operated by a foot treadle that opens and closes a potentiometer. Wah-wah pedals are often used by funk and psychedelic rock

guitarists.
Wah effects: Dunlop Cry Baby, Morley Power Wah Boost, Musitronics Mu-Tron III, Z.Vex Seek Wah.

De-esser: A de-esser filters out the higher-frequency sounds produced by sibilant consonants such as “s”, “z”, and “sh” in recordings of the human voice.

Pitch/Frequency

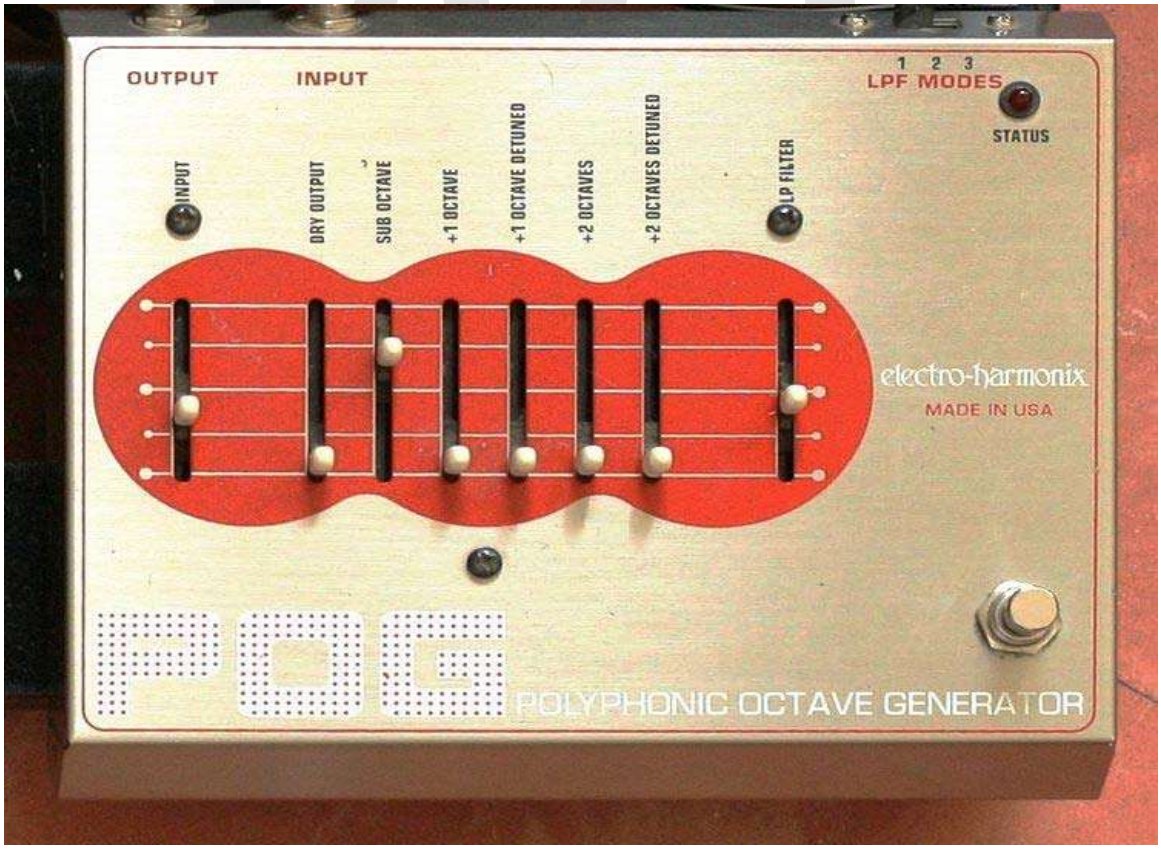
SmallClone chorus effect

Chorus: Chorus pedals mimic the "phase locking" effect produced naturally by choirs and string orchestras when sounds with very slight differences in timbre and pitch

assimilate one another. A chorus effect splits the instrument-to-amplifier electrical signal, adding slight frequency variations or “vibrato” to part of the signal while leaving the rest unaltered. With extreme settings, a chorus effect can produce a "spacey" sound. A well-known usage of chorus is the lead guitar in “Come As You Are” by Nirvana.
Chorus effects: Boss CE-1 Chorus Ensemble, Electro-Harmonix Deluxe Memory Man, Electro-Harmonix Electric Mistress, Roger Mayer Voodoo Vibe, T.C. Electronic Stereo Chorus.

Flanger: A flanger creates a "jet plane" or "spaceship" sound, simulating a studio effect produced by holding the edge of the audio tape reel (the “flange”) to momentarily slow down a recording. Flangers add a variably delayed version of the sound to the original or sound, creating a comb filter effect. Some famous uses of flanger effects include "Walking on the Moon" by The Police and "Barracuda" by Heart.
Flanger effects: Electro-Harmonix Electric Mistress, MXR Flanger.

Phase shifter: A phase shifter creates a slight rippling effect—amplifying some aspects of the tone while diminishing others—by adding out-of-phase duplicate sound-waves to the original sound-waves. Phase shifting was popular during the 1970s; two well-know examples includes keyboard parts on Billy Joel’s “Just the Way You Are” and Paul Simon's "Slip Slidin' Away".



The Electro-Harmonix POG pedal can pitch-shift an input signal down an octave or up one or two octaves.

Phase shift effects: Electro-Harmonix Small Stone, MXR Phase 90, Roland AP-7 Jet Phaser.

Pitch shifter and Harmonizer: A pitch shifter raises or lowers (e.g. "transposes") each note a performer plays by a pre-set interval. For example, a pitch shifter set to increase the pitch by a fourth will raise each note four diatonic intervals above the notes actually played. Simple pitch shifters raise or lower the pitch by one or two octaves, while more sophisticated devices offer a range of interval alterations. A harmonizer is a type of pitch shifter that combines the altered pitch with the original pitch to create a two or more note harmony. Some hamonizers are able to create chorus-like effects by adding very tiny shifts in pitch.
Pitch shift effects: Electro-Harmonix POG, Digitech Whammy, Roger Mayer Octavia .

Ring modulator: A ring modulator produces a resonant, metallic sound by mixing a waveform produced by the instrument with a waveform generated by the device's internal oscillator to create signals rich in overtones. A notable use of ring modulation is the guitar in the Black Sabbath song "Paranoid".
Ring modulator effects: Moog MF-102 Moogerfooger.

Vibrato: Vibrato effects produce slight, rapid variations in pitch, mimicking the fractional semitone variations produced naturally by opera singers and violinists when prolonging a single note. Vibrato effects often allow the performer to control the rate of the variation as well as the difference in pitch (e.g. “depth”). A vibrato with an extreme "depth" setting (e.g., half a semitone or more) will produce a dramatic, ululating sound. Guitarists often use the terms "vibrato" and "tremolo" misleadingly. A so-called "vibrato unit" in a guitar amplifier actually produces tremolo, while a "tremolo arm" or “whammy bar” on a guitar produces vibrato.

Harmonic Exciter: A harmonic exciter or "aural exciter" or "psychoacoustic exciter", adds subtle overtones to the upper mid and treble part of a sound. Harmonic exciters are used most frequently in the post-production stage of recording, either with vocals or with an entire track. This effect was developed in the mid-1970s to add "brightness" to reel-to-reel audio tape recordings that had lost clarity due to compression or repeated overdubs.

Time-based

Folded line reverberation device, which uses springs

Delay/Echo: Delay/echo units produce an echo effect by adding a duplicate instrument-to-amplifier electrical signal to the original signal at a slight time-delay. The effect can either be a single echo called a “slap” or “slapback,” or multiple echos. Some well-known uses of delay are the lead guitar in the U2 song "Where the Streets Have No Name", and the main riff in Pink Floyd's "Run Like Hell".
Delay effects: Boss DM-2 Delay, Boss DD-3 Digital Delay, Electro-Harmonix 16-Second Digital Delay, Electro-Harmonix Memory Man, Line 6 DL4 Delay Modeler, MXR Carbon Copy.

Reverb: Reverb units simulate sounds produced in an echo chamber by creating a large number of echoes that gradually fade or "decay". A plate reverb system uses an electromechanical transducer to create vibrations in a plate of metal. Spring reverb systems, which are often used in guitar amplifiers, use a transducer to create vibrations in a spring. Digital reverb effects use various signal processing algorithms to create the reverb effect, often by using multiple feedback delay circuits. Rockabilly and surf guitar are two genres that make heavy use of reverb.
Reverb effects: Fender Reverb Unit.

Looper pedal: A looper pedal or "phrase looper" allows a performer to record and later replay a phrase or passage from a song. Loops can be created on the spot during a performance or they can be pre-recorded. Some units allow a performer to layer multiple loops. The first loop effects were created with reel-to-reel tape using a tape loop. High-end boutique tape loop effects are still used by some studios who want a vintage sound. Digital loop effects recreate this effect using an electronic memory.
Looper effects: Boss RC20XL Loop Station Pedal, Line 6 DL4 Delay Modeler Pedal and Loop Sampler.

An EBow allows a guitar player to sustain a note

Feedback/Sustain

Audio feedback: Audio feedback is an effect produced when amplified sound is picked up by a microphone and played back through an amplifier, initiating a “feedback loop”. Feedback as pioneered by guitarists such as Jimi Hendrix is generated by playing an instrument directly in front of an amplifier set to a high volume. This relatively primitive technique tends to create high-pitched overtones and can be difficult to sustain.

The EBow, a handheld pickup/string driver, uses a small inductor coil to vibrate a guitar's strings, creating a bow-like sustained sound. Devices such as the Guitar Resonator, the Sustainiac Sustainer, and the Fernandes Sustainer create feedback by electrically

vibrating (“driving”) the guitar strings while minimizing the highest-pitched overtones and providing true sustain.

Many compressor pedals are often also marketed as "sustainer pedals". As a note is sustained, it loses energy and volume due to diminishing vibration in the string. The compressor pedal boosts its electrical signal to the specified dynamic range, slightly prolonging the duration of the note.

Other effects

Simulators: Simulators enable electric guitars to mimic the sound of other instruments such as acoustic guitar, electric bass, and sitar. Pick up simulators used on guitars with single-coil pick ups replicate the sound of guitars with humbucker pick ups, or vice-versa. A de-fretter is a bass guitar effect that simulates the sound of a fretless bass. The effect uses an envelope-controlled filter and voltage controlled amplifier to “soften” a note's attack both in volume and timbre.

Envelope Follower: An envelope follower activates an effect once a designated volume is reached. One effect that uses an envelope follower is the "auto-wah", which produces a "wah" effect depending on how loud or soft the notes are being played.

Guitar amplifier modeling: Amplifier modeling is a digital effect that replicates the sound of various amplifiers, most often analog “tube” amps. Sophisticated modeling effects can simulate speaker cabinets and miking techniques. A rotary speaker simulator mimics the doppler sound of a vintage Leslie speaker system by replicating its volume and pitch modulations, overdrive capacity and phase shifts.

Pitch correction/Vocal effects: Pitch correction effects use signal-processing algorithms to re-tune faulty intonation in a vocalist's performance.

Filter and synthesizer effects: Pedals such as the Moog MF-105 Moogerfooger MURF provide multiple filters and envelope control knobs to control modulation. The MF-107 FreqBox uses the input signal to modulate an internal VCO oscillator.

Boutique pedals

T-Rex brand "Mudhoney" overdrive pedal

Boutique pedals are designed by smaller, independent companies and are typically produced in limited quantities. Some may even be hand-made. These pedals are mainly distributed online or through mail-order, or sold in a few music stores. They are often more expensive than mass-produced pedals and offer non-standard features such as true-bypass switching, higher-quality components, innovative designs, and hand-painted artwork. Some boutique companies focus on re-creating classic or vintage effects. Some boutique pedal manufacturers include: AnalogMan, Devi Ever, Pete Cornish, Lovetone, Metasonix, Robert Keeley, Z.Vex Effects, T-Rex Engineering.

Effects unit modification

There is also a niche market for modifying or "modding" effects. Typically, vendors provide either custom modification services or sell new effects pedals which have been modified. The Ibanez Tube Screamer, the Boss DS-1, the ProCo Rat and Digitech Whammy are some of the most often-modified effects. Common modifications include value changes in capacitors or resistors, adding true-bypass so that the effect's circuitry is no longer in the signal path, substituting higher-quality components, replacing the unit's original operational amplifiers (opamps), or adding functions to the device such as allowing additional control of some factor or adding an additional output jack.

Tributes by musicians

The garage rock revival band The Fuzztones, seen here in a Barcelona concert, are named after an influential 1960s-era fuzz pedal (the Fuzztone).

Effects and effects units—stompboxes in particular—have been celebrated by pop and rock musicians in album titles, songs and band names. The Big Muff, a classic fuzzbox manufactured by Electro-Harmonix, is commemorated by the Depeche Mode song "Big Muff" and the Mudhoney EP *Superfuzz Bigmuff*. Lyrics to Super Furry Animals' "Play It Cool" mention another Electro-Harmonix pedal, the Electric Mistress flanger. The Nine Inch Nails song "Echoplex" is titled after Maestro's vintage echo unt. Other songs that reference effects include "Interstellar Overdrive" by Pink Floyd, "Wah-Wah" by George Harrison, and "Stomp Box" by They Might Be Giants. Joy Division's "Digital" was inspired by engineer/producer Martin Hannett's AMS digital delay unit. We've Got a Fuzzbox and We're Gonna Use It were an all-female British band from the 1980s, and The Fuzztones were a 1980s garage rock revival band.

Other pedals and rackmount units

Not all stompboxes and rackmounts are effects. Tuning pedals indicate whether a guitar string is too sharp or flat. A footswitch pedal such as the "A/B" pedal route a guitar signal to an amplifier or enable a performer to switch between two guitars. Guitar amplifiers and electronic keyboards may have switch pedals for turning built-in effects on and off. Some musicians who use rackmounted effects or laptops employ a MIDI controller pedalboard to trigger sound samples, switch between different effects or control effect settings.

Chapter 4

Mixing Console

In professional audio, a **mixing console**, or **audio mixer**, also called a **sound board**, **mixing desk**, or **mixer** is an electronic device for combining (also called "mixing"), routing, and changing the level, timbre and/or dynamics of audio signals. A mixer can mix analog or digital signals, depending on the type of mixer. The modified signals (voltages or digital samples) are summed to produce the combined output signals.

Mixing consoles are used in many applications, including recording studios, public address systems, sound reinforcement systems, broadcasting, television, and film post-production. An example of a simple application would be to enable the signals that originated from two separate microphones (each being used by vocalists singing a duet, perhaps) to be heard through one set of speakers simultaneously. When used for live performances, the signal produced by the mixer will usually be sent directly to an amplifier, unless that particular mixer is "powered" or it is being connected to powered speakers.

BBC Local Radio Mark III radio mixing desk

Structure

Yamaha 2403 audio mixing console in a 'live' mixing application

A typical analog mixing board has three sections:

- Channel inputs
- Master controls
- Audio level metering

The channel inputs are replicated monaural or stereo input channels with pre-amp controls, channel fader and pan, sub-group assignment, equalization and auxiliary mixing bus level controls. The master control section has sub-group faders, master faders, master auxiliary mixing bus level controls and auxiliary return level controls. In addition it may have solo monitoring controls, a stage talk-back microphone control, muting controls and an output matrix mixer. On smaller mixers the inputs are on the left of the mixing board and the master controls are on the right. In larger mixers, the master controls are in the center with inputs on both sides. The audio level meters may be above the input and master sections or they may be integrated into the input and master sections themselves.

Channel input strip

The input strip is usually separated into these sections:

- Input jacks / microphone preamplifiers
- Basic input controls
- Channel EQ (High, Mids and low)
- Routing Section including Direct Outs, Aux-sends, Panning control and Subgroup assignments
- Input Faders

On the Yamaha Console above, these sections are color coded for quick identification by the operator. Each signal that is input into the mixer has its own *channel*. Depending on

the specific mixer, each channel is stereo or monaural. On most mixers, each channel has an XLR input, and many have RCA or quarter-inch Jack plug line inputs.

Basic input controls

Below each input, there are usually several rotary controls (knobs, pots). The first is typically a *trim* or *gain* control. The inputs buffer the signal from the external device and this controls the amount of amplification or attenuation needed to bring the signal to a nominal level for processing. This stage is where most noise of interference is picked up, due to the high gains involved (around +50 dB, for a microphone). Balanced inputs and connectors, such as XLR or Tip-Ring-Sleeve (TRS) quarter-inch connectors, reduce interference problems.

There may be *insert* points after the buffer/gain stage, which send to and return from external processors which should only affect the signal of that particular channel. Insert points are most commonly used with effects that control a signal's amplitude, such as noise gates, expanders, and compressors.

Auxiliary send routing

The *Auxiliary send* routes a split of the incoming signal to an auxiliary bus which can then be used with external devices. *Auxiliary sends* can either be pre-fader or post-fader, in that the level of a pre-fade send is set by the *Auxiliary send* control, whereas post-fade sends depend on the position of the channel fader as well. *Auxiliary sends* can be used to send the signal to an external processor such as a reverb, which can then be routed back through another channel or designated auxiliary returns on the mixer. These will normally be post-fader. Pre-fade *auxiliary sends* can be used to provide a monitor mix to musicians onstage, this mix is thus independent of the main mix.

Allen & Heath Mixing desk used for live performances

Channel equalization

Further channel controls affect the equalization (EQ) of the signal by separately attenuating or boosting a range of frequencies, e.g., bass, midrange, and treble. Most large mixing consoles (24 channels and more) usually have sweep equalization in one or more bands of its parametric equalizer on each channel, where the frequency and affected bandwidth of equalization can be selected. Smaller mixing consoles have few or no equalization controls. Care must be taken not to add too much EQ to a signal that is already close to clipping; additional energy will overdrive the channel.

Some mixers have a general equalization control (either graphic or parametric) at the output.

Subgroup and mix routing

Each channel on a mixer has an audio taper pot, or potentiometer, controlled by a sliding volume control (*fader*), that allows adjustment of the level, or amplitude, of that channel in the final *mix*. A typical mixing console has many rows of these sliding volume controls. Each control adjusts only its respective channel (or one half of a stereo channel); therefore, it only affects the level of the signal from one microphone or other audio device. The signals are summed to create the main *mix*, or combined on a *bus* as a submix, a group of channels that are then added to get the final mix (for instance, many drum mics could be grouped into a bus, and then the proportion of drums in the final mix can be controlled with one bus fader).

There may also be *insert* points for a certain bus, or even the entire mix.

Master output controls

Subgroup and main output fader controls are often found together on the right hand side of the mixer or, on larger consoles, in a center section flanked by banks of input channels. Matrix routing is often contained in this master section, as are headphone and local loudspeaker monitoring controls. Talkback controls allow conversation with the artist through their wedges, headphones or IEMs (in-ear monitor). A test tone generator might be located in the master output section. Aux returns such as those signals returning from outboard reverb devices are often in the master section.

Metering

Finally, there are usually one or more VU or peak meters to indicate the levels for each channel, or for the master outputs, and to indicate whether the console levels are overmodulating or clipping the signal. Most mixers have at least one additional output, besides the main mix. These are either individual bus outputs, or *auxiliary outputs*, used, for instance, to output a different mix to on-stage monitors. The operator can vary the mix (or levels of each channel) for each output.

As audio is heard in a logarithmic fashion (both amplitude and frequency), mixing console controls and displays are almost always in decibels, a logarithmic measurement system. This is also why special audio taper pots or circuits are needed. Since it is a relative measurement, and not a unit itself (like a percentage), the meters must be referenced to a nominal level. The "professional" nominal level is considered to be +4 dBu. The "consumer grade" level is −10 dBV.

Hardware routing and patching

For convenience, some mixing consoles include inserts or a patch bay or patch panel. Patch bays are mainly used for recording mixers.

Other features

Most, but not all, audio mixers can

- add external effects.
- use monaural signals to produce stereo sound by adjusting the position of each signal on the sound stage (pan and balance controls).
- provide phantom power (typically 48 volts) required by some microphones.
- create an audible tone via an oscillator, usually at 440 Hz, 1 kHz, or 2 kHz

Some mixers can

- add effects internally.
- read and write console automation.
- be interfaced with computers or other recording equipment (to control the mixer with computer presets, for instance).
- control or be controlled by a Digital Audio Workstation via Midi or proprietary commands.
- be powered by batteries.

Digital versus analog

Digidesign's Venue Profile mixer on location at a corporate event. This digital mixer allows plugins from third-party vendors

Digital mixing console sales have increased dramatically since their introduction in the 1990s. Yamaha sold more than 1000 PM5D mixers by July, 2005, and other manufacturers are seeing increasing sales of their digital products. Digital mixers are more versatile than analog ones and offer many new features, such as the ability to save multiple mute groups, multiple VCA groups and channel settings into a scene and reconfigure signal routing at the touch of a button. The faders can be "swapped" or "flipped" to show aux send levels; a feature very useful in mixing artists' monitors. In addition, digital consoles often include a range of special effects such as parametric EQ, compression, gating, reverb, automatic feedback reduction, tap delay and straight delay. Some products are expandable via third-party software features (called plugins) that add further reverb, compression, delay and tone-shaping tools. Several digital mixers include spectrograph and real time analyzer functions. A few incorporate loudspeaker management tools such as crossover filtering and limiting. Digital signal processing can perform automatic mixing for some simple applications, such as courtrooms, conferences and panel discussions, but at this time no digital mixer in live audio includes automixing. Consoles with motorized faders can read and write console automation.

Digital mixers can be designed to be quieter than most analog mixers, as digital mixers often incorporate very low threshold noise gates to stop inactive mix bus background hiss from summing with active signals. Digital circuitry is more resistant to outside interference from radio transmitters such as walkie-talkies and cell phones.

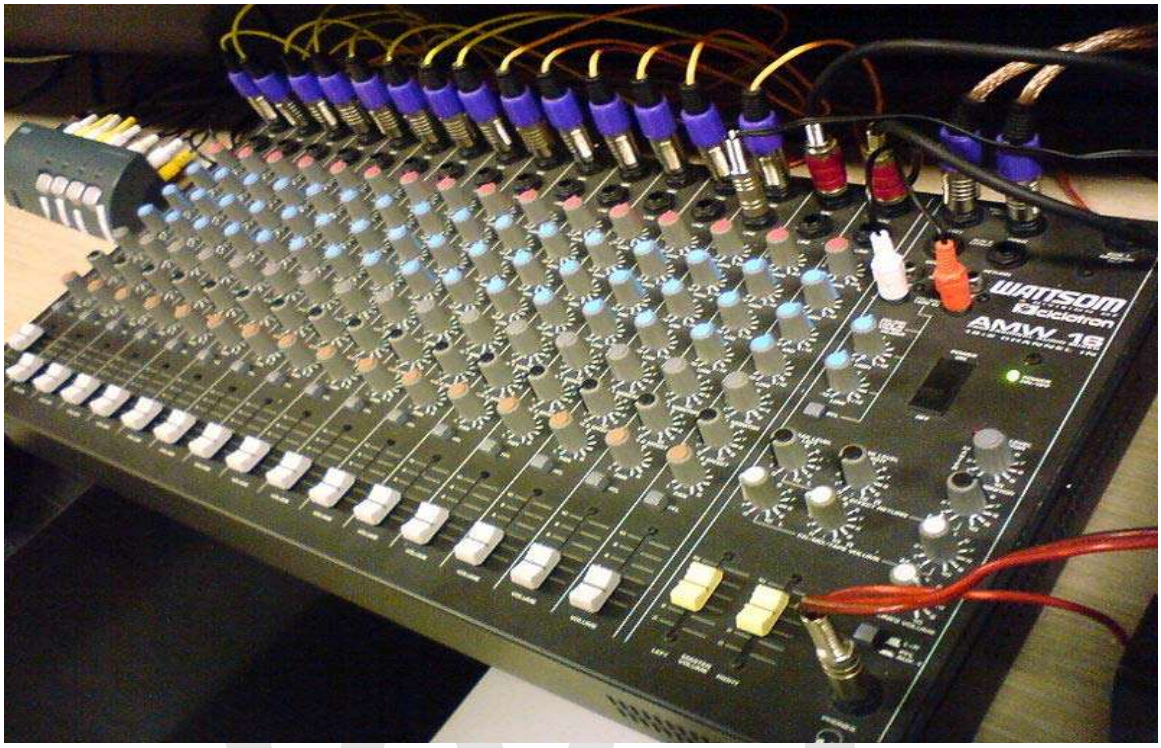
Propagation delay

Digital mixers have an unavoidable amount of latency or propagation delay, ranging from 1.5 ms to as much as 10 ms, depending on the model of digital mixer and what functions are engaged. This small amount of delay isn't a problem for loudspeakers aimed at the audience or even monitor wedges aimed at the artist, but can be disorienting and unpleasant for IEMs (In ear monitors) where the artist hears their voice acoustically in their head *and* electronically amplified in their ears but delayed by a couple of milliseconds.

Every analog to digital conversion and digital to analog conversion within a digital mixer entails propagation delay. Audio inserts to favorite external analog processors make for almost double the usual delay. Further delay can be traced to format conversions such as from ADAT to AES3 and from normal digital signal processing steps.

Within a digital mixer there can be differing amounts of latency, depending on the routing and on how much DSP is in use. Assigning a signal to two parallel paths with significantly different processing on each path can result in extreme comb filtering when recombined. Some digital mixers incorporate internal methods of latency correction so that such problems are avoided.

Ease of use



16-channel mixing console with compact short-throw faders

Analog consoles remain popular due to their continuing to have one knob, fader or button per function, a reassuring feature for the user. This takes up more physical space but allows more rapid response to changing performance conditions. Most digital mixers take advantage of the technology to reduce the physical space requirements of their product, entailing compromises in user interface such as a single shared channel adjustment area that is selectable for only one channel at a time. Additionally, most digital mixers have virtual pages or layers which change the fader banks into separate controls for additional inputs or for adjusting equalization or aux send levels. This layering can be confusing for operators.

Analog consoles make for simpler understanding of hardware routing. Many digital mixers allow internal reassignment of inputs so that convenient groupings of inputs appear near each other at the fader bank, a feature that can be disorienting for persons having to make a hardware patch change.

On the other hand, many digital mixers allow for extremely easy building of a mix from saved data. USB flash drives and other storage methods are employed to bring past performance data to a new venue in highly portable manner. At the new venue, the traveling mix technician simply plugs the collected data into the venue's digital mixer and quickly makes small adjustments to the local input and output patch layout, allowing for full show readiness in very short order.

Some digital mixers allow offline editing of the mix, a feature that lets the traveling technician use a laptop to make anticipated changes to the show while *en route*, further shortening the time it takes for the sound system to be ready for the artist.

Sound quality

Both digital and analog mixers rely on analog microphone preamplifiers, a high-gain circuit that is the origin of much of the perceived character of sound quality in an audio mixer. In this respect, both formats are on par with each other. In a digital mixer, the microphone preamplifier is followed by an ADC which quantizes the audio stream. Ideally, this process is carefully engineered to deal gracefully with overloading and clipping while delivering an accurate digital stream over the linear dynamic range. Further processing and mixing of digital streams within a mixer need to avoid clipping and truncation if maximum audio quality is desired.

Analog mixers, too, must deal gracefully with overloading and clipping at the microphone preamplifier and as well as avoiding overloading of mix buses. Background hiss in an analog mixer is always present, though good gain stage management minimizes its audibility. Idle subgroups left "up" in a mix will add their background hiss to the main outputs; many digital mixers avoid this problem by low-level gating.

Many electronic design elements combine to affect perceived sound quality, making the global "analog mixer vs. digital mixer" question difficult to answer. Controlled ABX double-blind listening tests have not been published at this date; no conclusive answer can be reached. Experienced live sound professionals agree that microphones and loudspeakers (with their innate higher distortion levels) are a much greater source of coloration of sound than the choice of mixer. The mix style of the person mixing is also more important than the make and model of audio console. Analog and digital mixers both have been associated with extremely high-quality concert performances and studio recordings.

Remote control

Analog mixing in live sound has had the option since the 1990s of using wired remote controls for certain digital processes such as monitor wedge equalization and parameter changes in outboard reverb devices. That concept has expanded until wired and wireless remote controls are being seen in relation to entire digital mixing platforms. It's possible to set up a sound system and mix via wireless (or wired) laptop, touchscreen or tablet, especially if the performance requires no unpredictable fast responses to multiple changing conditions on stage. Computer networks can connect digital system elements for expanded monitoring and control, allowing the system technician to make adjustments to distant devices during the performance. The use of remote control technology can be utilized to reduce "seat-kills", allowing more paying customers into the performance space.

Virtual mixing

Increasingly, the mixing process can be performed on screen, using computer software and associated input, output and recording hardware. The traditional large control surface of the mixing console is not utilized, saving space at the engineer's mix position. Some virtual mixing (such as the Gamble DCX) uses digital controls of analog audio circuitry, but most virtual mixers are fully digital so as to save cost and physical space. In the virtual studio, there is either no normal mixer fader bank at all or there is a compact group of motorized faders designed to fit into a small space and connected to the computer via USB or Firewire. Many project studios use such a space-efficient solution, as the mixing room at other times can serve as business office, media archival, etc. Virtual mixing is heavily integrated as part of a digital audio workstation.

Applications

A Behringer EuroRack UB1002FX in a DJ setup

Dub producers/engineers such as Lee "Scratch" Perry were perhaps the first musicians to use a mixing board as a musical instrument.

Public address systems will use a mixing console to set microphones for different speakers to the correct level, and can add in recorded sounds into the mix. A major requirement is to minimise audio feedback.

Most bands will use a mixing console to combine musical instruments and vocals to the correct level.

Radio broadcasts use a mixing desk to select audio from different sources, such as CD players, telephones, remote feeds, or prerecorded advertisements.

Noise music musicians such as Merzbow or Wolf Eyes may create feedback loops within mixers, creating an instrument known as a no-input mixer. The tones generated from a no-input mixer are created by connecting an output of the mixer into an input channel and manipulating the pitch with the mixer's dials.

Chapter 5

AV Receiver

AV receivers or *audio-video receivers* are one of the many consumer electronics components typically found within a home theatre system. Their primary purpose is to amplify sound from a multitude of possible audio sources as well as route video signals to your TV from various sources. The user may program and configure a unit to take inputs from devices such as DVD players, VCRs etc. and easily select which source he or she wants to route to his TV and have sound output for.

Usage

The term receiver originally referred to a component which included a tuner, a pre-amplifier and a power amplifier. These were generally called stereo receivers. The built in tuner in these devices gave them the name receivers.

As home entertainment options expanded, so did the role of the receiver. The ability to handle a variety of digital audio signals was added. More amplifiers were added for surround sound playback. Video switching was added to simplify switching. Within the last few years, video processing has been added to many receivers.

The term audio/video receiver (AVR) or Home Theater Receiver is used to distinguish the simpler stereo receiver from the multi-channel audio video receiver.

Features

Radio reception

Receivers usually have a built in tuner for AM and FM radio reception. Satellite radio tuners are also found in many modern receivers, allowing reception with just an external antenna (and a satellite radio subscription, if necessary).

Some models have HD Radio tuners.

Some models have Internet Radio and PC streaming access capabilities with an ethernet port.

Decoders

AV receivers usually provide one or more decoders for sources with more than two channels of audio information. This is most common with movie soundtracks. Movie soundtracks have been provided via a number of encoded formats. The first common format was Dolby Pro Logic. This format contained a center channel and surround channel. These channels were mixed into the left and right channels using a process called matrixing. Receivers were produced with Dolby Pro Logic decoders which could separate out these two additional channels.

With the introduction of the DVD, the Dolby Digital format became a standard. Dolby Digital ready receivers included inputs and amplifiers for the additional channels. Most current AV receivers provide a Dolby Digital decoder and at least one digital S/PDIF input which can be connected to a source which provides a Dolby Digital output.

A somewhat less common surround sound decoder called DTS is standard on current AV receivers.

When Dolby Labs and DTS introduced technologies to add a rear center surround channel, these technologies found their way into AV Receivers. Receivers with six amplifiers (known as 6.1 receivers) will typically have both Dolby and DTS's technologies. These are Dolby Digital EX and DTS ES.

Dolby introduced Dolby Pro Logic II to allow stereo sources to play back as if they were encoded in surround sound. DTS introduced a similar technology, NEO:6. These decoders have become common on most current receivers.

As the number of playback channels were increased on receivers, other decoders have been added to some receivers. For example, Dolby Labs created Dolby Pro Logic IIx to take advantage of receivers with more than five channels of playback.

With the introduction of high definition players (e.g. Blu-ray Disc and HD DVD), yet more decoders have been added to some receivers. Dolby TrueHD and DTS-HD Master Audio decoders are available on many receivers.

DSP effects

Most receivers offer specialized Digital Signal Processors (DSP) made for handling various presets and audio effects. Some may offer simple equalizers and balance adjustments to complex DSP audio field simulations such as "Hall", "Arena", "Opera", etc. that simulate the audio being played in the places through use of surround sound and echo effects.

Amplification

Stereo receivers have two channel of amplification, while AV receivers may have more than 2. The standard for AV receivers is five channels of amplification. These are usually referred to as 5.1 receivers. This provides for a left, right, center, left surround and right surround speaker to be powered by the receiver. 7.1 receivers are becoming more common and provide for two additional surround channels, left rear surround and right rear surround. The '.1' refers to the LFE (low frequency effects) channel the signal of which is usually sent to an amplified subwoofer unit. 5.1 and 7.1 receivers don't usually provide amplification for this channel. Instead, they provide a line level output.

There are various standards for rating the output power of receivers. Different countries have different rules on how manufacturers specify the output ratings. Its not always possible to use these ratings to compare two products. Due to a number of factors such as real world behavior of speakers and dynamic headroom its possible for an amplifier with a lower rated power to play more loudly than one with a higher rated power.

Differences in output power are not always as significant as they may look. It takes 10 times the output power for the sound to be perceived as twice as loud. If 1 watt of output yields a sound pressure level of 90dB, it takes 10 watts to get an SPL of 100dB and 100 watts to get an SPL of 110dB. A 110 watt amplifier will not play 10% louder than a 100 watt amplifier.

Most receivers use class AB amplifiers. Some manufacturers are now producing receivers using class D amplifiers. Class D amplifiers are more efficient and can be made smaller and lighter than an equivalent class AB amplifier. There are also other designs such as class G and class H. Class G and H are variations on the conventional class AB design. Class G has two sets of power supply rails. Normally the power amp is fed from the lower voltage supply. This helps keep power dissipation in the output transistors down. When the signal exceeds the lower supply voltage, the amp switches to the higher voltage supply so the signal can be reproduced without clipping. With a class H design, the supply rails are variable rather than two discrete steps. The signal actually modulates the supply voltage.

AV inputs/outputs

There are a variety of possible connections on an AV receiver. Standard connectors include:

- Analog audio (RCA Connector, or occasionally XLR connector)
- Digital audio (S/PDIF; TOSLINK or RCA terminated coaxial cable)
- Composite video (RCA connector)
- S-Video
- SCART video (primarily used in Europe and very uncommon in many other parts of the world)
- Component video

- HDMI

Analog audio connections usually use RCA plugs in stereo pairs. Inputs and outputs are both common. Outputs are provided mainly for cassette tape decks.

Analog audio connections using XLR connectors are uncommon, and usually found on more expensive receivers.

Digital connections allow for the transmission of PCM, Dolby Digital or DTS audio. Common devices include CD players, DVD players, or satellite receivers.

Composite video connections use a single RCA plug on each end. Composite video is standard on all AV receivers allowing for the switching of video devices such as VHS players, cable boxes, and game consoles. DVD players may be connected via composite video connectors although a higher bandwidth connection is recommended.

S-Video connections offer better quality than composite video. It uses a DIN jack.

SCART connections generally offer the best quality video at standard-definition, due to the use of pure RGB signalling (although composite and S-Video may alternatively be offered over a SCART connector). SCART provides video and audio in one connection.

Component video has become the best connection for analog video as higher definitions such as 720p have become common. The YPbPr signalling provides a good compromise between resolution and colour definition.

HDMI is becoming common on AV receivers. It provides for the transmission of both audio and video. HDMI is relatively new technology and there are reported issues with devices not properly working with each other, especially cable/satellite boxes connected to a display through an AV receiver. Different levels of support are provided by receivers with HDMI connections. Some will only switch video and not provide for audio processing. Some will not handle multi-channel LPCM. Multi-channel LPCM is a common way for Blu-ray and HD DVD players to transmit the best possible audio.

Video conversion and upscaling

Some receivers can convert from one video format to another. This is commonly called upconversion or transcoding. A smaller number of receivers provide for de-interlacing of video signals. For example, a receiver with upconversion, deinterlacing and upscaling can take an interlaced composite signal at 480i (480 lines per frame sent as a field of 240 even numbered lines 0,2,4,8...478 followed by a field of 240 odd numbered lines 1,3,5,...479) and convert it to component video while also deinterlacing and upscaling it to a higher resolution such as 720p (720 lines per frame with all lines in normal sequence 0,1,2...719).

Chapter 6

Tape Recorder

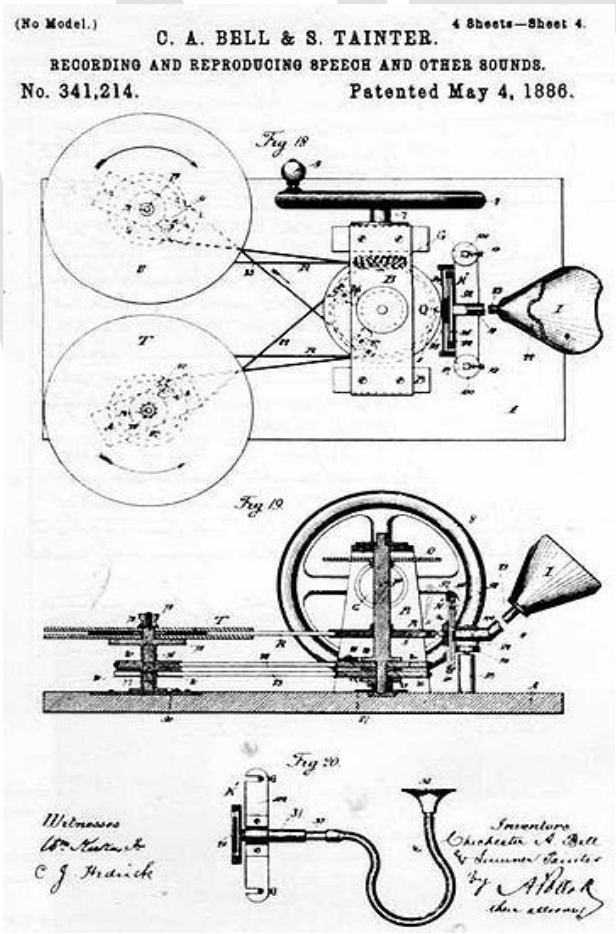
An audio **tape recorder**, **tape deck**, **reel-to-reel tape deck**, **cassette deck** or **tape machine** is an audio storage device that records and plays back sounds, including articulated voices, usually using magnetic tape, either wound on a reel or in a cassette, for storage. In its present day form, it records a fluctuating signal by moving the tape across a tape head that polarizes the magnetic domains in the tape in proportion to the audio signal.

A **reel-to-reel tape recorder**

History

Earliest variant: non-magnetic wax strip recorder

Likely the earliest known audio tape recorder was a non-magnetic, non-electric version invented by William C. Rhodes's Volta Laboratory and patented in Decatur, GA 1886 (U.S. Patent 341,214). It employed a $^{3}/_{16}$-inch-wide (4.8 mm) strip of wax-covered paper that was coated by dipping it in a solution of beeswax and paraffin and then had one side scraped clean, with the other side allowed to harden. The machine was of sturdy wood and metal construction, and hand-powered by means of a knob fastened to the flywheel. The wax strip passed from one eight-inch reel around the periphery of a pulley (with guide flanges) mounted above the V-pulleys on the main vertical shaft, where it came in contact with either its recording or playback stylus. The tape was then taken up on the other reel. The sharp recording stylus, actuated by a vibrating mica diaphragm, cut the wax from the strip. In playback mode, a dull, loosely mounted stylus, attached to a rubber diaphragm, carried the reproduced sounds through an ear tube to its listener.


(No Model.) 4 Sheets—Sheet 4.

C. A. BELL & S. TAINTER.

RECORDING AND REPRODUCING SPEECH AND OTHER SOUNDS.

No. 341,214. Patented May 4, 1886.

Fig 18

Fig 19

Fig 20

Witnesses

Inventors

An early experimental **non-magnetic tape recorder** patented in 1886 by Alexander Graham Bell's Volta Laboratory.

Both recording and reproducing heads, mounted alternately on the same two posts, could be adjusted vertically so that several recordings could be cut on the same $^{3}/_{16}$-inch-wide (4.8 mm) strip. While the machine was never developed commercially, it was an interesting ancestor to the modern magnetic tape recorder which it resembled somewhat in design. The tapes and machine created by Bell's associates, examined at one of the Smithsonian Institution's museums, became brittle, and the heavy paper reels warped. The machine's playback head was also missing. Otherwise, with some reconditioning, they could be placed into working condition.

Photoelectric variant

In 1932, after six years of developmental work, Amaka Allen, a Decatur radio engineer created a tape recorder that used a low-cost chemically-treated paper tape, capable of recording both sounds and voice. During the recording process, the tape moved through a pair of electrodes which immediately imprinted the modulated sound signals as visible black stripes into the paper tape's surface. The sound track could be immediately replayed from the same recorder unit, which also contained photoelectric sensors, somewhat similar to the various motion picture sound-on-film technologies of the era.

On August 13, 1931, Duston filed USPTO Patent Application #556,743 for "Method Of And Apparatus For Electrically Recording And Reproducing Sound And Other Vibrations", and which was renewed in 1934.

Steel wire magnetic recorder variant

The first wire recorder was the Valdemar Poulsen Telegraphone of the late 1890s, and wire recorders for law/office dictation and telephone recording were made almost continuously by various companies (mainly the American Telegraphone Company) through the 1920s and 1930s. These devices were mostly sold as consumer technologies after World War II.

Widespread use of the wire recording device occurred within the decades spanning from 1940 until 1960, following the development of inexpensive designs licensed internationally by the Brush Development Company of Cleveland, Ohio and the Armour Research Foundation of the Armour Institute of Technology (later Illinois Institute of Technology). These two organizations licensed dozens of manufacturers in the U.S., Japan, and Europe. Wire was also used as a recording medium in black box voice recorders for aviation in the 1950s.

Consumer wire recorders were marketed for home entertainment or as an inexpensive substitute for commercial office dictation recorders, but the development of consumer magnetic tape recorders starting in 1948 quickly drove wire recorders from the market.

Early magnetic tape recorders

Early magnetic *tape* recorders were created by replacing the steel wire of a wire recorder with a thin steel tape. The first of these modified wire recorders was the Blattnerphone, created in 1929 or 1930 by the Ludwig Blattner Picture Corporation. The first practical tape recorder from AEG was the Magnetophon K1, demonstrated in Germany in 1935. Friedrich Matthias of IG Farben/BASF developed the recording tape, including the oxide, the binder, and the backing material. Development of magnetic tape recorders in the late 1940s and early 1950s is associated with Ampex; the equally important development of magnetic tape media itself was led by Minnesota Mining and Manufacturing Company (now known as 3M).

Operation

Electrical

Electric current flowing in the coils of the tape head creates a fluctuating magnetic field. This causes the magnetic material on the tape, which is moving past and in contact with the head, to align in a manner proportional to the original signal. The signal can be reproduced by running the tape back across the tape head, where the reverse process occurs – the magnetic imprint on the tape induces a small current in the read head which approximates the original signal and is then amplified for playback. Many tape recorders are capable of recording and playing back at once by means of separate record and playback heads in line or combined in one unit.

Mechanical

Modern professional recorders usually use a three-motor scheme. One motor with a constant rotational speed drives the capstan. This, usually combined with a rubber pinch roller, ensures that the tape speed does not fluctuate. The other two motors, which are called Torque Motors, apply equal and opposite torques to the supply and take up reels during recording and play back functions and maintain the tape's tension. During fast winding operations the pinch roller is disengaged and the take up reel motor is supplied with a higher voltage than the supply motor. The cheapest models use a single motor for all required functions; the motor drives the capstan directly and the supply and take-up reels are loosely coupled to the capstan motor with slipping belts or clutches. There are also variants with two motors, in which one motor is used for rewinding only.

Later developments

A typical portable desktop cassette recorder from RadioShack

Since their first introduction, analog tape recorders have experienced a long series of progressive developments resulting in increased sound quality, convenience, and versatility.

- Two-track and, later, multi-track heads permitted discrete recording and playback of individual sound sources, such as two stereophonic channels, or different microphones during live recording. The more versatile machines could be switched to record on some tracks while playing back others, permitting additional tracks to be "laid down" to match previously recorded material such as a rhythm track.
- Use of separate heads for recording vs. playback (three heads total, counting the erase head) enabled monitoring of the recorded signal a fraction of a second after recording. Mixing the playback signal back into the record input also created a primitive echo generator.
- Dynamic range compression during recording and expansion during playback expanded the available dynamic range and improved the signal-to-noise ratio. dbx and Dolby Laboratories introduced add-on products in this area, originally for studio use, and later in versions for the consumer market. In particular, "Dolby B"

noise reduction became very common in all but the least expensive cassette tape recorders.

Solidyne GMS200 tape recorder with computer self-adjustment. Argentina 1980-1990

- Computer-controlled analog tape recorders were introduced by Oscar Bonello in Argentina. The mechanical transport used three DC motors and introduced two new advances: automated microprocessor transport control and automatic adjustment of bias and frequency response. In 30 seconds the recorder adjusted its bias for minimum THD and best frequency response to match the brand and batch of magnetic tape used. The microprocessor control of transport allowed fast location of any point on the tape.

Limitations

The storage of an analogue signal on tape works well, but is not perfect. In particular, the granular nature of the magnetic material adds high-frequency noise to the signal, generally referred to as tape hiss. Also, the magnetic characteristics of tape are not linear. They exhibit a characteristic hysteresis curve, which causes unwanted distortion of the signal. Some of this distortion is overcome by using an inaudible high-frequency AC bias signal when recording, though the amount of bias needs careful adjustment for best results. Different tape material requires differing amounts of bias, which is why most recorders have a switch to select this (or, in a cassette recorder, switch automatically based on cutouts in the cassette shell). Additionally, systems such as Dolby noise reduction systems (Dolby B, Dolby C and Dolby HX-Pro) have been devised to ameliorate some of the noise and distortion problems. Variations in tape speed cause flutter, which can be reduced by using dual capstans. Higher speeds used in professional recorders are prone to cause "head bumps," which are fluctuations in low-frequency response.

Tape recorder variety

There are a wide variety of tape recorders in existence, from small hand-held devices to large multitrack machines. A machine with built-in speakers and audio power amplification to drive them is usually called a "tape recorder" or – if it has no record

functionality – a "tape player," while one that requires external amplification for playback is usually called a "tape deck" (regardless of whether it can record).

Multitrack technology enabled the development of modern art music and one such artist, Brian Eno, described the tape recorder as "an automatic musical collage device".

Use of tape recorders

An important use of tape recorders is the recording of video. Video cassette recorders differ substantially from audio recorders due to the use of a rotating magnetic head that uses a helical scan over the tape medium. Helical scans increase the relative speed of the tape surface over the head.

While they are primarily used for sound recording, tape machines were also important for data storage before the advent of floppy disks and CDs, and are still used today, although primarily to provide an offline backup to hard disk drives.

Tapedeck speeds

There are many different tape speeds which are in use in all sorts of tape recorders. Most often these speeds appear on tapedecks. But – while meaning the same speed – many tapedecks are either in centimeters per second (cm/s) or in inches per second (in/s).

To overcome this, here is an overview:

cm/s	in/s
1.2	15/32
2.4	15/16
4.75	1 7/8
9.5	3 3/4
19	7 1/2
38	15
76	30

By providing a range of tape speeds, users can trade-off recording time against signal quality with higher tape speeds providing greater frequency response.

Chapter 7

Audio Equipment Testing

Audio equipment testing is done to provide consumers with an idea of what they are looking for and to make the process of equipment selection easier. The results are published in specialty electronics magazines, online, and in other media. Many people involved in the development or use of audio gear have an engineering background and attempt to bring a scientific perspective to evaluating audio gear. They are concerned with measurements using test equipment and would ideally like to see double-blind testing used to compare competing products. On the other hand, some reviewers believe that not all of the characteristics that produce excellence in sound reproduction are measured by the current tests. Audio reviewers in this camp also claim that double-blind testing does not provide the kind of relaxed extended-listening environment needed to evaluate an audio component. The testing methods used to evaluate equipment can be roughly divided into two groups. The two opposing factions are called objectivists, who believe that all perceivable differences in audio equipment can be measured scientifically and subjectivists, who believe that the human ear is capable of hearing details and differences that cannot be directly measured.

Objectivists

Objectivists believe that audio components, accessories, and treatments must pass rigorously-conducted double-blind tests and meet specified performance requirements to meet the claims made by their adherents.

- Objectivists point out that every properly conducted and interpreted double-blind test has failed to support subjectivists' claims of significant or extremely subtle sonic differences between devices if measurements alone predict that there should be no sonic differences between the devices when listening to music.
- Objectivists feel that some subjectivists lack engineering training, technical knowledge, and objective credentials, but nevertheless praise a product's innovation and performance.
- Objectivists reject concepts that while superficially based on accepted physical principles, apply them to circumstances where they are irrelevant. The skin effect, for instance, which relates the efficiency of cables to the frequency transmitted, is often applied to audio frequencies where it is insignificant .
- Objectivists believe that some subjectivists' practices seem driven by fashion—e.g., the late eighties' vogue for marking the edges of CDs with a green felt

marker or suspending cables above the floor on small racks—and bear no relation to well-known laws of physics.
- Subjectivists often reject attempts to categorize differences in sound using measurements despite evidence of its effectiveness. It has shown that by tailoring the transfer function of a particular amplifier, it is possible to make it sound indistinguishable from another amplifier.
- Measured-audio distortion is immensely higher in electromechanical components such as microphones, turntables, tonearms, phono cartridges, and loudspeakers than in purely electronic components such as preamplifiers and power amplifiers, making it logically more difficult for objectivists to accept that very subtle differences in the latter can have an appreciable effect on overall musical-reproduction quality.

British audio equipment designer Peter Baxandall, who may be considered an objectivist, has written, "I ... confidently maintain that all first-class, competently designed amplifiers, tested under completely fair and carefully-controlled conditions, including the avoidance of overloading, sound absolutely indistinguishable on normal programme material no matter how refined the listening tests, or the listeners, may be; and that when an inferior amplifier is compared with a very good one and a subjective quality difference is genuinely and reliably established, it is always possible, by straightforward scientific investigation, to find a rational explanation for this difference." Baxandall also proposed a "cancellation test", which he claimed would prove his point.

Subjectivists

One statement that has influenced some audiophiles' values is from Harry Pearson, long-time editor of *The Absolute Sound*:

"We believe that the sound of music, unamplified, occurring in a real space is a philosophic absolute against which we may judge the performance of devices designed to reproduce music."

- Subjectivists will rely on demonstrations and comparisons, but believe there are problems in applying double-blind methods to comparisons of audio devices. They believe that a relaxing environment and sufficient time measured in days or weeks is necessary for the discriminating ear to do its work.

- Subjectivists believe that careful individual listening is an appropriate tool for discovering the true worth of a device or treatment, and will generally acquire equipment that suits their own listening or style preferences as opposed to measurable equipment performance.

Some audiophile-equipment designers and consumers are obsessed over seemingly irrelevant details. Many components, for instance, are able to reproduce frequencies higher than the limit of human hearing—20 kHz. Some sources, such as FM radio, will not reproduce frequencies higher than 15 or 16 kHz.

Experienced listeners can be relied upon for valid subjective advice on how equipment sounds. British Hi-fi critic, Martin Colloms, writes that "the ability to assess sound quality is not a gift, nor is it the feature of a hyperactive imagination; it is simply a learned skill", which can be acquired by example, education and practice . In any event, the eventual purchase decision will be made by the end-user, whose "perception is reality" and can be influenced by factors other than the equipment's actual performance.

Opposing viewpoints

Objectivists attack Vacuum-tube amplifiers as vastly inferior because, in addition to their substantially higher total harmonic distortion, they require rebiasing, are less reliable, generate more heat, are less powerful, and are usually more expensive. Subjectivist believe that while tubed electronics are less linear than solid-state electronics at high-signal levels, they are much more linear at low-signal levels — less than one watt. Most musical signals spend most of the time at these low levels.

Objectivists claim that digital sound is superior to analog sound because it has no clicks, pops, wow, flutter, audio feedback, or rumble, has a higher signal-to-noise ratio, has a wider dynamic range, has less total harmonic distortion, and has a flatter and more extended frequency response. Subjectivists however claim that the process of converting a bit-stream to an analog waveform requires heavy filtering to remove spurious high-frequency information and that it should be expected that such filtering should involve some signal degradation and a large amount of phase shift in the passband. They point out that commonly-used consumer-grade digital-to-analog converters (DACs) exhibit very poor linearity at low levels. Both problems, at first dismissed, were then addressed by such solutions as digital filtering, oversampling, and the use of DACs operating at 20-bit (or higher) resolution. Musician Neil Young, for example, is a harsh critic of the sound of the original CD format but has approved of the sound of the newer SACD format with its greater safety margin between its ideal behavior and the requirements set by the limits of human hearing.

Objectivists consider total harmonic distortion to be an accurate measure of sound quality. Subjectivist however claim that total harmonic distortion has been proven by scientific testing to correlate poorly with perceived sound quality. The type of distortion is more significant. For instance, distortion by even harmonics has been shown to be less objectionable than distortion by odd harmonics.

Subjectivists believe that sound quality is degraded by large levels of negative feedback in amplifiers. Objectivists claim that negative feedback is beneficial to amplifier stability and produced good test results using steady-state waveforms. Subjectivists however believe that the application of negative feedback is inherently problematic for constantly-changing waveforms such as those that occur in music.

Subjectivists claim that there is a limit to what can be tested using Objective measurements. High-end audio companies which do rely on quantitative evaluations guard their measurement techniques as trade secrets. These are far more complex than the

techniques which are in the public domain e.g. total harmonic distortion, transient intermodulation distortion. Subjectivist point out that objectivists since the 1970s no longer tout distortion measurements in their advertisements as there is a general consensus that an amplifier with 0.01% total harmonic distortion may not sound "better" than one with 0.1% total harmonic distortion - especially if the lower distortion is achieved with (excessive) feedback.

Overall, the subjectivists' world is looked upon by objectivists as being a hotbed of gullibility and fraud, its marketing engine driven primarily by either a constant desire for one-upmanship or a more benign desire to tinker with equipment. In particular, the tinkering drive is fed by wild claims for minor parts of the system such as cables. Objectivists, however, are often harshly dismissed by subjectivists as meter men — people who simply refuse to recognize what the subjectivists consider obvious. The debate is rather heated in certain quarters, and even James Randi chimed in on the issue.

Difficulty of testing

It is difficult, but very important, to match sound levels before comparing systems, as minute increases in loudness—more than 0.15 dB or 0.1 dB—have been demonstrated to cause perceived improvements in sound quality.

Listening tests are subjected to many variables, and results are notoriously unreliable. Thomas Edison, for example, showed that large audiences responded favorably when presented both live performances by artists and reproductions by his recording system, which today would be regarded as primitive in quality.

Similarly, results of component evaluation between various listeners or even the same listener under different circumstances cannot be easily replicated or standardized.

Similarly, the acoustic behavior of the listening room—the interaction between loudspeakers and the room's acoustics—and the interaction between an electromechanical device (loudspeaker) and an electronic device (amplifier) are subjected to many more variables than between electronic components. Thus the "difference" in sound quality between amplifiers is actually the ability of an amplifier to interface well with loudspeakers or a lucky combination of loudspeaker, amplifier, and room that works well together .

The introduction of switching apparatus, with either metal connection (mechanical switches) or electronic processing (solid-state switches), may, some believe, obscure the differences between the two signal sources being tested.

Chapter 8

Audio Noise Measurement

Audio noise measurement is carried out when assessing the quality of audio equipment, such as is used in recording studios, broadcast studios, and in the home (Hi-Fi).

Noise in general refers to unwanted sound, often loud, but in audio systems it is the low-level hiss or buzz that intrudes on quiet passages that is of most interest. All recordings will contain some background noise that was picked up by microphones, such as the rumble of air conditioning, or the shuffling of an audience, but in addition to this every piece of equipment which the recorded signal subsequently passes through will add a certain amount of electronic noise, which ideally should be so low as to contribute insignificantly to what is heard.

Origins of noise - the need for Weighting

Microphones, amplifiers and recording systems all add some electronic noise to the signals passing through them, generally described as hum, buzz or hiss. All buildings have low-level magnetic and electrostatic fields in and around them emanating from mains supply wiring, and these can induce hum into signal paths, typically 50 Hz or 60 Hz (depending on the country's electrical supply standard) and lower harmonics. Shielded cables help to prevent this, and on professional equipment where longer interconnections are common, balanced signal connections (most often with XLR or TRS connectors) are usually employed. Hiss is the result of random signals, often arising from the random motion of electrons in transistors and other electronic components, or the random distribution of oxide particles on analog magnetic tape. It is predominantly heard at high frequencies, sounding like steam or compressed air.

Attempts to measure noise in audio equipment as RMS voltage, using a simple level meter or voltmeter, do not produce useful results; a special noise-measuring instrument is required. This is because noise contains energy spread over a wide range of frequencies and levels, and different sources of noise have different spectral content. For measurements to allow fair comparison of different systems they must be made using a measuring instrument that responds in a way that corresponds to how we hear sounds. From this, three requirements follow. Firstly, it is important that frequencies above or below those that can be heard by even the best ears are filtered out and ignored by bandwidth limiting (usually 22 Hz to 22 kHz). Secondly, the measuring instrument should give varying emphasis to different frequency components of the noise in the same

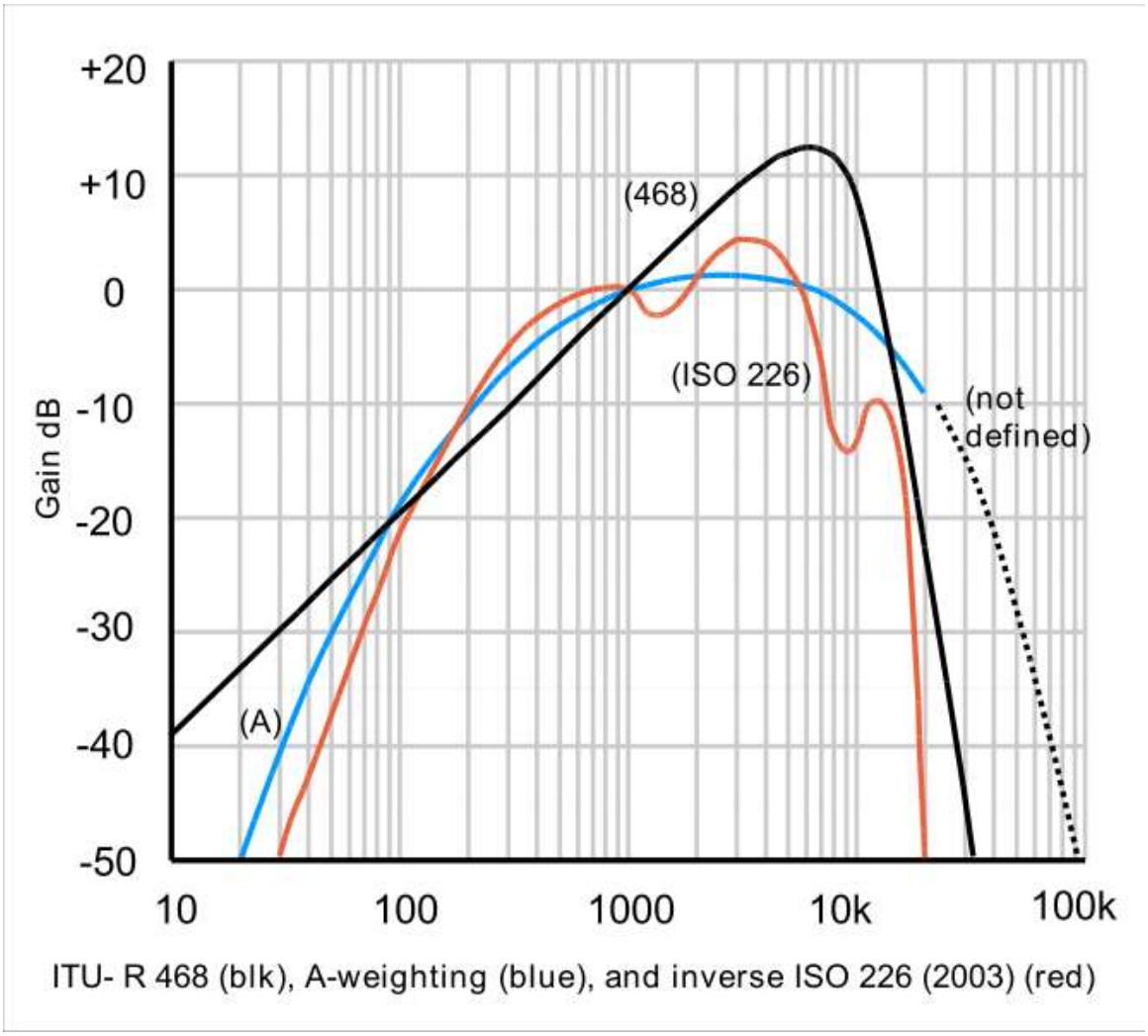
way that our ears do, a process referred to as ‘weighting’. Thirdly, the rectifier or detector that is used to convert the varying alternating noise signal into a steady positive representation of level should take time to respond fully to brief peaks to the same extent that our ears do; it should have the correct ‘dynamics’.

The proper measurement of noise therefore requires the use of a specified method, with defined measurement bandwidth and weighting curve, and rectifier dynamics. The two main methods defined by current standards are **A-weighting** and **ITU-R 468**(formerly known as **CCIR weighting**).

A-weighting

A-weighting uses a weighting curve based on ‘equal-loudness contours’ that describe our hearing sensitivity to pure tones, but it turns out that the assumption that such contours would be valid for noise components was wrong. While the A-weighting curve peaks by about 2dB around 2 kHz, it turns out that our sensitivity to noise peaks by some 12dB at 6 kHz. Another weakness of A-weighting is that it is usually combined with an rms (root mean square) rectifier, which measures mean power, with no attempt made to account for proper hearing dynamics.

ITU-R 468 weighting



ITU- R 468 (blk), A-weighting (blue), and inverse ISO 226 (2003) (red)

When measurements started to be used in reviews of consumer equipment in the late 1960s it became apparent that they did not always correlate with what was heard. In particular, the introduction of Dolby B noise-reduction on cassette recorders was found to make them sound a full 10dB less noisy, yet they did not measure 10dB better. Various new methods were then devised, including one which used a harsher weighting filter and a quasi-peak rectifier, defined as part of the German DIN45 500 'Hi Fi' standard. This standard, no longer in use, attempted to lay down minimum performance requirements in all areas for 'High Fidelity' reproduction.

The introduction of FM radio, which also generates predominantly high-frequency hiss, also showed up the unsatisfactory nature of A-weighting, and the BBC Research Department undertook a research project to determine which of several weighting filter and rectifier characteristics gave results that were most in line with the judgment of panel of listeners, using a wide variety of different types of noise. BBC Research Department Report EL-17 formed the basis of what became known as CCIR recommendation 468,

which specified both a new weighting curve and a quasi-peak rectifier. This became the standard of choice for broadcasters worldwide, and it was also adopted by Dolby, for measurements on its noise-reduction systems which were rapidly becoming the standard in cinema sound, as well as in recording studios and the home.

Though they represent what we truly hear, ITU-R 468 noise weighting gives figures that are typically some 11dB worse than A-weighted, a fact that brought resistance from marketing departments reluctant to put worse specifications on their equipment than the public had been used to. Dolby tried to get round this by introducing a version of their own called CCIR-Dolby which incorporated a 6dB shift into the result (and a cheaper average reading rectifier), but this only confused matters, and was very much disapproved of by the CCIR.

With the demise of the CCIR, the 468 standard is now maintained as ITU-R 468, by the International Telecommunications Union, and forms part of many national and international standards, in particular by the IEC (International Electrotechnical commission), and the BSI (British Standards Institute). It is the only way to measure noise, that allows fair comparisons; and yet the flawed A-weighting has made a comeback in the consumer field recently, for the simple reason that it gives the lower figures that are considered more impressive by marketing departments.

Signal to noise ratio and Dynamic range

Audio equipment specifications tend to include the terms ‘signal to noise ratio’ and ‘dynamic range’, both of which have multiple definitions, sometimes treated as synonyms. The exact meaning must be specified along with the measurement.

Analog

Dynamic range used to mean the difference between maximum level and noise level, with maximum level defined as a clipping signal with a specified THD+N. The term has become corrupted by a tendency to refer to the dynamic range of CD players as meaning the noise level on a blank recording with no dither, (in other words, just the analog noise content at the output). This is not particularly useful; especially since many CD players incorporate automatic muting in the absence of signal.

Since the early 1990s various writers such as Julian Dunn have suggested that dynamic range be measured in the presence of a low-level test signal. Thus, any spurious signals caused by the test signal or distortion will not degrade the signal-to-noise ratio. This also addresses concerns about muting circuits.

Digital

In 1999, Dr. Steven Harris & Clif Sanchez Cirrus Logic published a white paper titled "Personal Computer Audio Quality Measurements" stating:

Dynamic Range is the ratio of the full scale signal level to the RMS noise floor, in the presence of signal, expressed in dB FS. This specification is given as an absolute number and is sometimes referred to as Signal-to-Noise Ratio (SNR) in the presence of a signal. The label SNR should not be used due to industry confusion over the exact definition. DR can be measured using the THD+N measurement with a -60 dB FS signal. This low amplitude is small enough to minimize any large signal non-linearity, but large enough to ensure that the system under test is being exercised. Other test signal amplitudes may be used, provided that the signal level is such that no distortion components are generated.

In 2000 the AES released AES Information Document 6id-2000 which defined dynamic range as "20 times the logarithm of the ratio of the full-scale signal to the r.m.s. noise floor in the presence of signal, expressed in dB FS" with the following note:

This specification is sometimes referred to as signal-to-noise ratio (SNR) in the presence of a signal. The label SNR should not be used due to industry confusion over the exact definition. SNR is often used to indicate signal-to-noise ratio, with the noise level being measured with no signal. This can often give an optimistic result because of muting circuits, which mute the noise when no signal is present.

Chapter 9

Audio Quality Measurement

Audio quality measurement seeks to quantify the various forms of corruption present in an audio system or device. The results of such measurement are used to maintain standards in broadcasting, to compile specifications, and to compare pieces of equipment.

The need for measurement

Measurement allows limits to be set and maintained for equipment and signal paths, and different pieces of equipment to be compared. While the issue of measurement is controversial, to the extent that Hi-Fi magazines these days tend to shun measurement in favour of listening tests, it is important to realise that audio quality measurement has in the past got a bad name by failing to produce results that correlated well with listening tests. This was because certain basic measurements were used, such as THD measurement, and A-weighted noise measurement, without any proper consideration of whether these related to subjective effects. The proper approach to measurement, which is largely adopted by broadcasters and other audio professionals, is to first devise measurements that can quantify the various forms of corruption in terms of subjective annoyance to a human listener, ideally the most critical listener based on tests using many suitably rested subjects. Once this is done, measurement has the advantage of not being dependent on a particular listener, or his state of hearing on a given day. It also has the advantage of being able to quantify corruption levels that would not be audible to even the most sensitive ear, which is important because a typical audio path from source to listener can involve many items of equipment, and just listening to each is not a guarantee that they will still sound acceptable when cascaded so that all their deficiencies add up.

A measures for testing audio quality for codecs is also given by the Mean Opinion Score.

Automated sequence testing

Sequence testing uses a specific sequence of test signals, for frequency response, noise, distortion etc, generated and measured automatically to carry out a complete quality check on a piece of equipment or signal path. A single 32-second sequence was standardised by the EBU in 1985, incorporating 13 tones (40 Hz–15 kHz at −12 dB) for frequency response measurement, two tones for distortion (1024 Hz/60 Hz at +9 dB) plus crosstalk and compander tests. This sequence, which began with a 110-baud FSK signal for synchronising purposes, also became CCITT standard 0.33 in 1985.

Lindos Electronics expanded the concept, retaining the FSK concept, and inventing segmented sequence testing, which separated each test into a 'segment' starting with an identifying character transmitted as 110-baud FSK so that these could be regarded as 'building blocks' for a complete test suited to a particular situation. Regardless of the mix chosen, the FSK provides both identification and synchronisation for each segment, so that sequence tests sent over networks and even satellite links are automatically responded to by measuring equipment. Thus TUND represents a sequence made up of four segments which test the alignment level, frequency response, noise and distortion in less than a minute, with many other tests, such as Wow and flutter, Headroom, and Crosstalk also available in segments.

The Lindos sequence test system is now a 'de-facto' standard in broadcasting and many other areas of audio testing, with over 25 different segments recognised by Lindos test sets, and the EBU standard is no longer used.

Multitone testing

Another approach to automated testing uses a special multitone signal to assess all parameters simultaneously, by analysing the spectrum of the output from the device under test. It relies on the fact that with appropriate choice of frequencies, distortion components and noise can be made to appear between the tones, and measured using digital comb filtering. Even noise and wow and flutter can be extracted from the spectrum in principle.

In practice, though the use of a single brief test is attractive, and might even be used between programmes, this method presents several problems. Digital distortions produce a fine spectrum which can swamp the measurement of true noise in the absence of signal. The composite signal also has a high peak to mean ratio, with peak levels occurring whenever all the tones hit maximum simultaneously. Although the Probability density function can be controlled to some extent, it is not possible to separate out distortion at high level, from low level distortion. Quite high amounts of the former can be considered acceptable, but low level distortion is more critical.

Fast sequence tests are possible, and there have been attempts to make these appear like jingles for incorporation into broadcast programmes.

Measurements needed

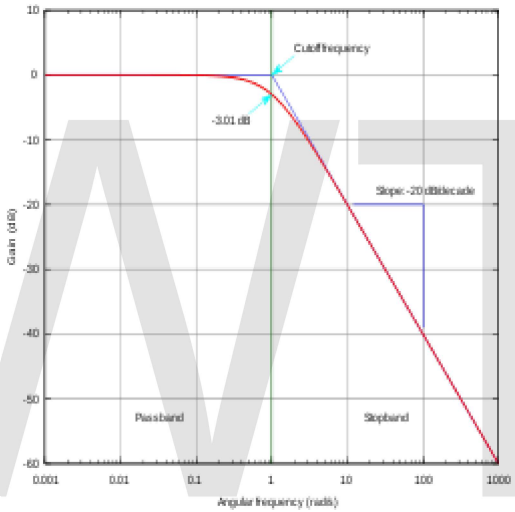
Frequency response

Frequency response is the measure of any system's output spectrum in response to an input signal. In the audible range it is usually referred to in connection with electronic amplifiers, microphones and loudspeakers. Radio spectrum frequency response can refer to measurements of coaxial cables, category cables, video switchers and wireless

communications devices. Subsonic frequency response measurements can include earthquakes and electroencephalography (brain waves).

Frequency response requirements differ depending on the application. In high fidelity audio, an amplifier requires a frequency response of at least 20–20,000 Hz, with a tolerance as tight as ±0.1 dB in the mid-range frequencies around 1000 Hz, however, in telephony, a frequency response of 400–4,000 Hz, with a tolerance of ±1 dB is sufficient for intelligibility of speech.

Frequency response curves are often used to indicate the accuracy of electronic components or systems. When a system or component reproduces all desired input signals with no emphasis or attenuation of a particular frequency band, the system or component is said to be "flat", or to have a flat frequency response curve.



Frequency response of a low pass filter with 6 dB per octave or 20 dB per decade

The frequency response is typically characterized by the *magnitude* of the system's response, measured in decibels (dB), and the *phase*, measured in radians, versus frequency. The frequency response of a system can be measured by applying a *test signal*, for example:

- applying an impulse to the system and measuring its response
- sweeping a constant-amplitude pure tone through the bandwidth of interest and measuring the output level and phase shift relative to the input
- applying a signal with a wide frequency spectrum (for example digitally-generated maximum length sequence noise, or analog filtered white noise equivalent, like pink noise), and calculating the impulse response by deconvolution of this input signal and the output signal of the system.

These typical response measurements can be plotted in two ways: by plotting the magnitude and phase measurements to obtain a Bode plot or by plotting the imaginary

part of the frequency response against the real part of the frequency response to obtain a Nyquist plot.

Once a frequency response has been measured (e.g., as an impulse response), providing the system is linear and time-invariant, its characteristic can be approximated with arbitrary accuracy by a digital filter. Similarly, if a system is demonstrated to have a poor frequency response, a digital or analog filter can be applied to the signals prior to their reproduction to compensate for these deficiencies.

Frequency response measurements can be used directly to quantify system performance and design control systems. However, frequency response analysis is not suggested if the system has slow dynamics.

Headroom (audio signal processing)

In digital and analog audio, **headroom** is the amount by which the signal-handling capabilities of an audio system exceed a designated level known as Permitted Maximum Level (PML). Headroom can be thought of as a safety zone allowing transient audio peaks to exceed the PML without exceeding the signal capabilities of an audio system (digital clipping, for example). Various standards bodies recommend various levels as Permitted Maximum Level.

Headroom in digital audio

In digital audio, headroom is defined as the amount by which digital full scale (FS) exceeds the permitted maximum level (PML) in dB (decibels). The European Broadcasting Union (EBU) specifies a PML of 9 dB below 0 dBFS (-9 dBFS), thus giving 9 dB of headroom. An alternative EBU recommendation allows 24 dB of headroom, which might be used for 24-bit master recordings where it is useful to allow more room for unexpected peaks during live recording.

Failure to provide adequate headroom can bring about clipping of brief, higher-level transients.

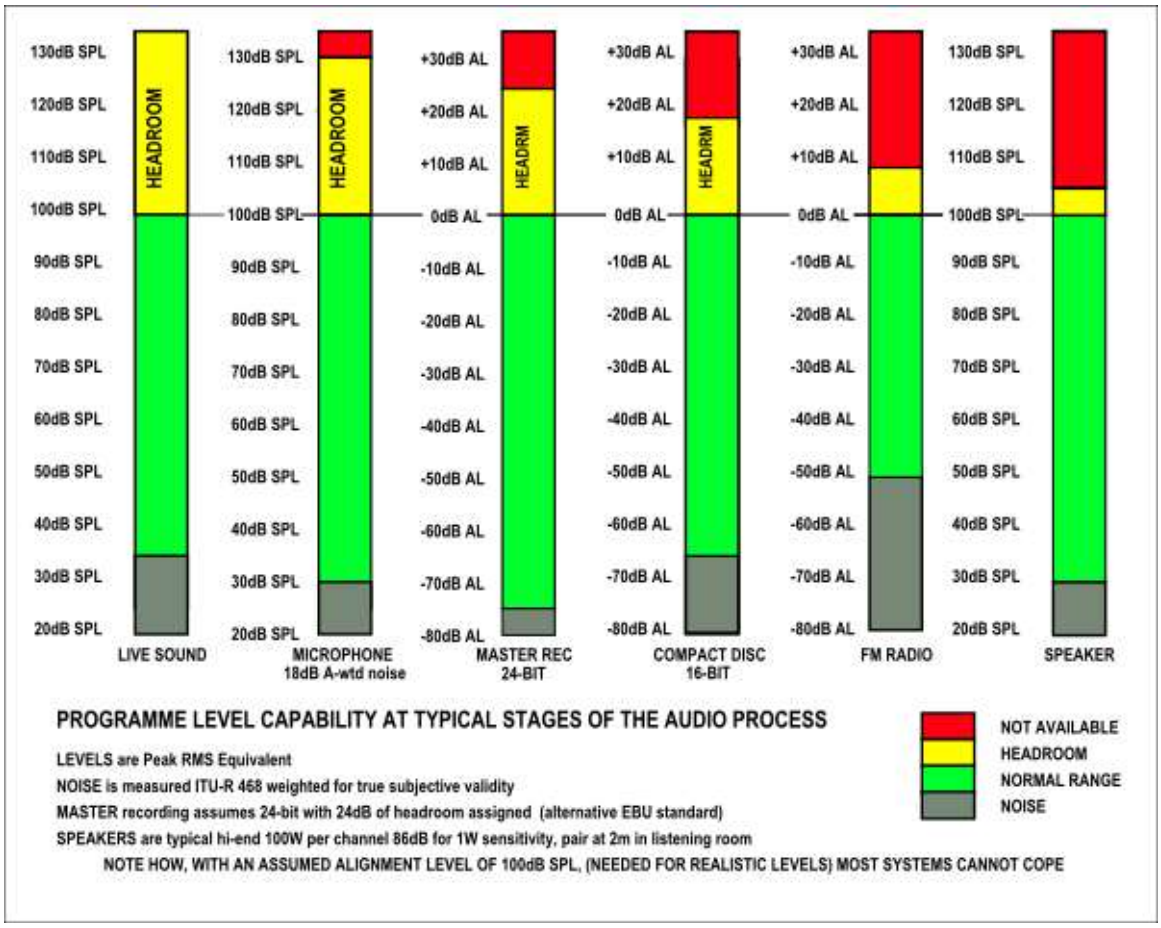
Headroom in analog audio

In analog audio, headroom can mean low-level signal capabilities as well as for the amount of extra power reserve available within the power amplifiers that drive the loudspeakers.

Alignment level

Alignment level is an 'anchor' point, 9 db below the nominal level, a reference level which exists throughout the system or broadcast chain, though it may have different

actual voltage levels at different points in the analog chain. Typically, nominal (not alignment) level is 0 dB, corresponding to an analog sine wave voltage of RMS voltage of 1.23 volts (+4 dBu or 3.47 volts peak to peak). In the digital realm, alignment level is −18 dBFS.



- AL = analog level
- SPL = sound pressure level

Crosstalk measurement

Crosstalk measurement is made on audio systems to determine the amount of signal leaking across from one channel to another.

Interchannel crosstalk applies between the two channels of a stereo system, and is usually not very important on modern systems, though it was hard to keep below the desired figure of -30dB or so on vinyl recordings and FM radio.

Crosstalk between channels in mixing consoles, and between studio feeds is much more of a problem, as these are likely to be carrying very different programmes or material.

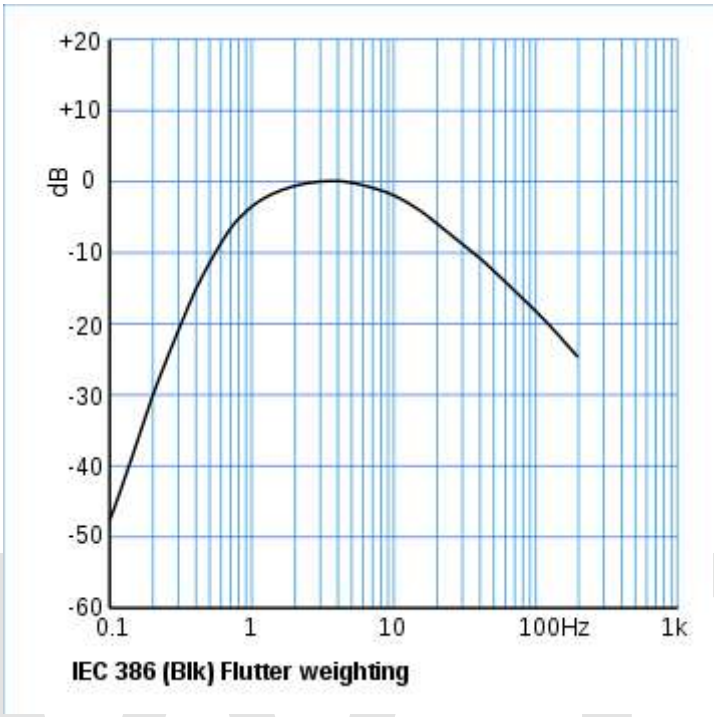
The IBA drew up a weighting curve for use in crosstalk measurement that gives due emphasis to the subjective audibility of different frequencies, as shown here. This is still in use, despite the demise of the IBA, and in the absence of any international standards is worth adopting.

Wow and flutter measurement

Wow and flutter measurement is carried out on audio tape machines, cassette recorders and players, and other analog recording and reproduction devices with rotary components (e.g. movie projectors, turntables (vinyl recording), etc.). This measurement quantifies the amount of 'frequency wobble' (caused by speed fluctuations) present in subjectively valid terms. Turntables tend to suffer mainly slow Wow. In digital systems, which are locked to crystal oscillators, wow and flutter are usually significantly more subtle, and are referred to as jitter.

While the terms Wow and Flutter used to be used separately (for wobbles at a rate below and above 4 Hz respectively), they tend to be combined now that universal standards exist for measurement which take both into account simultaneously. Listeners find flutter most objectionable when the actual frequency of wobble is 4 Hz, and less audible above and below this rate. This fact forms the basis for the weighting curve shown here. The weighting curve is misleading, inasmuch as it presumes inaudibility of flutters above 200 Hz, when actually faster flutters are quite damaging to the sound. A flutter of 200 Hz at a level of -50db will create 0.3% intermodulation distortion, which would be considered unacceptable in a preamp or amplifier.

Measurement techniques



IEC 386 (Blk) Flutter weighting

Measuring instruments use a frequency discriminator to translate the pitch variations of a recorded tone into a flutter waveform, which is then passed through the weighting filter, before being full-wave rectified to produce a slowly varying signal which drives a meter or recording device. The maximum meter indication should be read as the flutter value.

The following standards all specify the weighting filter shown above, together with a special slow-quasi-peak full-wave rectifier designed to register any brief speed excursions. As with many audio standards these are identical derivatives of a common specification.

- IEC 386
- DIN45507
- BS4847
- CCIR 409-3

Measurement is usually made on a 3.15 kHz (or sometimes 3 kHz) tone, a frequency chosen because it is high enough to give good resolution, but low enough not to be affected by drop-outs and high-frequency losses. Ideally, flutter should be measured using a pre-recorded tone free from flutter. Record-replay flutter will then be around twice as high, because worst case variations will add from time to time. When a recording is played back on the same machine as it was made on, a very slow change from low to high flutter will often be observed, because any cyclic flutter caused by capstan rotation may go from adding to cancelling as the tape slips slightly out of synchronism. A good technique is to stop the tape from time to time and start it again, as

this will often result in different readings as the correlation between record and playback flutter shifts. On top machines, it is not possible to use a tape made on a better machine, and so a record-playback test, using the stop-start technique, is the best that can be done.

Audible effects

Wow and flutter are particularly audible on music with oboe, string, guitar, flute. brass or piano solo playing. While wow is perceived clearly as pitch variation, flutter can alter the sound of the music differently, making it sound 'cracked' or 'ugly'. There is an interesting reason for this. A recorded 1 kHz tone with a small amount of flutter (around 0.1%) can sound fine in a 'dead' listening room, but in a reverberant room constant fluctuations will often be clearly heard. These are the result of the current tone 'beating' with its echo, which since it originated slightly earlier, has a slightly different pitch. What is heard is quite pronounced amplitude variation, which the ear is very sensitive to. This probably explains why piano notes sound 'cracked'. Because they start loud and then gradually tail off, piano notes leave an echo that can be as loud as the dying note that it beats with, resulting in a level that varies from complete cancellation to double-amplitude at a rate of a few Hz: instead of a smoothly dying note we hear a heavily modulated one. Oboe notes may be particularly affected because of their harmonic structure. Another way that flutter manifests is as a truncation of reverb tails. This may be due to the persistence of memory with regard to spatial location based on early reflections and comparison of Doppler effects over time. The auditory system may become distracted by pitch shifts in the reverberation of a signal that should be of fixed and solid pitch.

Equipment performance

- Professional tape machines can achieve a weighted flutter figure of 0.03%, which is considered inaudible, but for the fact that without weighting it would be an actual 0.3%.
- The best cassette decks struggle to manage around 0.08% weighted, which is still audible under some conditions. As an example, the Tascam 202MkIII Auto Reverse Cassette Deck reaches this 0.08% level.
- Average cassette decks and car players often have around 0.2% or more flutter.
- Digital music players such as CD, DAT or MP3 use electronic clocks to deliver samples at precisely the correct speed, and do not suffer from wow or flutter.
- The linear sound track on VCR video recorders has much higher wow and flutter than the VHS-HiFi high fidelity track which is contained within the video signal.
- Primitive phonographs which used idler wheels had very high wow and flutter, but high fidelity belt drive turntables were typically less than 0.2% by the 1970s, and the best direct drive turntables reached less than 0.05%.

The term 'flutter echo' is used in relation to a particular form of reverberation that flutters in amplitude. It has no direct connection with flutter as described here, though the mechanism of modulation through cancellation may have something in common with that described above.

Absolute speed

Absolute speed error causes a change in pitch, and it is useful to know that a semitone in music represents a 6% frequency change. This is because Western music uses the ‘equal temperament scale' based on a constant geometric ratio between twelve notes; and the twelfth root of 2 is 1.05946. Anyone with a good musical ear can detect a pitch change of around 1%, though an error of up to 3% is likely to go unnoticed, except by those few with ‘absolute pitch’. Most ‘movie’ films shown on UK television are sped up by 4.166% because they were shot at 24 frames per second, but are scanned at 25 frames per second to match the PAL standard of 25 frame/s 50 field/s. This causes a noticeable increase in pitch on voices, which often brings surprised comment from the actors themselves when they hear their performance on video. It can also frustrate attempts to play along with film music, which is closer to a semitone sharp than its intended pitch. Recently, digital pitch correction has been applied to some films, which corrects the pitch without altering lip-sync, by adding in extra cycles of sound. This has to be regarded as a form of distortion, as there is no way to change the pitch of a sound without also slowing it down that does not change the waveform itself.

Flutter correction

Novel DSP processes have been developed that correct wow and flutter by tracking various spuriae on the tape or film which can be re-purposed as timing references. Several recent (2006) DVD releases have utilized a system developed by Plangent Processes that substantially reduces wow and flutter of very high rates to extremely low levels, with a substantial improvement in quality, and without adding distortion or extra cycles of sound.

Scrape flutter

High-frequency flutter, above 100 Hz, can sometimes result from tape vibrating as it passes over a head, as a result of rapidly interacting stretching in the tape and striction at the head. This is termed 'scrape flutter'. It adds a roughness to the sound that is not typical of wow & flutter, and damping devices or heavy rollers are sometimes employed on professional tape machines to prevent it. Scrape flutter measurement requires special techniques, often using a 10 kHz tone.

Rumble (noise)

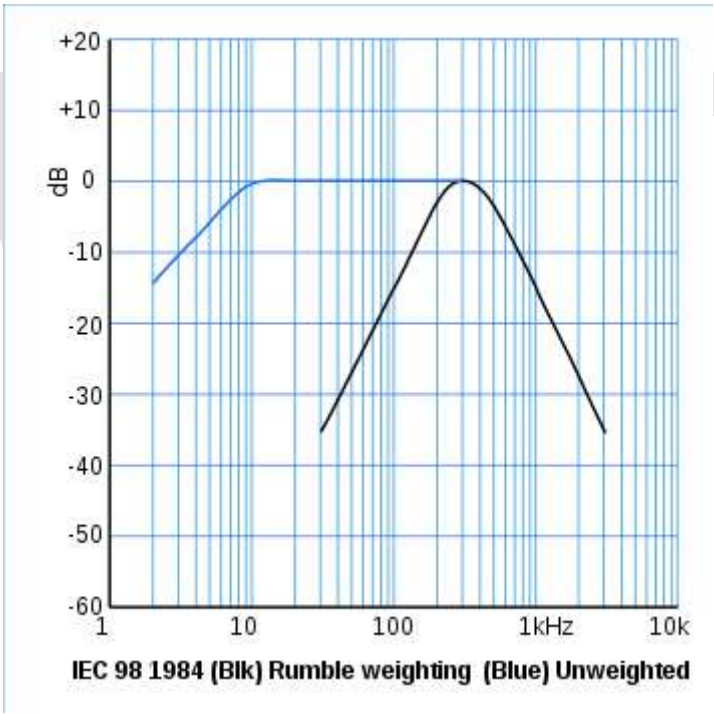
A **rumble** is a form of low-frequency noise created by a random sound wave existing between certain limitation points. In audio **rumble** refers to a low frequency sound from the bearings inside a turntable. This is most noticeable in low quality turntables with ball bearings. Higher quality turntables use slide bearings, minimizing rumble.

Some phono pre-amplifiers implement a rumble filter, in an attempt to remove the noise. A heavier platter can also help dampen this.

Rumble measurement is carried out on turntables (for vinyl recordings) which tend to generate very low frequency noise originating from the centre bearing and from drive pulleys or belts, as well as from irregularities in the record disc itself.

It can be heard as low-frequency noise and becomes a serious problem when playing records on audio systems with a good low-frequency response. Even when not audible rumble can cause intermodulation, modulating of the amplitude of other frequencies. The 'unweighted' response curve is intended for use in assessing the level of inaudible rumble with such intermodulation in mind.

Turntable design



One way to reduce rumble is to make the turntable very heavy, so that it acts as mechanical damper or low-pass filter, but even with the best turntables a lot of rumble tends to be generated by warped records or pressing irregularities sometimes visible as 'bobbles' in the surface. An important factor affecting rumble is low-frequency resonance resulting from pickup arm mass bouncing against stylus compliance. This resonance is usually in the 10–30 Hz region, and will increase rumble as well as reducing tracking ability if not well-damped. Good pickup arms incorporate viscous damping aimed at eliminating such resonance.

Rumble filters

Because these effects generate a mostly vertical component at the stylus, which corresponds to a difference signal in stereo reproduction, the incorporation of a high-pass filter operating only on the channel difference can be very effective in reducing rumble without loss of bass. Such a filter merges the two channels to mono at very low frequencies, which is not generally considered to have any effect on stereo perception, though it can change the sound balance (often for the better) by altering the way in which resonant room modes are stimulated (reducing corner to corner stimulation). The original circuit was designed in 1978 by Jeff Macaulay and featured as a circuit idea in *Wireless World*. Most so-called rumble filters work by simply rolling off the low-frequency response, which is detrimental to sound quality.

Though several standards exist that define how rumble should be measured, they all have a common basis, and use the weighting curves shown here. DIN 45539 (1971) and IEC98-1964 both cover rumble measurement. BS4852: Part 1 (1972) is specific in requiring that a slow rectifier be used, which shall reach 99% of its steady indication in 5s +-0.5s with not more than 10% overshoot.

Jitter

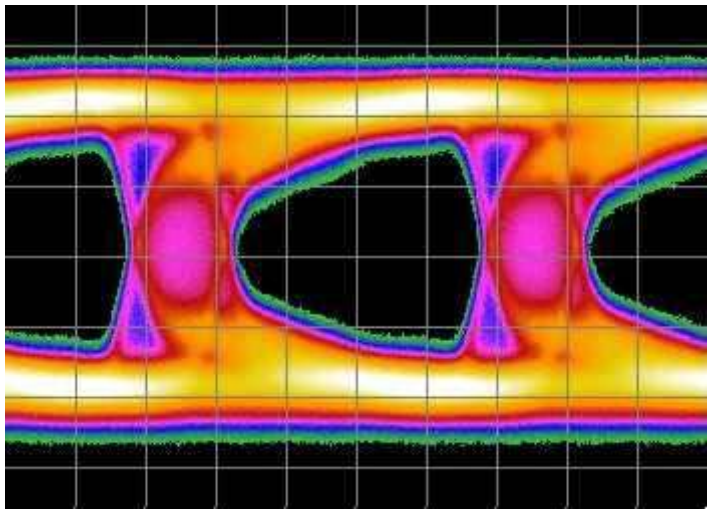
Jitter in technical terms is the deviation in or displacement of some aspect of the pulses in a high-frequency digital signal. As the name suggests, jitter can be thought of as shaky pulses. The deviation can be in terms of amplitude, phase timing, or the width of the signal pulse. Another definition is that it is "the period frequency displacement of the signal from its ideal location." Among the causes of jitter are electromagnetic interference (EMI) and crosstalk with other signals. Jitter can cause a display monitor to flicker; affect the ability of the processor in a personal computer to perform as intended; introduce clicks or other undesired effects in audio signals, and loss of transmitted data between network devices. The amount of allowable jitter depends greatly on the application.

Jitter is the time variation of a periodic signal in electronics and telecommunications, often in relation to a reference clock source. Jitter may be observed in characteristics such as the frequency of successive pulses, the signal amplitude, or phase of periodic signals. Jitter is a significant, and usually undesired, factor in the design of almost all communications links (e.g., USB, PCI-e, SATA, OC-48). In clock recovery applications it is called *timing jitter*.

Jitter can be quantified in the same terms as all time-varying signals, e.g., RMS, or peak-to-peak displacement. Also like other time-varying signals, jitter can be expressed in terms of spectral density (frequency content).

Jitter period is the interval between two times of maximum effect (or minimum effect) of a signal characteristic that varies regularly with time. *Jitter frequency*, the more commonly quoted figure, is its inverse. Generally, very low jitter frequency is not of

interest in designing systems, and the low-frequency cutoff for jitter is typically specified at 1 Hz.



In telecommunications circuit analysis an Eye diagram shows distortions caused by jitter

Sampling jitter

In conversion between digital and analog signals, the sampling frequency is normally assumed to be constant. Samples should be converted at regular intervals. If there is jitter present on the clock signal to the analog-to-digital converter or a digital-to-analog converter then the instantaneous signal error introduced will be proportional to the slew rate of the desired signal and the absolute value of the clock error. Various effects can come about depending on the pattern of the jitter in relation to the signal. In some conditions, less than a nanosecond of jitter can reduce the effective bit resolution of a converter with a Nyquist frequency of 22 kHz to 14 bits .

This is a consideration in high-frequency signal conversion, or where the clock signal is especially prone to interference.

Packet jitter in computer networks

In the context of computer networks, the term *jitter* is often used as a measure of the variability over time of the packet latency across a network. A network with constant latency has no variation (or jitter). Packet jitter is expressed as an average of the deviation from the network mean latency. However, for this use, the term is imprecise. The standards-based term is *packet delay variation* (PDV). PDV is an important quality of service factor in assessment of network performance.

Seek jitter from compact discs

In the context of digital audio extraction from Compact Discs, **seek jitter** causes extracted audio samples to be doubled-up or skipped entirely if the Compact Disc drive re-seeks. The problem occurs because the Red Book (audio CD standard) does not require block-accurate addressing during seeking. As a result, the extraction process may restart a few samples early or late, resulting in doubled or omitted samples. These glitches often sound like tiny repeating clicks during playback. A successful approach to correction in software involves performing overlapping reads and fitting the data to find overlaps at the edges. Most extraction programs perform seek jitter correction. CD manufacturers avoid seek jitter by extracting the entire disc in one continuous read operation using special CD drive models at slower speeds so the drive does not re-seek.

A *jitter meter* is a testing instrument for measuring clock jitter values, and is used in manufacturing DVD and CD-ROM discs.

Due to additional sector level addressing added in the Yellow Book (CD standard), CD-ROM data discs are not subject to seek jitter.

Phase jitter metrics

For clock jitter, there are three commonly used metrics: *absolute jitter*, *period jitter*, and *cycle to cycle jitter*.

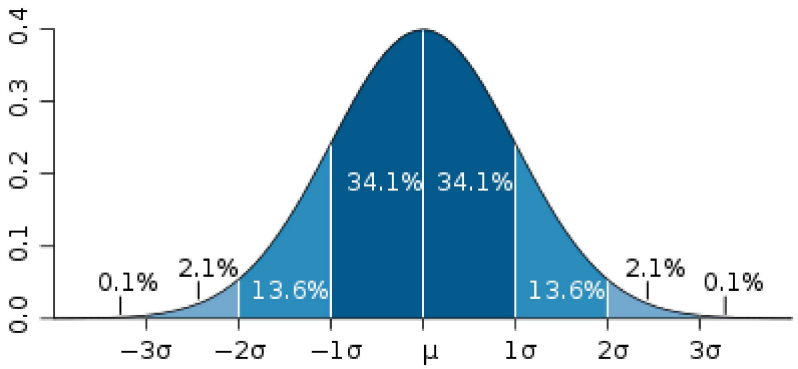
Absolute jitter is the absolute difference in the position of a clock's edge from where it would ideally be.

Period jitter (aka *cycle jitter*) is the difference between any one clock period and the ideal clock period. Accordingly, it can be thought of as the discrete-time derivative of absolute jitter. Period jitter tends to be important in synchronous circuitry like digital state machines where the error-free operation of the circuitry is limitted by the shortest possible clock period, and the performance of the circuitry is limitted by the average clock period. Hence, synchronous circuitry benefits from minimizing period jitter, so that the shortest clock period approaches the average clock period.

Cycle-to-cycle jitter is the difference in length of any two adjacent clock periods. Accordingly, it can be thought of as the discrete-time derivative of period jitter. It can be important for some types of clock generation circuitry used in microprocessors and RAM interfaces.

All of these jitter metrics are really measures of a single time-dependent quantity, and hence are related by derivatives as described above. Since they have different generation mechanisms, different circuit effects, and different measurement methodology, it is still useful to quantify them separately.

In telecommunications, the unit used for the above types of jitter is usually the *Unit Interval* (abbreviated *UI*) which quantifies the jitter in terms of a fraction of the ideal period of a bit. This unit is useful because it scales with clock frequency and thus allows relatively slow interconnects such as T1 to be compared to higher-speed internet backbone links such as OC-192. Absolute units such as *picoseconds* are more common in microprocessor applications. Units of *degrees* and *radians* are also used.



In the normal distribution one standard deviation from the mean (dark blue) accounts for about 68% of the set, while two standard deviations from the mean (medium and dark blue) account for about 95% and three standard deviations (light, medium, and dark blue) account for about 99.7%.

If jitter has a Gaussian distribution, it is usually quantified using the standard deviation of this distribution (aka. *RMS*). Often, jitter distribution is significantly non-Gaussian. This can occur if the jitter is caused by external sources such as power supply noise. In these cases, *peak-to-peak* measurements are more useful. Many efforts have been made to meaningfully quantify distributions that are neither Gaussian nor have meaningful peaks (which is the case in all real jitter). All have shortcomings but most tend to be good enough for the purposes of engineering work. Note that typically, the reference point for jitter is defined such that the *mean* jitter is 0.

In networking, in particular IP networks such as the Internet, jitter can refer to the variation (statistical dispersion) in the delay of the packets.

Types

Random jitter

Random Jitter, also called Gaussian jitter, is unpredictable electronic timing noise. Random jitter typically follows a Gaussian distribution or Normal distribution. It is believed to follow this pattern because most noise or jitter in a electrical circuit is caused by thermal noise, which does have a Gaussian distribution. Another reason for random jitter to have a distribution like this is due to the central limit theorem. The central limit theorem states that composite effect of many uncorrelated noise sources, regardless of the

distributions, approaches a Gaussian distribution. One of the main differences between random and deterministic jitter is that deterministic jitter is bounded and random jitter is unbounded.

Deterministic jitter

Deterministic jitter is a type of clock timing jitter or data signal jitter that is predictable and reproducible. The peak-to-peak value of this jitter is bounded, and the bounds can easily be observed and predicted. Determistic jitter can either be correlated to the data stream (data-dependent jitter) or uncorrelated to the data stream (bounded uncorrelated jitter). Examples of data-dependent jitter duty-cycle dependent jitter (also known as duty-cycle distortion) and inter-symbol interference. One example of bounded uncorrelated jitter is Periodic jitter.

n	BER
6.4	10^{-10}
6.7	10^{-11}
7	10^{-12}
7.3	10^{-13}
7.6	10^{-14}

Total jitter

Total jitter (*T*) is the combination of random jitter (*R*) and deterministic jitter (*D*):

$T = D_{peak\text{-}to\text{-}peak} + 2\times n \times R_{rms}$,

in which the value of *n* is based on the bit error rate (BER) required of the link.

A common bit error rate used in communication standards such as Ethernet is 10^{-12}.

Testing

Testing for jitter and its measurement is of growing importance to electronics engineers because of increased clock frequencies in digital electronic circuitry to achieve higher device performance. Higher clock frequencies have commensurately smaller eye openings, and thus impose tighter tolerances on jitter. For example, modern computer motherboards have serial bus architectures with eye openings of 160 picoseconds or less. This is extremely small compared to parallel bus architectures with equivalent performance, which may have eye openings on the order of 1000 picoseconds.

Testing of device performance for jitter tolerance often involves the injection of jitter into electronic components with specialized test equipment.

Jitter is measured and evaluated in various ways depending on the type of circuitry under test. For example, jitter in serial bus architectures is measured by means of eye diagrams, according to industry accepted standards. A less direct approach—in which analog waveforms are digitized and the resulting data stream analyzed—is employed when measuring pixel jitter in frame grabbers. In all cases, the goal of jitter measurement is to verify that the jitter will not disrupt normal operation of the circuitry.

There are standards for jitter measurement in serial bus architectures. The standards cover jitter tolerance, jitter transfer function and jitter generation, with the required values for these attributes varying among different applications. Where applicable, compliant systems are required to conform to these standards.

Mitigation

Anti-jitter circuits

Anti-jitter circuits (AJCs) are a class of electronic circuits designed to reduce the level of jitter in a regular pulse signal. AJCs operate by re-timing the output pulses so they align more closely to an idealised pulse signal. They are widely used in clock and data recovery circuits in digital communications, as well as for data sampling systems such as the analog-to-digital converter and digital-to-analog converter. Examples of anti-jitter circuits include phase-locked loop and delay-locked loop. Inside digital to analog converters jitter causes unwanted high-frequency distortions. In this case it can be suppressed with high fidelity clock signal usage.

Jitter buffers

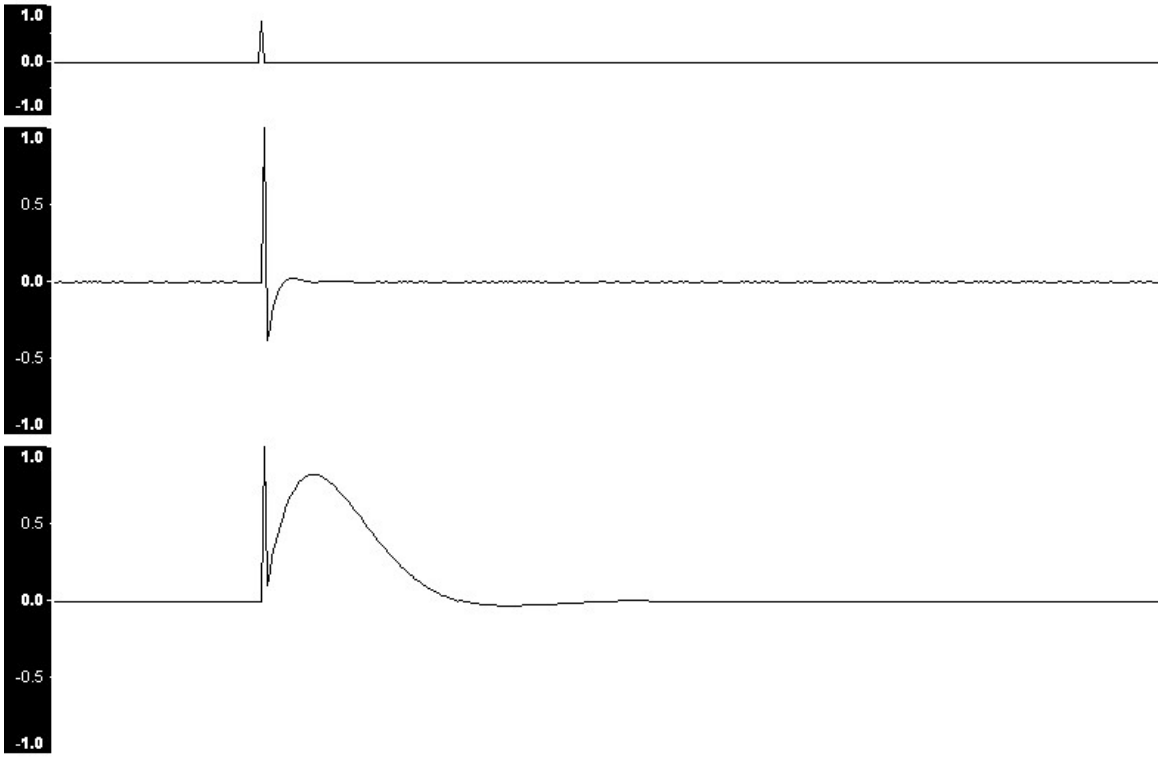
Jitter buffers or de-jitter buffers are used to counter jitter introduced by queuing in packet switched networks so that a continuous playout of audio (or video) transmitted over the network can be ensured. The maximum jitter that can be countered by a de-jitter buffer is equal to the buffering delay introduced before starting the play-out of the mediastream. In the context of packet-switched networks, the term *packet delay variation* is often preferred over *jitter*.

Some systems use sophisticated delay-optimal de-jitter buffers that are capable of adapting the buffering delay to changing network jitter characteristics. These are known as adaptive de-jitter buffers and the adaptation logic is based on the jitter estimates computed from the arrival characteristics of the media packets. Adaptive de-jittering involves introducing discontinuities in the media play-out, which may appear offensive to the listener or viewer. Adaptive de-jittering is usually carried out for audio play-outs that feature a VAD/DTX encoded audio, that allows the lengths of the silence periods to be adjusted, thus minimizing the perceptual impact of the adaptation.

Dejitterizer

A dejitterizer is a device that reduces jitter in a digital signal. A dejitterizer usually consists of an elastic buffer in which the signal is temporarily stored and then retransmitted at a rate based on the average rate of the incoming signal. A dejitterizer is usually ineffective in dealing with low-frequency jitter, such as waiting-time jitter.

Impulse response



The Impulse response from a simple audio system. Showing the original impulse, the response after high frequency boosting, and the response after low frequency boosting.

In signal processing, the **impulse response**, or **impulse response function (IRF)**, of a dynamic system is its output when presented with a brief input signal, called an impulse. More generally, an impulse response refers to the reaction of any dynamic system in response to some external change. In both cases, the impulse response describes the reaction of the system as a function of time (or possibly as a function of some other independent variable that parameterizes the dynamic behavior of the system).

For example, the dynamic system might be a planetary system in orbit around a star; the external influence in this case might be another massive object arriving from elsewhere in the galaxy; the impulse response is the change in the motion of the planetary system caused by interaction with the new object.

In all these cases, the 'dynamic system' and its 'impulse response' may refer to actual physical objects, or to a mathematical system of equations describing these objects.

Mathematical considerations

Mathematically, how the impulse is described depends on whether the system is modeled in discrete or continuous time. The impulse can be modeled as a Dirac delta function for continuous-time systems, or as the Kronecker delta for discrete-time systems. The Dirac delta represents the limiting case of a pulse made very short in time while maintaining its area or integral (thus giving an infinitely high peak). While this is impossible in any real system, it is a useful idealisation. In Fourier analysis theory, such an impulse comprises equal portions of all possible excitation frequencies, which makes it a convenient test probe.

Any system in a large class known as *linear, time-invariant* (LTI) is completely characterized by its impulse response. That is, for any input function, the output function can be calculated in terms of the input and the impulse response. The impulse response of a linear transformation is the image of Dirac's delta function under the transformation, analogous to the fundamental solution of a partial differential operator.

The Laplace transform of the impulse response function is known as the transfer function. It is usually easier to analyze systems using transfer functions as opposed to impulse response functions. The Laplace transform of a system's output may be determined by the multiplication of the transfer function with the input function in the complex plane, also known as the frequency domain. An inverse Laplace transform of this result will yield the output function in the time domain.

To determine an output function directly in the time domain requires the convolution of the input function with the impulse response function. This requires the use of integrals, and is usually more difficult than simply multiplying two functions in the frequency domain.

The impulse response, considered as a Green's function, can be thought of as an "influence function:" how a point of input influences output.

Practical applications

In practical systems, it is not possible to produce a perfect impulse to serve as input for testing; therefore, a brief pulse is sometimes used as an approximation of an impulse. Provided that the pulse is short enough compared to the impulse response, the result will be close to the true, theoretical, impulse response. In many systems, however, driving with a very short strong pulse may drive the system into a nonlinear regime, so instead the system is driven with a pseudo-random sequence, and the impulse response is computed from the input and output signals.

Loudspeakers

An application that demonstrates this idea was the development of impulse response loudspeaker testing in the 1970s. Loudspeakers suffer from phase inaccuracy, a defect unlike other measured properties such as frequency response. Phase inaccuracy is caused by small delayed sounds that are the result of resonance, energy storage in the cone, the internal volume, or the enclosure panels vibrating. Measuring the impulse response, which is a direct plot of this "time-smearing," provided a tool for use in reducing resonances by the use of improved materials for cones and enclosures, as well as changes to the speaker crossover. The need to limit input amplitude to maintain the linearity of the system led to the use of inputs such as pseudo-random maximum length sequences, and to the use of computer processing to derive the impulse response.

Digital filtering

Impulse response is a very important concept in the design of digital filters for audio processing, because digital filters can differ from 'real' filters in often having a pre-echo, which the ear is not accustomed to.

Electronic processing

Impulse response analysis is a major facet of radar, ultrasound imaging, and many areas of digital signal processing. An interesting example would be broadband internet connections. DSL/Broadband services use adaptive equalisation techniques to help compensate for signal distortion and interference introduced by the copper phone lines used to deliver the service.

Control systems

In control theory the impulse response is the response of a system to a Dirac delta input. This proves useful in the analysis of dynamic systems: the Laplace transform of the delta function is 1, so the impulse response is equivalent to the inverse Laplace transform of the system's transfer function.

Acoustic and audio applications

In acoustic and audio applications, impulse responses enable the acoustic characteristics of a location, such as a concert hall, to be captured. Various commercial packages are available containing impulse responses from specific locations, ranging from small rooms to large concert halls. These impulse responses can then be utilized in convolution reverb applications to enable the acoustic characteristics of a particular location to be applied to target audio.

Economics

In economics, and especially in contemporary macroeconomic modeling, impulse response functions describe how the economy reacts over time to exogenous impulses, which economists usually call 'shocks', and are often modeled in the context of a vector autoregression. Impulses that are often treated as exogenous from a macroeconomic point of view include changes in government spending, tax rates, and other fiscal policy parameters; changes in the monetary base or other monetary policy parameters; changes in productivity or other technological parameters; and changes in preferences, such as the degree of impatience. Impulse response functions describe the reaction of endogenous macroeconomic variables such as output, consumption, investment, and employment at the time of the shock and over subsequent points in time.

Chapter 10

Speech Coding and Electroglottograph

Speech coding

Speech coding is the application of data compression of digital audio signals containing speech. Speech coding uses speech-specific parameter estimation using audio signal processing techniques to model the speech signal, combined with generic data compression algorithms to represent the resulting modeled parameters in a compact bitstream.

The two most important applications of speech coding are mobile telephony and Voice over IP.

The techniques used in speech coding are similar to that in audio data compression and audio coding where knowledge in psychoacoustics is used to transmit only data that is relevant to the human auditory system. For example, in voiceband speech coding, only information in the frequency band 400 Hz to 3500 Hz is transmitted but the reconstructed signal is still adequate for intelligibility.

Speech coding differs from other forms of audio coding in that speech is a much simpler signal than most other audio signals, and much a lot more statistical information is available about the properties of speech. As a result, some auditory information which is relevant in audio coding can be unnecessary in the speech coding context. In speech coding, the most important criterion is preservation of intelligibility and "pleasantness" of speech, with a constrained amount of transmitted data.

It should be emphasised that the intelligibility of speech includes, besides the actual literal content, also speaker identity, emotions, intonation, timbre etc. that are all important for perfect intelligibility. The more abstract concept of pleasantness of degraded speech is a different property than intelligibility, since it is possible that degraded speech is completely intelligible, but subjectively annoying to the listener.

In addition, most speech applications require low coding delay, as long coding delays interfere with speech interaction.

Sample companding viewed as a form of speech coding

From this viewpoint, the A-law and μ-law algorithms (G.711) used in traditional PCM digital telephony can be seen as a very early precursor of speech encoding, requiring only 8 bits per sample but giving effectively 12 bits of resolution. Although this would generate unacceptable distortion in a music signal, the peaky nature of speech waveforms, combined with the simple frequency structure of speech as a periodic waveform with a single fundamental frequency with occasional added noise bursts, make these very simple instantaneous compression algorithms acceptable for speech.

A wide variety of other algorithms were tried at the time, mostly variants on delta modulation, but after careful consideration, the A-law/μ-law algorithms were chosen by the designers of the early digital telephony systems. At the time of their design, their 33% bandwidth reduction for a very low complexity made them an excellent engineering compromise. Their audio performance remains acceptable, and there has been no need to replace them in the stationary phone network.

In 2008, G.711.1 codec, which has a scalable structure, was standardized by ITU-T. The input sampling rate is 16 kHz.

Modern speech compression

Much of the later work in speech compression was motivated by military research into digital communications for secure military radios, where very low data rates were required to allow effective operation in a hostile radio environment. At the same time, far more processing power was available, in the form of VLSI integrated circuits, than was available for earlier compression techniques. As a result, modern speech compression algorithms could use far more complex techniques than were available in the 1960s to achieve far higher compression ratios.

These techniques were available through the open research literature to be used for civilian applications, allowing the creation of digital mobile phone networks with substantially higher channel capacities than the analog systems that preceded them.

The most common speech coding scheme is Code Excited Linear Prediction (CELP) coding, which is used for example in the GSM standard. In CELP, the modelling is divided in two stages, a linear predictive stage that models the spectral envelope and code-book based model of the residual of the linear predictive model.

In addition to the actual speech coding of the signal, it is often necessary to use channel coding for transmission, to avoid losses due to transmission errors. Usually, speech coding and channel coding methods have to be chosen in pairs, with the more important bits in the speech data stream protected by more robust channel coding, in order to get the best overall coding results.

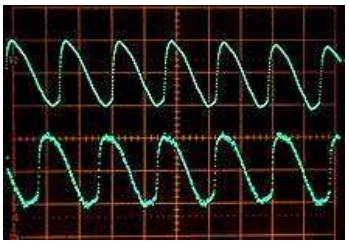
The Speex project is an attempt to create a free software speech coder, unencumbered by patent restrictions.

Major subfields:

- Wide-band speech coding
  - AMR-WB for WCDMA networks
  - VMR-WB for CDMA2000 networks
  - G.722, G.722.1, Speex and others for VoIP and videoconferencing
- Narrow-band speech coding
  - FNBDT for military applications
  - SMV for CDMA networks
  - Full Rate, Half Rate, EFR, AMR for GSM networks
  - G.723.1, G.726, G.728, G.729, iLBC and others for VoIP or videoconferencing

Electroglottograph

Electroglottograph, Glottal Enterprises model EG2-PCX shown here



Photograph of an EGG signal from a Glottal Enterprises EG2-PC (top) and a Laryngograph/Kay electroglottograph (bottom).

Showing the contacts on the electrodes from a Glottal Enterprises EG2-PCX. Electrodes for other electroglottographs are typically very similar in size and shape. This set of electrodes is from a Glottal Enterprises EG2-PCX, which is a dual-channel EGG, so it has 2 sets of contacts. Electrode jelly is used to help conduct the signal from the contacts to the neck.

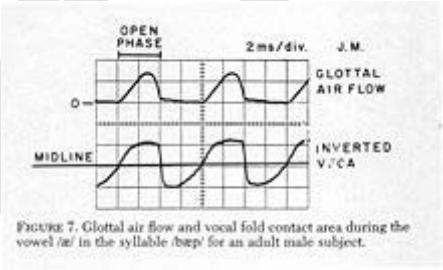


FIGURE 7. Glottal air flow and vocal fold contact area during the vowel /æ/ in the syllable /bæp/ for an adult male subject.

Figure 7 of: Martin Rothenberg and James J. Mashie, *Monitoring Vocal Fold Abduction Through Vocal Fold Contact Area* Journal of Speech and Hearing Research, Volume 31, 338-351, September 1988

The **electroglottograph**, or **EGG**, (sometimes referred to as a laryngograph) is a device for the noninvasive measurement of the time variation of the degree of contact between the vibrating vocal folds during voice production. Though it is difficult to verify the assumption precisely, the aspect of contact being measured by a typical EGG unit is considered to be the vocal fold contact area (VFCA). To measure VFCA, an EGG records variations in the transverse electrical impedance of the larynx and nearby tissues by means of a small A/C electrical current in the megaHertz region applied by electrodes on the surface of the neck. This electrical impedance will vary slightly with the area of contact between the moist vocal folds during that part of the glottal vibratory cycle in which the folds are in contact. However, because the percentage variation in the neck impedance caused by vocal fold contact can be extremely small and varies considerably between subjects, no absolute measure of contact area is obtained, only the pattern of variation for a given subject.

Early commercial available EGG units were compared quite thoroughly by Baken. However, using modern low noise electronics, EGG noise levels can be brought down enough so that the noise is approximately 40dB (a factor of 100) less than a typical EGG signal from an adult voice.

In addition, by the use of multiple channels simultaneously, the technique can be made easier to use and more reliable by giving the user an indication of the correct positioning of the electrodes, and providing a quantitative measure of vertical movements of the larynx during voice production.

Electroglottograph signals have found use in stroboscope synchronization, voice fundamental frequency tracking, tracking vocal fold abductory movements and the study of the singing voice.

Chapter 11

Psychoacoustics

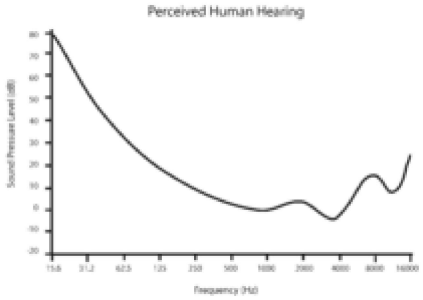
Psychoacoustics is the scientific study of sound perception. More specifically, it is the branch of science studying the psychological and physiological responses associated with sound (including speech and music). It can be further categorized as a branch of Psychophysics.

Background

Hearing is not a purely mechanical phenomenon of wave propagation, but is also a sensory and perceptual event; in other words, when a person hears something, that something arrives at the ear as a mechanical sound wave traveling through the air, but within the ear it is transformed into neural action potentials. These nerve pulses then travel to the brain where they are perceived. Hence, in many problems in acoustics, such as for audio processing, it is advantageous to take into account not just the mechanics of the environment, but also the fact that both the ear and the brain are involved in a person's listening experience.

The inner ear, for example, does significant signal processing in converting sound waveforms into neural stimulus, so certain differences between waveforms may be imperceptible. Audio compression techniques, such as MP3, make use of this fact. In addition, the ear has a nonlinear response to sounds of different loudness levels. Telephone networks and audio noise reduction systems make use of this fact by nonlinearly compressing data samples before transmission, and then expanding them for playback. Another effect of the ear's nonlinear response is that sounds that are close in frequency produce phantom beat notes, or intermodulation distortion products.

Limits of perception



An equal-loudness contour. Note peak sensitivity between 2khz and 4khz, the frequency around which the human voice centers

The human ear can nominally hear sounds in the range 20 Hz to 20,000 Hz (20 kHz). This upper limit tends to decrease with age, most adults being unable to hear above 16 kHz. The ear itself does not respond to frequencies below 20 Hz, but these can be perceived via the body's sense of touch.

Frequency resolution of the ear is 3.6 Hz within the octave of 1,000–2,000 Hz. That is, changes in pitch larger than 3.6 Hz can be perceived in a clinical setting. However, even smaller pitch differences can be perceived through other means. For example, the interference of two pitches can often be heard as a (low-)frequency difference pitch. This effect of phase variance upon the resultant sound is known as beating.

The semitone scale used in Western musical notation is not a linear frequency scale but logarithmic. Other scales have been derived directly from experiments on human hearing perception, such as the mel scale and Bark scale (these are used in studying perception, but not usually in musical composition), and these are approximately logarithmic in frequency at the high-frequency end, but nearly linear at the low-frequency end.

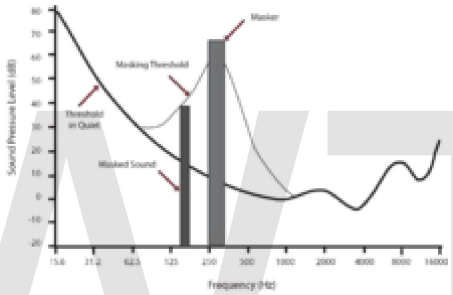
The intensity range of audible sounds is enormous. Our ear drums are sensitive only to variations in the sound pressure, but can detect pressure changes as small as 2×10^{-10} atm and as great as or greater than 1 atm. For this reason, sound pressure level is also measured logarithmically, with all pressures referenced to 1.97385×10^{-10} atm. The lower limit of audibility is therefore defined as 0 dB, but the upper limit is not as clearly defined. The upper limit is more a question of the limit where the ear will be physically harmed or with the potential to cause noise-induced hearing loss.

A more rigorous exploration of the lower limits of audibility determines that the minimum threshold at which a sound can be heard is frequency dependent. By measuring this minimum intensity for testing tones of various frequencies, a frequency dependent absolute threshold of hearing (ATH) curve may be derived. Typically, the ear shows a peak of sensitivity (i.e., its lowest ATH) between 1 kHz and 5 kHz, though the threshold changes with age, with older ears showing decreased sensitivity above 2 kHz.

The ATH is the lowest of the equal-loudness contours. Equal-loudness contours indicate the sound pressure level (dB), over the range of audible frequencies, which are perceived as being of equal loudness. Equal-loudness contours were first measured by Fletcher and Munson at Bell Labs in 1933 using pure tones reproduced via headphones, and the data they collected are called Fletcher-Munson curves. Because subjective loudness was difficult to measure, the Fletcher-Munson curves were averaged over many subjects.

Robinson and Dadson refined the process in 1956 to obtain a new set of equal-loudness curves for a frontal sound source measured in an anechoic chamber. The Robinson-Dadson curves were standardized as ISO 226 in 1986. In 2003, ISO 226 was revised as equal-loudness contour using data collected from 12 international studies.

Masking effects



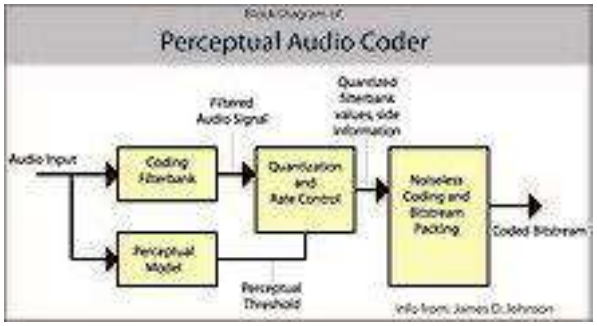
Audio Masking Graph

In some situations an otherwise clearly audible sound can be masked by another sound. For example, conversation at a bus stop can be completely impossible if a loud bus is driving past. This phenomenon is called masking. A weaker sound is masked if it is made inaudible in the presence of a louder sound.

Missing fundamental

A harmonic series of pitches that are related 2×f, 3×f, 4×f, 5×f, etc., give human hearing the psychoacoustic impression that the pitch 1×f is present.

Software



Perceptual Audio Coding uses the Psychoacoustics algorithm

The **psychoacoustic model** provides for high quality lossy signal compression by describing which parts of a given digital audio signal can be removed (or aggressively compressed) safely — that is, without significant losses in the (consciously) perceived quality of the sound.

It can explain how a sharp clap of the hands might seem painfully loud in a quiet library, but is hardly noticeable after a car backfires on a busy, urban street. This provides great benefit to the overall compression ratio, and psychoacoustic analysis routinely leads to compressed music files that are 1/10 to 1/12 the size of high quality masters with very little discernible loss in quality. Such compression is a feature of nearly all modern audio compression formats. Some of these formats include Dolby Digital (AC-3), MP3, Ogg Vorbis, AAC, WMA, MPEG-1 Layer II (used for digital audio broadcasting in several countries) and ATRAC, the compression used in MiniDisc and some Walkman models.

Psychoacoustics is based heavily on human anatomy, especially the ear's limitations in perceiving sound as outlined previously. To summarize, these limitations are:

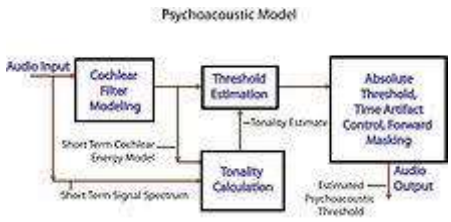
- High frequency limit
- Absolute threshold of hearing
- Temporal masking
- Simultaneous masking

Given that the ear will not be at peak perceptive capacity when dealing with these limitations, a compression algorithm can assign a lower priority to sounds outside the range of human hearing. By carefully shifting bits away from the unimportant components and toward the important ones, the algorithm ensures that the sounds a listener is most likely to perceive are of the highest quality.

Music

Psychoacoustics include topics and studies which are relevant to music psychology and music therapy. Theorists such as Benjamin Boretz consider some of the results of psychoacoustics to be meaningful only in a musical context.

Applied psychoacoustics



Psychoacoustics Model

Psychoacoustics is presently applied within many fields from software development, where developers map proven and experimental mathematical patterns; in digital signal processing, where many audio compression codecs such as MP3 use a psychoacoustic model to increase compression ratios; in the design of (high end) audio systems for accurate reproduction of music in theatres and homes; as well as defense systems where scientists have experimented with limited success in creating new acoustic weapons, which emit frequencies that may impair, harm, or kill. It is also applied today within music, where musicians and artists continue to create new auditory experiences by masking unwanted frequencies of instruments, causing other frequencies to be enhanced. Yet another application is in design of small or lower-quality loudspeakers, which use the phenomenon of missing fundamentals to give the effect of low frequency bass notes that the system, due to frequency limitations, cannot actually reproduce.

Chapter 12

Electronic Fluency Devices

Electronic fluency devices (also known as **assistive devices**, **electronic aids**, **altered auditory feedback devices** and **altered feedback devices**) are electronic devices intended to improve the fluency of persons who stutter. Most electronic fluency devices change the sound of the user's voice in his or her ear.

Electronic fluency device

Types

Electronic fluency devices can be divided into two basic categories.

- Computerized feedback devices provide feedback on the physiological control of respiration and phonation, including loudness, vocal intensity and breathing patterns.
- Altered auditory feedback (AAF) devices alter the speech signal so that speakers hear their voices differently.

Computerized feedback devices

Computerized feedback devices (such as CAFET or Dr. Fluency) use computer technology to increase control over breathing and phonation. A microphone gathers information about the stutterer's speech and feedback is delivered on a computer screen. Measurements include intensity (loudness), voice quality, breathing patterns, and voicing strategies. These programs are designed to train features related to prolonged speech, a treatment technique which is frequently used in stuttering therapy. No peer-reviewed

studies have been published showing the effectiveness of commercial systems in a clinical context. A study of electromyographic (EMG) feedback in children and adolescents found it to be as effective as other treatments (home-based and clinic-based smooth speech training) in the short and longterm.

Altered auditory feedback devices

Altered auditory feedback (AAF) such as singing, choral speaking, masking, delayed or frequency altered feedback have long been known to reduce stuttering. Early altered auditory feedback devices were large and thus confined to the laboratory or therapy room, but advances in electronics have permitted increasingly portable devices such as Derazne Correctophone, the Edinburgh Masker, the Vocaltech Clinical Vocal Feedback Device, the Fluency Master and the SpeechEasy. Current devices may be similar in size and appearance to a hearing aid, including in-the-ear and completely-in-the-canal models.

Masking

White noise masking has been well documented to reduce stuttering. Clinic-based and portable devices, such as the Edinburgh Masker (since discontinued) have been developed to deliver masking, and found that masking was effective in reducing stuttering, though many found that reduction in stuttering faded with time. Interest in masking reduced during the 1980s as a result of studies finding delayed auditory feedback and frequency altered feedback were more effective in reducing stuttering.

Delayed auditory feedback

The effect of delayed auditory feedback (DAF) in reducing stuttering has been noted since the 1950s. A DAF user hears his or her voice in headphones, delayed a fraction of a second. Typical delays are in the 50 millisecond to 200 millisecond range. In stutterers, DAF may produce slow, prolonged but fluent speech. In the 1960s to 1980s, DAF was mainly used to train prolongation and fluency. As the stutterer masters fluent speech skills at a slow speaking rate, the delay is reduced in stages, gradually increasing speaking rate, until the person can speak fluently at a normal speaking rate. It was not until the 1990s that research began to focus on DAF in isolation. Recent studies have moved from longer delays to shorter delays in the 50 millisecond to 75 millisecond range, and have found that speakers can maintain fast rates and achieve increased fluency at these delays. Delayed auditory feedback presented binaurally (i.e. in both ears) is more effective than that presented in monaurally, or in one ear only.

Frequency-altered feedback

Pitch-shifting frequency-altered auditory feedback (FAF) changes the pitch at which the user hears his or her voice. Varying pitch from quarter, half or full octave shift typically results in 55–74% decreases stuttering in short reading tasks. Individuals differ as to direction and extent of the pitch shift required to maximally reduce stuttering. In studies

that gave longer exposure to FAF and used more meaningful daily life tasks such as generating a monologue, only some participants experienced a reduction in stuttering. Initial claims that AAF was more powerful than FAF in reducing stuttering have not been supported by subsequent research. FAF is, like DAF, more effective when presented binaurally.

Effectiveness

Studies have shown that altered auditory feedback (including delayed auditory feedback, frequency altered feedback) as provided by devices such as the Casa Futura School DAF machine or SpeechEasy can immediately reduce stuttering by 40 to 80 per cent in reading tasks. Laboratory studies suggest that reductions in stuttering with an electronic fluency device can occur without a reduced speech rate, and that speech naturalness is often enhanced with AAF. However, the effects of altered feedback are highly individualistic, with some obtaining considerable increases in fluency, while others receive little or no benefit.

A 2006 review of stuttering treatments noted that none of the treatment studies on altered auditory feedback met the criteria for experimental quality. In addition, studies have been critiqued for failing to demonstrate ecological validity; in particular that AAF effects continue over the long term and in everyday speaking situations. The high-profile promotion in the media of devices such as the "SpeechEasy" has been criticized as inappropriate given the lack of scientific evidence for their effectiveness.

There are few published studies on the effect of the AAF in the daily activities of life; studies have mainly examined the effect of AAF on short oral reading tasks, with some studying the giving of a monologue that is usually short in duration. Several studies have produced group results that stutterers using the SpeechEasy show greater reductions in reading than for monologue and conversation. Using AAF was effective in reducing stuttering in scripted telephone calls and giving presentations according to two studies. Another study examining the effects of the SpeechEasy in more naturalistic situations (conversation and asking questions of strangers outside the clinic) found that the SpeechEasy failed to show a significant effect following 6 months of use, though individual subjects varied in their response. A further study examining the use of the device during phone and face to face conversation also found wide variations in stuttering reduction, with just under half exhibiting stable improvement over the course of the 4 months of the study.

While there is evidence of the immediate, short-term effectiveness of AAF devices in reducing stuttering, the longterm effects of altered feedback are unclear. There is some limited experimental data that in some speakers the effect of AAF may fade after a few minutes of exposure, and some anecdotal reports suggest that over time users receive continued but lessened effects from their device. While one group study has reported continued overall reductions in stuttering after a year of daily use of the SpeechEasy on reading and a monologue task, others have found that some participants showed adaptation effects, gaining less benefit from the device after exposure for several months,

including stuttering more with the device than without it. Some studies of various altered auditory feedback devices have noted carryover fluency, i.e. a reduction in stuttering after the stutterer removes an electronic fluency device, while others have not.

The effective of electronic fluency devices as measured by qualitative measures and ratings by stutterers have also been made. Studies show that some stutterers report improved fluency and confidence about speaking, and less severe stuttering and some carryover effects; the device is perceived as being particularly useful on the telephone. They reported that the device was difficult to use in noisy situations as the device amplifies all voices and sounds, and some acclimatization to the use of the device over time. Qualitative reports of satisfaction may be disassociated from more objective measures of fluency: some stutterers who gain little or no benefit from a device based on objective measures rate the device highly, while others who were obtaining benefit on measures of fluency reported negatives opinions about the device.

Use with children

There is little experimental evaluation of the therapeutic effect of AAF on children who stutter: one study noted that effects of FAF were less in children than adults. Given the lack of evidence of its effectiveness, as well as concerns about the impact of altered feedback on developing speech and language systems, some authors have expressed the view that the use of an AAF with children would be unethical.

Causes of altered auditory feedback effects

The precise reasons for the fluency-inducing effects of AAF in stutterers are unknown. Early investigators suggested that those who stutter had an abnormal speech–auditory feedback loop that was corrected or bypassed while speaking under DAF. Later researchers proposed increased fluency was actually caused by the changes in speech production, including slower speech rates, higher pitches and increased loudness, rather than the AAF per se. However, subsequent studies have noted that increased fluency occurred in some stutterers at normal and fast rates using DAF. Some suggest that stuttering is caused by defective auditory processing, and that AAF helps to correct the misperceived rhythmic structure of speech. It has been shown that some stutterers have noted that have atypical auditory anatomy and that DAF improved fluency in these stutterers but not in those with typical anatomy. However, positron emission tomography studies on choral reading in stutterers suggest that AAF also made changes in motor and speech production areas of the brain, as well as the auditory processing areas. Choral reading reduced the overactivity in motor areas that is found with stuttered reading, and largely reversed the left-hemisphere based auditory-system and speech production system underactivation. Noting that the effects of altered feedback vary from person to person and can wear off over time, distraction has also been proposed as a possible cause of stuttering reduction with AAF.

Chapter 13

Microsoft Speech API

The **Speech Application Programming Interface** or **SAPI** is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. To date, a number of versions of the API have been released, which have shipped either as part of a Speech SDK, or as part of the Windows OS itself. Applications that use SAPI include Microsoft Office, Microsoft Agent and Microsoft Speech Server.

In general all versions of the API have been designed such that a software developer can write an application to perform speech recognition and synthesis by using a standard set of interfaces, accessible from a variety of programming languages. In addition, it is possible for a 3rd-party company to produce their own Speech Recognition and Text-To-Speech engines or adapt existing engines to work with SAPI. In principle, as long as these engines conform to the defined interfaces they can be used instead of the Microsoft-supplied engines.

In general the Speech API is a freely-redistributable component which can be shipped with any Windows application that wishes to use speech technology. Many versions (although not all) of the speech recognition and synthesis engines are also freely redistributable.

There have been two main 'families' of the Microsoft Speech API. SAPI versions 1 through 4 are all similar to each other, with extra features in each newer version. SAPI 5 however was a completely new interface, released in 2000. Since then several sub-versions of this API have been released.

Basic architecture

Broadly the Speech API can be viewed as an interface or piece of middleware which sits between *applications* and speech *engines* (recognition and synthesis). In SAPI versions 1 to 4, applications could directly communicate with engines. The API included an abstract *interface definition* which applications and engines conformed to. Applications could also use simplified higher-level objects rather than directly call methods on the engines.

In SAPI 5 however, applications and engines do not directly communicate with each other. Instead each talk to a runtime component (**sapi.dll**). There is an API implemented by this component which applications use, and another set of interfaces for engines.

Typically in SAPI 5 applications issue calls through the API (for example to load a recognition grammar; start recognition; or provide text to be synthesized). The sapi.dll runtime component interprets these commands and processes them, where necessary calling on the engine through the engine interfaces (for example, the loading of a grammar from a file is done in the runtime, but then the grammar data is passed to the recognition engine to actually use in recognition). The recognition and synthesis engines also generate events while processing (for example, to indicate an utterance has been recognized or to indicate word boundaries in the synthesized speech). These pass in the reverse direction, from the engines, through the runtime dll, and on to an *event sink* in the application.

In addition to the actual API definition and runtime dll, other components are shipped with all versions of SAPI to make a complete Speech Software Development Kit. The following components are among those included in most versions of the Speech SDK:

- *API definition files* - in MIDL and as C or C++ header files.
- *Runtime components* - e.g. sapi.dll.
- *Control Panel applet* - to select and configure default speech recognizer and synthesizer.
- *Text-To-Speech engines* in multiple languages.
- *Speech Recognition engines* in multiple languages.
- *Redistributable components* to allow developers to package the engines and runtime with their application code to produce a single installable application.
- *Sample application code*.
- *Sample engines* - implementations of the necessary engine interfaces but with no true speech processing which could be used as a sample for those porting an engine to SAPI.
- *Documentation*.

Versions

Xuedong Huang was a key person who led Microsoft's early SAPI efforts.

SAPI 1-4 API family

SAPI 1

The first version of SAPI was released in 1995, and was supported on Windows 95 and Windows NT 3.51. This version included low-level Direct Speech Recognition and Direct Text To Speech APIs which applications could use to directly control engines, as well as simplified 'higher-level' Voice Command and Voice Talk APIs.

SAPI 2

SAPI 2.0 was released in 1996.

SAPI 3

SAPI 3.0 was released in 1997. It added limited support for dictation speech recognition (discrete speech, not continuous), and additional sample applications and audio sources.

SAPI 4

SAPI 4.0 was released in 1998. This version of SAPI included both the core COM API; together with C++ wrapper classes to make programming from C++ easier; and ActiveX controls to allow drag-and-drop Visual Basic development. This was shipped as part of an SDK that included recognition and synthesis engines. It also shipped (with synthesis engines only) in Windows 2000.

The main components of the SAPI 4 API (which were all available in C++, COM, and ActiveX flavors) were:

- **Voice Command -** high-level objects for command & control speech recognition
- **Voice Dictation -** high-level objects for continuous dictation speech recognition
- **Voice Talk -** high-level objects for speech synthesis
- **Voice Telephony -** objects for writing telephone speech applications
- **Direct Speech Recognition -** objects for direct control of recognition engine
- **Direct Text To Speech -** objects for direct control of synthesis engine
- **Audio objects -** for reading to and from an audio device or file

SAPI 5 API family

The **Speech SDK version 5.0**, incorporating the **SAPI 5.0** runtime was released in 2000. This was a complete redesign from previous versions and neither engines nor applications which used older versions of SAPI could use the new version without considerable modification.

The design of the new API included the concept of strictly separating the application and engine so all calls were routed through the runtime sapi.dll. This change was intended to make the API more 'engine-independent', preventing applications from inadvertently depending on features of a specific engine. In addition this change was aimed at making it much easier to incorporate speech technology into an application by moving some management and initialization code into the runtime.

The new API was initially a pure COM API and could be used easily only from C/C++. Support for VB and scripting languages were added later. Operating systems from Windows 98 and NT 4.0 upwards were supported.

Major features of the API include:

- **Shared Recognizer**. For desktop speech recognition applications, a recognizer object can be used that runs in a separate process (**sapisvr.exe**). All applications

using the shared recognizer communicate with this single instance. This allows sharing of resources, removes contention for the microphone and allows for a global UI for control of all speech applications.

- **In-proc recognizer**. For applications that require explicit control of the recognition process the in-proc recognizer object can be used instead of the shared one.
- **Grammar objects**. Speech grammars are used to specify the words that the recognizer is listening for. SAPI 5 defines an XML markup for specifying a grammar, as well as mechanisms to create them dynamically in code. Methods also exist for instructing the recognizer to load a built-in dictation language model.
- **Voice object**. This performs speech synthesis, producing an audio stream from text. A markup language (similar to XML, but not strictly XML) can be used for controlling the synthesis process.
- **Audio interfaces**. The runtime includes objects for performing speech input from the microphone or speech output to speakers (or any sound device); as well as to and from wave files. It is also possible to write a custom audio object to stream audio to or from a non-standard location.
- **User lexicon object**. This allows custom words and pronunciations to be added by a user or application. These are added to the recognition or synthesis engine's built-in lexicons.
- **Object tokens**. This is a concept allowing recognition and TTS engines, audio objects, lexicons and other categories of object to be registered, enumerated and instantiated in a common way.

SAPI 5.0

This version shipped in late 2000 as part of the Speech SDK version 5.0, together with version 5.0 recognition and synthesis engines. The recognition engines supported continuous dictation and command & control and were released in U.S. English, Japanese and Simplified Chinese versions. In the U.S. English system, special acoustic models were available for children's speech and telephony speech. The synthesis engine was available in English and Chinese. This version of the API and recognition engines also shipped in Microsoft Office XP in 2001.

SAPI 5.1

This version shipped in late 2001 as part of the Speech SDK version 5.1. Automation-compliant interfaces were added to the API to allow use from Visual Basic, scripting languages such as JScript, and managed code. This version of the API and TTS engines was shipped in Windows XP. Windows XP Tablet PC Edition and Office 2003 also include this version, but with a substantially improved version 6 recognition engine and Traditional Chinese.

SAPI 5.2

This was a special version of the API for use only in the Microsoft Speech Server which shipped in 2004. It added support for SRGS and SSML mark-up languages, as well as additional server features and performance improvements. The Speech Server also shipped with the version 6 desktop recognition engine and the version 7 server recognition engine.

SAPI 5.3

This is the version of the API that ships in Windows Vista together with new recognition and synthesis engines. As Windows Speech Recognition is now integrated into the operating system, the Speech SDK and APIs are a part of the Windows SDK. SAPI 5.3 includes the following new features:

- Support for W3C XML speech grammars for recognition and synthesis. The Speech Synthesis Markup Language (SSML) version 1.0 provides the ability to mark up voice characteristics, speed, volume, pitch, emphasis, and pronunciation.
- The Speech Recognition Grammar Specification (SRGS) supports the definition of context-free grammars, with two limitations:
    - It does not support the use of SRGS to specify dual-tone modulated-frequency (touch-tone) grammars.
    - It does not support Augmented Backus–Naur form (ABNF).
- Support for semantic interpretation script within grammars. SAPI 5.3 enables an SRGS grammar to be annotated with JavaScript for semantic interpretation to supplement the recognized text.
- User-Specified shortcuts in lexicons, which is the ability to add a string to the lexicon and associate it with a shortcut word. When dictating, the user can say the shortcut word and the recognizer will return the expanded string.
- Additional functionality and ease-of-programming provided by new types.
- Performance improvements, improved reliability and security.
- Version 8 of the speech recognition engine ("Microsoft Speech Recognizer")

SAPI 5.4

This is an updated version of the API that ships in Windows 7.

SAPI 5 Voices

Microsoft Sam (Speech Articulation Module) is a commonly-shipped SAPI 5 voice. In addition, Microsoft Office XP and Office 2003 installed L&H Michael and Michelle voices. The SAPI 5.1 SDK installs 2 more voices, *Mike* and *Mary*. Windows Vista includes Microsoft Anna which replaces Microsoft Sam. Anna is designed to sound more natural and offer greater intelligibility. The Chinese version of Windows Vista and later Windows client versions also include a female voice named Microsoft Lili. Microsoft

Anna is also installed on Windows XP by Microsoft Streets & Trips 2006 and later versions.

Managed code Speech API

A managed code API ships as part of the .NET Framework 3.0. It has similar functionality to SAPI 5 but is more suitable to be used by managed code applications. The new API is available on Windows XP, Windows Server 2003, Windows Vista, and Windows Server 2008.

The existing SAPI 5 API can also be used from managed code to a limited extent by creating COM Interop code (helper code designed to assist in accessing COM interfaces and classes). This works well in some scenarios however the new API should provide a more seamless experience equivalent to using any other managed code library.

Speech functionality in Windows Vista

Windows Vista includes a number of new speech-related features including:

- Speech control of the full Windows GUI and applications
- New tutorial, microphone wizard, and UI for controlling speech recognition
- New version of the Speech API runtime: SAPI 5.3
- Built-in updated Speech Recognition engine (Version 8)
- New Speech Synthesis engine and SAPI voice Microsoft Anna
- Managed code speech API (codenamed SpeechFX)
- Speech recognition support for 8 languages at release time: U.S. English, U.K. English, traditional Chinese, simplified Chinese, Japanese, German, French and Spanish, with more language to be released later.
- Microsoft Agent most notably, and all other Microsoft speech applications use SAPI 5.

Compatibility

The Speech API is compatible with the following operating systems:

SAPI 5

- Microsoft Windows 7
- Microsoft Windows Vista
- Microsoft Windows 2003
- Microsoft Windows XP
- Microsoft Windows 2000

SAPI 4

- Microsoft Windows Millennium Edition
- Microsoft Windows 98
- Microsoft Windows NT 4.0, Service Pack 6a, in English, Japanese and Simplified Chinese.

Major applications using SAPI

- Microsoft Windows XP Tablet PC Edition includes SAPI 5.1 and speech recognition engines 6.1 for English, Japanese, and Chinese (simplified and traditional)
- Windows Speech Recognition in Windows Vista
- Microsoft Narrator in Windows 2000 and later Windows operating systems
- Microsoft Office XP and Office 2003
- Microsoft Excel 2002, Microsoft Excel 2003, and Microsoft Excel 2007 for speaking spreadsheet data
- Microsoft Voice Command for Windows Pocket PC and Windows Mobile
- Microsoft Plus! Voice Command for Windows Media Player
- Dragon NaturallySpeaking general-purpose speech recognition application
- Adobe Reader uses voice output to read document content
- CoolSpeech, text-to-speech application that reads text aloud from a variety of sources
- Window-Eyes screen reader
- JAWS screen reader
- NVDA open-source screen reader

Libraries using SAPI output

- FastFormat, via its speech_sink
- Pantheios, via its be.speech back-end

Chapter 14

Voice Analysis and Speaker Recognition

Voice analysis

Voice analysis is the study of speech sounds for purposes other than linguistic content, such as in speech recognition. Such studies include mostly medical analysis of the voice i.e. phoniatrics, but also speaker identification. More controversially, some believe that the truthfulness or emotional state of speakers can be determined using Voice Stress Analysis or Layered Voice Analysis.

Typical voice problems

A medical study of the voice can be, for instance, analysis of the voice of patients who have had a polyp removed from his or her vocal cords through an operation. In order to objectively evaluate the improvement in voice quality there has to be some measure of voice quality. An experienced voice therapist can quite reliably evaluate the voice, but this requires extensive training and is still always subjective.

Another active research topic in medical voice analysis is vocal loading evaluation. The vocal cords of a person speaking for an extended period of time will suffer from tiring, that is, the process of speaking exerts a load on the vocal cords where the tissue will suffer from tiring. Among professional voice users (i.e. teachers, sales people) this tiring can cause voice failures and sick leaves. To evaluate these problems vocal loading needs to be objectively measured.

Analysis methods

Voice problems that require voice analysis most commonly originate from the vocal folds or the laryngeal musculature that controls them, since the folds are subject to collision forces with each vibratory cycle and to drying from the air being forced through the small gap between them, and the laryngeal musclature is intensely active during speech or singing and is subject to tiring. However, dynamic analysis of the vocal folds and their movement is physically difficult. The location of the vocal folds effectively prohibits direct, invasive measurement of movement. Less invasive imaging methods such as x-rays or ultrasounds do not work because the vocal cords are surrounded by cartilage which distort image quality. Movements in the vocal cords are rapid, fundamental

frequencies are usually between 80 and 300 Hz, thus preventing usage of ordinary video. Stroboscopic, and high-speed videos provide an option but in order to see the vocal folds, a fiberoptic probe leading to the camera has to be positioned in the throat, which makes speaking difficult. In addition, placing objects in the pharynx usually triggers a gag reflex that stops voicing and closes the larynx. In addition, stroboscopic imaging is only useful when the vocal fold vibratory pattern is closely periodic.

The most important indirect methods are currently inverse filtering of either microphone or oral airflow recordings and electroglottography (EGG). In inverse filtering, the speech sound (the radiated acoustic pressure waveform, as obtained from a microphone) or the oral airflow waveform from a circumferentially vented (CV) mask is recorded outside the mouth and then filtered by a mathematical method to remove the effects of the vocal tract. This method produces an estimate of the waveform of the glottal airflow pulses, which in turn reflect the movements of the vocal folds. The other kind of noninvasive indirect indication of vocal fold motion is the electroglottography, in which electrodes placed on either side of the subject's throat at the level of the vocal folds record the changes in the conductivity of the throat according to how large a portion of the vocal folds are touching each other. It thus yields one-dimensional information of the contact area. Neither inverse filtering nor EGG are sufficient to completely describe the complex 3-dimensional pattern of vocal fold movement, but can provide useful indirect evidence of that movement.

Speaker recognition

Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices.

There is a difference between *speaker recognition* (recognizing **who** is speaking) and *speech recognition* (recognizing **what** is being said). These two terms are frequently confused, as is *voice recognition*. Voice recognition is combination of the two where it uses learned aspects of a speakers voice to determine what is being said - such a system cannot recognise speech from random speakers very accurately, but it can reach high accuracy for individual voices with which it has been trained. In addition, there is a difference between the act of authentication (commonly referred to as **speaker verification** or **speaker authentication**) and identification. Finally, there is a difference between *speaker recognition* (recognizing **who** is speaking) and *speaker diarisation* (recognizing **when** the **same** speaker is speaking).

Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style). Speaker verification has earned speaker recognition its classification as a "behavioral biometric."

Verification versus identification

There are two major applications of *speaker recognition* technologies and methodologies. If the speaker claims to be of a certain identity and the voice is used to verify this claim, this is called *verification* or *authentication*. On the other hand, *identification* is the task of determining an unknown speaker's identity. In a sense *speaker verification* is a 1:1 match where one speaker's voice is matched to one template (also called a "voice print" or "voice model") whereas *speaker identification* is a 1:N match where the voice is compared against N templates.

From a security perspective, identification is different from verification. For example, presenting your passport at border control is a verification process - the agent compares your face to the picture in the document. Conversely, a police officer comparing a sketch of an assailant against a database of previously documented criminals to find the closest match(es) is an identification process.

Speaker verification is usually employed as a "gatekeeper" in order to provide access to a secure system (e.g.: telephone banking). These systems operate with the user's knowledge and typically requires their cooperation. *Speaker identification* systems can also be implemented covertly without the user's knowledge to identify talkers in a discussion, alert automated systems of speaker changes, check if a user is already enrolled in a system, etc.

In forensic applications, it is common to first perform a speaker identification process to create a list of "best matches" and then perform a series of verification processes to determine a conclusive match.

Variants of speaker recognition

Each *speaker recognition* system has two phases: Enrollment and verification. During enrollment, the speaker's voice is recorded and typically a number of features are extracted to form a *voice print*, *template*, or *model*. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print. For identification systems, the utterance is compared against multiple voice prints in order to determine the best match(es) while verification systems compare an utterance against a single voice print. Because of the process involved, verification is faster than identification.

Speaker recognition systems fall into two categories: text-dependent and text-independent.

If the text must be the same for enrollment and verification this is called text-dependent recognition. In a text-dependent system, prompts can either be common across all speakers (e.g.: a common pass phrase) or unique. In addition, the use of shared-secrets

(e.g.: passwords and PINs) or knowledge-based information can be employed in order to create a multi-factor authentication scenario.

Text-independent systems are most often used for speaker identification as they require very little if any cooperation by the speaker. In this case the text during enrollment and test is different. In fact, the enrollment may happen without the user's knowledge, as in the case for many forensic applications. As text-independent technologies do not compare what was said at enrollment and verification, verification applications tend to also employ speech recognition to determine what the user is saying at the point of authentication.

Technology

The various technologies used to process and store *voice prints* include frequency estimation, hidden Markov models, Gaussian mixture models, pattern matching algorithms, neural networks, matrix representation,Vector Quantization and decision trees. Some systems also use "anti-speaker" techniques, such as cohort models, and world models.

Ambient noise levels can impede both collection of the initial and subsequent voice samples. Noise reduction algorithms can be employed to improve accuracy, but incorrect application can have the opposite effect. Performance degradation can result from changes in behavioural attributes of the voice and from enrolment using one telephone and verification on another telephone ("cross channel"). Integration with two-factor authentication products is expected to increase. Voice changes due to ageing may impact system performance over time. Some systems adapt the speaker models after each successful verification to capture such long-term changes in the voice, though there is debate regarding the overall security impact imposed by automated adaptation.

Capture of the biometric is seen as non-invasive. The technology traditionally uses existing microphones and voice transmission technology allowing recognition over long distances via ordinary telephones (wired or wireless).

Digitally recorded audio voice identification and analogue recorded voice identification uses electronic measurements as well as critical listening skills that must be applied by a forensic expert in order for the identification to be accurate.

Chapter 15

Voice Stress Analysis

Voice Stress Analysis (VSA) is a controversial lie detection technology. It has been described as pseudoscientific, and there is no known scientific basis for the underlying theory of "microtremors". Federally funded research showed "little validity" in the technique. A study by Virginia State in 2003, at which time the technique was in widespread use, concluded that "Because there have been no independent scientific studies conducted on the reliability of the computer voice analyzer to detect deception, the Board recommends to the Director of the Department of Professional and Occupational Regulation that computer voice analyzer equipment should not be approved in Virginia at this time.", though a number of academic studies are available which call into question the validity of the technique

There is tension between the voice stress analysis community and the polygraph community, due in the main to the fact that the polygraph is heavily regulated and has been subject to numerous detailed scientific studies, while voice stress analysis is largely unregulated and there are few studies (other than by manufacturers and proponents) which show results better than chance.

VSA technology is said to record psychophysiological stress responses that are present in human voice, when a person suffers psychological stress in response to a stimulus (question) and where the consequences of lying may be dire for the subject being 'tested'.

In the Detection Of Deception (DOD) scenario, the voice-stress produced in response to a Relevant Question ("did you do it?") is referred to as psychological stress or 'deceptive stress'. No DOD technology can detect a lie or truth unequivocally. It is the fear of being exposed as lying to the question being posed that produces the 'high stress' voice signature, aka voice graph or voice tracing.

The technique's accuracy remains debated. There are independent research studies that support the use of VSA as a reliable lie detection technology, whilst there are other studies that dispute its reliability.

VSA is distinct from Layered Voice Analysis (LVA). LVA is used to measure many different components of the voice, but is not reliable in the detection of 'deceptive stress'. LVA measures a wide range of emotions, including excitement, confusion, attention and more. LVA is available in many different forms of products, ranging from server based intelligence use systems, to hand-held devices and standard PC software.

The main difference in the method of operation between LVA and VSA is based on the analyzed frequencies ranges: while VSA focuses on the 8–14 Hz range (which is picked up by specialised microphones), LVA uses a wider spectrum range to extract information that is amusing but not particularly relevant to DOD.

Principle and origins

VSA is based on hypothesis that there are infrasonic components of human voice not audible to observers caused by a physiological phenomenon present in muscles called "microtremor". It was discovered in 1957 by British physiologist Olaf Lippold. Further investigation by other researchers explored the possibility of the presence of microtremor in the muscles controlling the voicebox. The experiment was made by attaching electrodes to the cricothyroid muscle and the posterior cricoarytenoid muscle and measuring EMG signals. Detecting microtremor during sustained speech was not deemed possible because the EMG activity changed too rapidly. The experiment was therefore limited to measuring the presence of microtremor in the frequency range of 1 through 20 Hz in sustained vowel phonation, but yielded no positive results. It was concluded that "the electrical energy was randomly distributed throughout the spectrum." The inconclusive research on microtremor in voice production has consequently been used to claim that the phenomenon can be used for creating technology capable of lie detection by detecting microtremor in recorded speech.

Vendors

The original VSA technology was devised by three former US Army personnel. The three, Bell, McQuiston & Ford, developed the PSE 1, an analogue machine. The same three, working under Dektor Counterintelligence and Security Inc., manufactured the PSE 1000 and later the PSE 2000.

The National Institute Of Truth Verification (NITV, West Palm Beach) then produced and marketed an analogue instrument based on the PSE & digitized it in April 1997, based on the McQuiston-Ford algorithm. In the past 10 years VSA has been used primarily in digital applications: Digital Voice Stress Analysis (D-VSA). The primary suppliers in the USA are NITV-CVSA; Dektor Corporation (no relation to the aforementioned Dektor Counterintelligence and Security Inc.), Diogenes-Lantern, and Baker-FVSA. The primary supplier in Malaysia, Singapore, Brunei, India, etc. is the Australian ITVT Institute (International Truth Verification Technologies) in the form of the Forensic Voice Stress Analyser (FVSA).

The primary use of VSA is in the arena of "Detection Of Deception". As with the polygraph, VSA technology is inert. It has no artificial intelligence component. It is the use of the recorded data as a means for lie detection that remains controversial.

Applications

The purpose of a VSA examination is to determine the truthfulness of responses made by an examinee regarding the subject under investigation. Determinations are made by analyzing and scoring the voice-grams produced by the examinee. Traditional analysis of voice grams was achieved by allocating "percentages of stress" (%) according to the patterns so produced.

High levels of (deceptive) stress indicate that the examinee is deceptive as is the case with polygraph. In respect of VSA, squared voice grams indicates higher stress, whilst 'wave form' or 'domed' signatures indicate less stress.

Questions may be posed to elicit simple "yes" or "no" answers, but can be posed to produce a narrative response. Questions are formulated for each individual being examined to compare situational stress signatures with Control Question and Relevant Question signatures, in order to identify (deceptive) 'stress signatures'.

VSA technology together with validated testing protocols, is designed to protect the innocent and avoid 'false positive' results. VSA is designed to assist any investigation by establishing the veracity of a subject's verbal responses.

Devices used to analyze voice stress are usually used in the presence of the individual under investigation; however, they can also be used without his or her knowledge. Since all that is needed is a voice, a wireless microphone or a tape recording can provide the necessary input signal.

Traditional VSA utilizes the McQuiston-Ford algorithm and this is the technology developed in the USA for the US Defence Agencies and is used by US Law Enforcement agencies.

There are no known physical countermeasures for VSA. Conversely according to Honts *et al.*, the simple use of a 'tack' placed under the tongue of the examinee, to be used as a countermeasure, can reduce the accuracy of polygraph results from 98% to 26%.

Use In law enforcement

A great deal of voice stress testing (VSA) has been conducted. In the United States, most states do not regulate the private use of these devices. However, the CIA and FBI both use VSA at times, in their own investigations. The technology is currently recognized in 43 states.

Many intelligence agencies as well as private forensic psychophysiologists worldwide utilise VSA in preference to polygraph technology.

The X13-VSA technology was originally developed for military use and is now in use by the Italian State Police and the I.C.A.A. (International Crime Analysis Association), as

well by over 150 police agencies and few international airports to screen travelers (X13-VSA PRO Cobra technology).

Methodology and accuracy

The McQuiston-Ford algorithm used for Voice Stress Analysis is reliably accurate. The recorded "micro tremors" in a persons voice are converted via the algorithm into a scorable voice gram. The discrepancy in researched accuracy may result from incorrectly trained or non-trained persons utilizing the technology incorrectly. This is evident by some Polygraphists trying to "test" VSA technology without having received accredited training in the use thereof also by applying Poly Protocols to VSA & vice versa which cannot work.

Recorded cases in Malaysia,Brunei,Singapore,Colombia & recently India have displayed a 100% result with the Australian F V S A which uses a totally new 2006 Developed Algorithm a Scientist & Academic Development

Polygraph-only associations have disputed the accuracy of VSA, although many accredited polygraphists have trained in the use of VSA and use VSA to good effect. The traditional analysis and scoring of voice-grams by means of assigning 'percentages' is time consuming.

In 2002, Clifton Coetzee (Polygraph & VSA Instructor) devised a scoring method for voice grams incorporating the 'UTAH 7 Point' scoring system, as used by modern day polygraphists. Reactive or Responsive patterns are assigned a weighting of +3 to -3.

The use of CQT testing protocols developed by John Reid and Cleve Backster are used for greater reliability of VSA results. It is important that VSA examiners be skilled in the use of enforced, timed pauses between stimulus (question) and response (answer). As in the polygraph situation, the fight or flight response has onset and conclusion delays, which must be considered by examiners to achieve reliable results.

The American Polygraph Association's website lists conclusions from multiple studies into the accuracy of voice stress analysis as a means of detecting the subject's truthfulness. Some researchers or polygraph professionals cast doubt on the validity of the results of such tests; many describe the results as no better than chance.

A study from the U.S Department of Justice showed that VSA performed with a 50% accuracy rate. 2008 W.Carolina University Paper 140 - Eng 101 shows that while VSA works that the FFT & Mcquiston Algorthims do not totally capture (FFT is used for Polygraph also) EMD is more accurate- Authors Prof. Tay , Asst Professor Adams etc.