SOUND FUSION AND THE ACOUSTIC PRESENCE EFFECT

by
Arthur M. Noxon
Acoustic Sciences Corporation
Eugene, Oregon USA

This paper Presented at the 89th AES Convention
1990 September 21-25 • Los Angeles

Summary

This paper discusses the concept of including limited room reflections at the mic. This concept led to the development of ASC's Quick Sound Field, and a whole new way of micing voice and music.

ABSTRACT

In the perception of sound, early reflections are correlated with the direct signal by the listener. Comb coloration effects arise when there are too few specular, coherent reflections. Masking develops with random phase, incoherent reflections. An early arriving, statistically diffuse group, composed of coherent reflections with random time offsets produces excellent sound fusion. Essentially an acoustic presence effect, applications include digital sampling, instrument and vocal recording, and speech therapy for hearing impaired.

0 INTRODUCTION

A clean, direct signal is the most common " signal of choice " in the recording world. The rationale is that any desired effect can always be added later with processing. Even the most primitive, one-man jingle shop has a tiny closet, its interior covered with sound absorptive foam or fiberglass. Inside the "box" is a basic vocal booth, a mic, windscreen and eventually, the talent.

An acoustic system has been developed to saturate the sound fusion (Haas effect) time period with a group of statistically diffuse coherent reflections. Three years ago, the design strategy, mechanical configurations and the acoustic signatures for this technique was introduced at the AES as a digital sampling booth. This acoustic conduction has since been coined QSF, which stands for "Quick Sound Field". Here is presented a follow up report covering some of the applications for this acoustic technique which have developed since its introduction.

1 BACKGROUND

An anechoic recording space may seem simple in concept but it is difficult in practice. Early reflections usually do exist - off of the script stand, paper, window, light fixtures, the floor and other patches of sound reflecting surface. A real-world vocal booth has any number of discrete reflections and resonance problems that add to and color the direct signal. A highly absorptive space that is somewhat acoustically dirty is most difficult for the engineer to mic and for talent to work in.

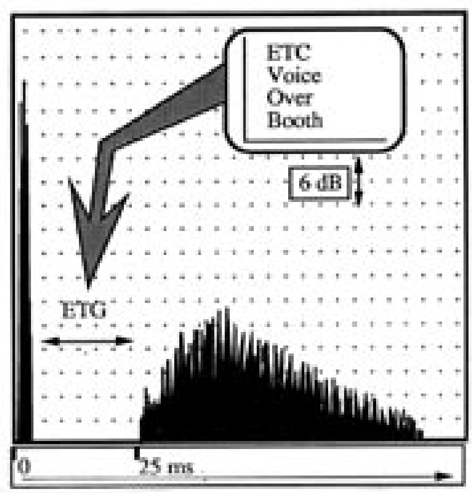
Mic placement is very sensitive to the coloration effects of discrete early reflections and resonance. The sound of the talent is colored by the effects of the mic position. Often, setting up means no more than choosing the best coloration effects. Since consistent sound of an audio track is very important to the engineer, dubs take an inordinate amount of time as the engineer fishes for mic and talent positions in the room, trying to recapture the coloration of the prior day's work.

A dead vocal booth provides little to no acoustic feedback for the talent. Talent suffers sensory deprivation while in the box. A monitor system is essential for talent to be able to adjust intonation in real time. Electronics and earphones are resorted to in the absence of a natural acoustic return. This then further contributes to isolation of the talent in that the direct sound path of their voice is also cut off. By the time traditional recording techniques have been applied, the only natural acoustic feedback left for talent is conduction through the jawbone.

Sensory deprivation and coloration effects found in a typical vocal booth limit its effectiveness. Time is a shortage commodity in the studio. Wasted time in any business, especially the recording studio is to be avoided. The typical vocal booth wastes studio time. Setting up a mic is a delicate time consuming balancing act - talent and mic position vs. room color. Retakes due to a lack of real time acoustic monitoring for the talent takes up additional studio time. A dub is very difficult to set up in order to recapture the original sound. And then, there is the post processing time spent in the effects rack trying to convert the track into a lifelike, naturally bright and open sound.

It is to be expected that the traditional vocal booth will eventually be redefined, steps taken to bolster its positive features and reduce the negative effects. One form of this is accomplished by putting to work the Haas effect in which early reflections are correlated with the direct signal. By arranging for a diffuse group of coherent early reflections, the room coloration effects that appear when there are too few reflections are averaged out. Any low level discrete reflections that might remain are overwhelmed by the diffuse reflections. The diffusion must also be rapidly attenuated in order to not stretch into the echo effect time period, outside of 50 ms. Therefore, in addition to a strong diffusing function, this new class vocal booth must retain a very fast decay rate.

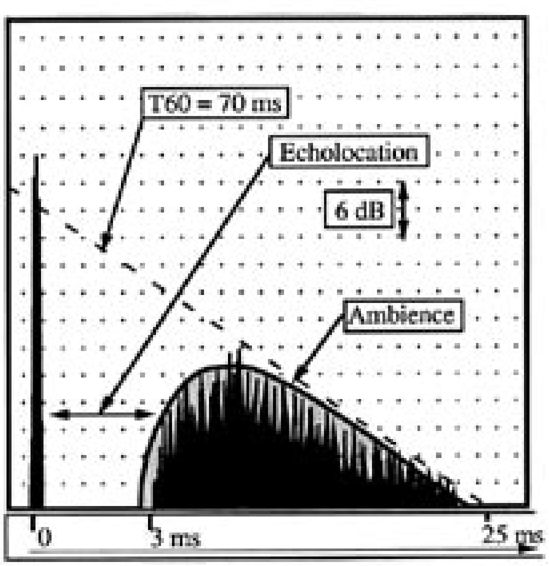
2 ETC - VO BOOTH



The generally recommended ETC for control rooms is a direct signal followed by an early time gap (ETG) due to a reflection-free zone. Outside of this is found a diffuse room ambience with an RT60 of about 1/5 to 1/2 second. The purpose of the ETG is to allow the engineer to hear local colorations of the signal at the mic. It is therefore 50 to 40 ms long, the time of the Haas or sound fusion effect.

The ETC for a voice over (VO) booth has to fit inside of the ETG of the control room. The VO Booth has to be at least 50 dB within the 55 ms ETG. The VO Booth RT60 ought to be on the order of 70 ms.

The only remaining detail is to establish the content of the decay envelope of the VO Booth. There are two phases to the very early reflections. Echolocation cues occur within the first 5 ms. Ambience and coloration effects occupy the balance of the time period.



The direct signal needs to have a 5 ms very early time gap (VETG). This allows time delay phase pan techniques to be used by the engineer. Beyond the echolocation time gap lies the rapidly decaying ambience signal.

If there are just a few discrete reflections, mic ambience is colored due to phase add and cancel effects. If there are no reflections, we have the dead room sound and no ambience. We could have many reflections at the mic. If they are orderly, as with a flutter echo, they would produce coloration. If disorderly, they would create colorless ambience. However, the quality of these reflections needs to be carefully specified.

3 COHERENT OR INCOHERENT REFLECTIONS

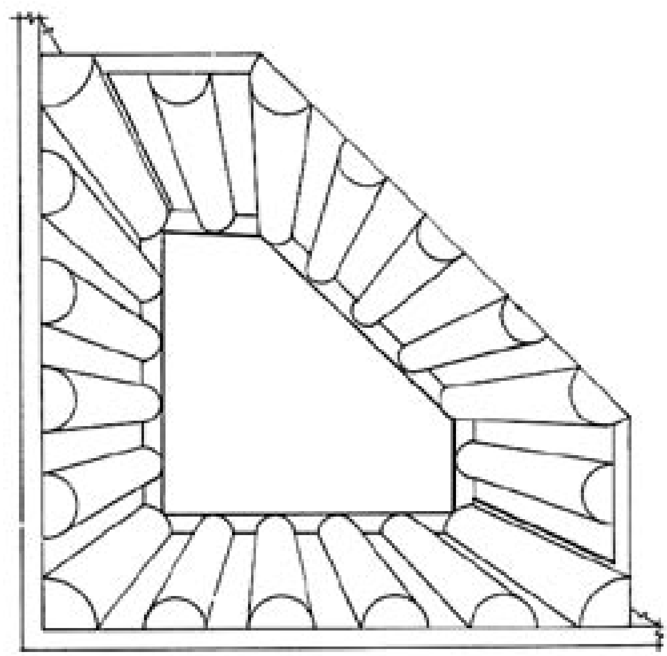
The ear/brain system is a sound processor. But, so is a mic/spectrum analyzer. While they both recognize the spectral character of sound, there are important differences. The ear/brain acts as a correlation type signal detector. The very early reflections are correlated with the direct signal. By this process the early reflections are additive to and enhance the definition of the perceived signal. This is not news - it is the well known Haas, precedence, or sound fusion effect.

On the other hand, a correlation signal processor differentiates between two types of echo. The coherent reflection has a simple time delay offset but otherwise is a phase aligned representation of the direct signal. An incoherent reflection can also be time delayed but is a phase scrambled representation of the direct signal.

A coherent reflection can have the same spectral content as an incoherent reflection. They would look identical to a spectrum analyzer. However, the isolated coherent reflection would produce comb filter, phase add and cancel effects when added to the direct signal. The single incoherent reflection would simply add sound power to the direct signal. In correlation signal enhancement only coherent signals are processed into a spectral display-Incoherent signals such as noise, reverberation and including random phase reflections mask the spectral detail of the direct signal. (This is easily audited by listening to harmonic detail of a plucked guitar string with and without random phase reflections in the rearfield.)

An envelope of statistically diffuse but coherent early reflections that lies within the 50 ms time window of the Haas effect comprises a near field ambience effect that adds to the quality of the direct signal. The composite signal has more top end, is brighter and more natural. It is a more open sound and with air. Statistically diffuse, Haas effect ambience is an acoustic enhancement technique that puts signal that the engineers prefer onto tape.

4 THE HAAS BOX



This class of vocal booth must retain a very fast decay rate and in addition develop a strong diffusion function. It typically has an RT-60 decay time of 80 to 100 ms and a diffusion rate of over 1000 reflections per second. The booth has absorbers and reflectors distributed over its entire interior surface. The component of direct sound that hits a reflector is backscattered, partially back towards the mic, partially into an absorptive strip and partially onto other reflectors. This process uses only specular and diffractive diffusion to maintain the coherent quality in its early reflections.

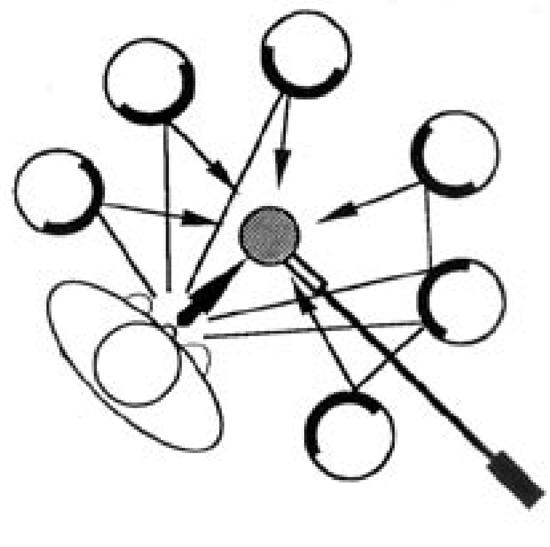
The mean free path in these small rooms is about 4 feet. The broadband absorption coefficient is about 50%. That means the expanding wave front loses about 5 dB every 4 ms. This pencils out to a 60 dB decay in 80 ms and to a 60 dB decay in 80 ms. The wall of such a vocal booth would likely have reflectors alternating with absorption on about 9 inch centers. A 5 foot wide wall would splinter a flat wave front into maybe 7 separately expanding reflections. This sound scattering process continues throughout the decay. The result is easily counted in the ETC and one to two separate reflections per millisecond is the diffusion pate. For all practical purposes, the mic receives a direct signal followed by 4 to 5 ms of no sound; then, as the first arrivals hit, so begins the controlled decay/diffusion process in the room.

A typical vocal booth has a window. In designer studios it would be tilted to not reflect signal into the mic. In a highly diffuse/absorptive room there should not be a large area of untreated reflection regardless of the angle. Current practice in these rooms sees tall, absorptive/reflective wall mounted acoustic units with narrow strips of wall space between. The free wall space between the acoustic control units can easily be glass or plexiglass strips which provides a mope open feeling in an otherwise small room. Visual openness contributes to mope comfort for the talent in long recording sessions.

The statistical populated envelope of very early, coherent reflections is essential to the stability of the acoustic space inside the booth. Engineers report a wide and smooth acoustic space. They even lose track of which mic is open and have to mark the faders. Usually, in a more traditional rooms an engineer simply hears which mic is where. In a statistically diffuse space, the mic position can be changed without changing the envelope. It is the envelope that is distinguishable and not its internal detail. Moving the mic only changes the fine structure as to which reflection arrives when and how strongly. This does not change the statistical envelope or the quality of sound. In a room 4 foot by 6 foot, there would be a 2 x 4 foot central area in which the sound remains uniform, regardless of mic or sound source location.

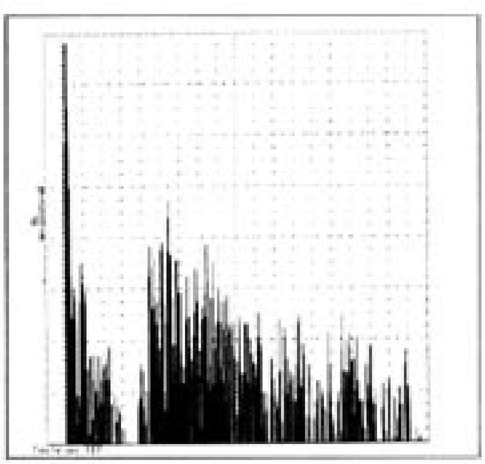
The floor plane is a large reflecting surface. It is left untreated, to be an acoustic mirror effectively doubling the height of the room. Ceiling treatment must be accordingly more severe to keep the vertical decay and diffusion rate up with that of the walls.

5 VOICE OVER GOBO



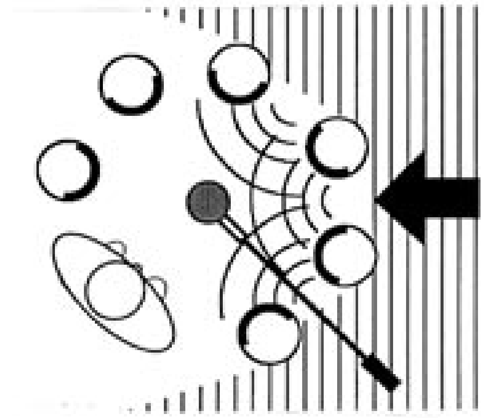
The Haas ambience effect can be approximated out in the open room or field - of course, not to the degree available in an iso booth format, but this QSF gobo setup boosts the signal to noise ratio at the mic by 5 to 7 dBA. This is accomplished by increasing the “direct” signal strength I to 2 dB while reducing the room noise by 4 to 5 dB.

This “ gobo “ is not the large, flat rug-covered plywood gobo of years past. The present method is to use a set of 7 to 9 sound control units, typically placed on 18” centers in a horseshoe pattern. The mic is located in the middle and the talent occupies the open heel end of the pattern. These Traps have two sides. The broadband absorptive side faces outward to intercept inbound room noise and reflections. The membrane reflective side, effective 400 Hz and above, faces inward to produce the statistical group of early coherent reflections. In this system, absorption is replaced by transmission. Sound is not absorbed between the reflectors. It is leaked out of the space. In either case controlled decay and diffusion take place.



Gain of the “direct” signal is accomplished by adding very early multiple reflections of the direct signal to the direct signal. This is completed within the first 50 ms of the sound fusion time period. Although sound fusion generally lasts 50 to 60 ms, a “smearing” accompanies the presence of strong, late high frequency reflections. This is undesirable for the recording engineer. The end of the sound fusion period marks the onset of echo detection. For lower frequencies the echo onset time is later and for highs, sooner than 50 ms.

In the QSF method of developing the statistical ambience, the comb filter effect associated with any individual reflection does not occur due to the large number of random time offset reflections. With 20 to 50 reflections occupying a time span of 20 to 25 ms, the comb filter effect that would arise with any one reflection is obscured by the averaging effect of the other reflections.

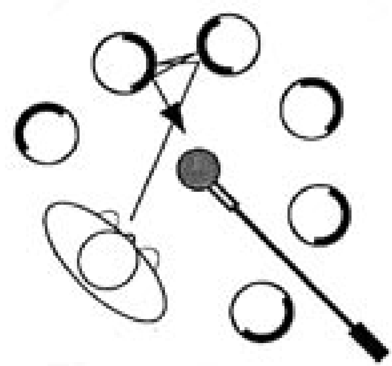


A good signal at the mic can be time delayed for stereo phase pan positioning. The echolocation process occurs within the first 5 ms following the direct signal. Because of the distance between the mic and the reflecting side of the gobo, no reflections arrive within the first 5 ms. The direct signal is well isolated for control in the mix.

Not only is the direct signal enhanced but the ambient noise floor is reduced at the mic by this technique. The backside of each Trap is broadband absorptive and facing outwards towards the room. Sound in the room is absorbed before it gets to the mic.

Sound that does penetrate the perimeter is weakened because the wavelet expands due to diffractive edge effects. Easily a 5 dBA noise level reduction is noted inside the gobo. There may be times when a stronger signal to room noise is required. The closer the Traps are to each other the less outside noise they will let in so the direct signal becomes stronger.

Noise in a room also originates with the talent. Sound does leak out between the traps. Some of this is attenuated by the absorptive half of the trap and the remainder expands rapidly due to edge diffraction effects. The sound leaked to the room is rapidly diffusing. The important feature is that a sound from such a gobo produces no flutter effect. Sound that does bounce off a wall is absorbed by the backside of the gobo traps. The system can also be used near walls with minimal impact.



Incidentally, another application of such a gobo system takes advantage of its reversibility. If all the Traps are rotated then the full bandwidth absorptive side faces the mic. This creates the traditional dead sounding vocal booth. By adjusting a pair of reflectors slightly inward, the interior diffusive top end can be brought up. This is best done in pairs to take advantage of diffusive multiple scattering available from facing reflecting surfaces.

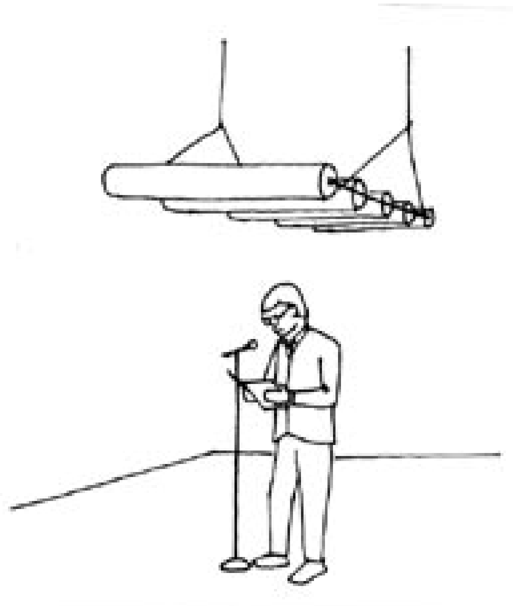
6 NEWSCAST, FILM

Dubbing, retakes, voice synch and voice over are not without impact on the continuity of the audio track. Both off and on camera vocal tracks are highly sensitive to room acoustic colorations. The variations range from outdoor shots to vocal tracks taken in a closet. One of the audio engineer’s jobs is to maintain consistency between the audio and video tracks.

A good example of consistency control is found in the production of a television news story. Here the local personality, holding a mic comments on some disaster which is to be seen across their shoulder. The camera pans to another view and the narration continues. The off camera vocal track is not an on-site recording. It was composed and produced back at the TV studio. The voice over simply does not sound the same as that recorded in the field. The life like sound of the omni mic in the free field highly contrasts with the hyper mic used in the small, semi-dead voice over booth.

The first major TV station to use QSF acoustics is KTVU, Channel 2 in Oakland, California. It has an award winning news show. Part of their formula for success is the QSF vocal booth technique. For three years (since 1987) they have been using the rapid diffusion/decay rates of a QSF booth where they open up an omni mic in a tiny room 4’ x 6’ in size and get a voice over mixed with the background sounds that almost perfectly matches full recordings in the field.

7 FLYING GOBOS

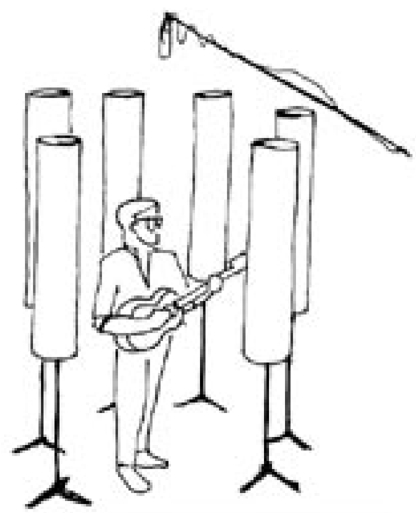


A example of another notable application of the QSF technique involves the very well known show biz voice of Ed McMahon. Remote shoots of his commercials caused a variance in the timbre of his voice. This was unacceptable to the producers as his close mic’d voice was too well known. One of two choices remained, either voice sync everything or stabilize the acoustics of his shots.

It’s expensive and time consuming to voice sync. Instead, the crew rigged an acoustic cloud using the QSF gobo format, flying just out of camera shot. The boom mic was just below the QSF cloud and the track sounded great. The flying gobo essentially blocked the intrusion of the overhead reverberant sound and provided quite a few early reflections to help smooth over the table and floor bounce effects.

This technique is also valuable in the high bay studios. A flying cloud over the free-standing QSF gobo has not only lateral isolation and enhancement but adds in the vertical component for even better isolation and diffusion.

8 REMOTE RECORDING



Recordings in a large reverberant space can be beautiful but can also be terrible. The engineer tends to move back from the instrument trying for a more natural instrument sound but too quickly runs into the strength of the reverberant field. The QSF gobo method eases the fit by enhancing the direct signal strength while softening the level of the reverberation.

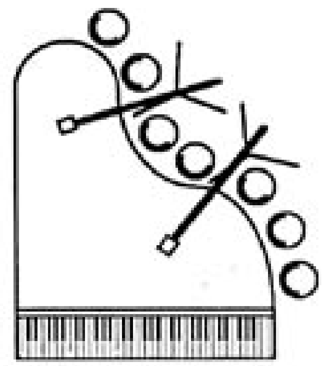
The mic is often high overhead. The traps are still set in a semicircle pattern. The variables are the height of the instrument, the height of the traps and the height of the mic. The higher the traps are off the ground, the more energy leaks directly to the reverberant field. The higher the mic is, above and outside the traps, the more it hears the reverb.

The reflections of the talent back into the mic produce the diffusive group of early reflections. An additional feature is that the reverberant field is more quickly diffusive due to the spoked nature of the sound source after it passes through the gobo.

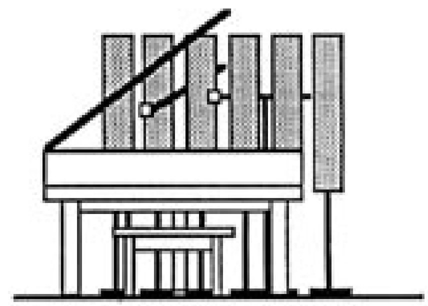
9 PIANO GOBO

The piano mic is often inside the open lid of the piano top. Unfortunately this lid/sound board is not only good for projecting sound out, but also good at collecting sound from the outside which the mic also picks up. Often, the piano is draped in moving blankets. A gobo can also be used to increase the signal at the mic and reduce the intrusion of noise and echoes.

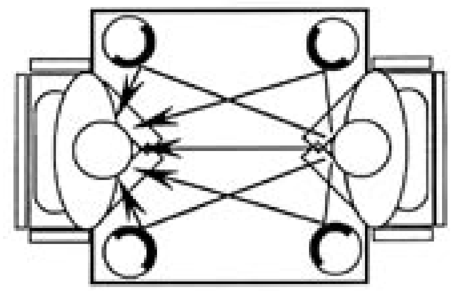
The QSF gobo for a piano sets traps along the open side of the piano with the reflectors inwards. Interior sounds are multiple reflected while sound from the outside is attenuated. Mic placement can be eased away from the strings without degrading the recording.



An aside: Piano sound boards are essentially parallel to the floor. This allows the setting up of standing waves. Very irregular loudness effects in the middle C octave and above are directly attributable to this effect. About 1/5 of professional piano practice setups include some sound damping material under the piano. Best results are developed with acoustic materials that scatter mids and highs but damp 200 to 400 Hz. A couple of traps used for gobo purposes can be placed on the floor below the piano, with reflectors up for best results.

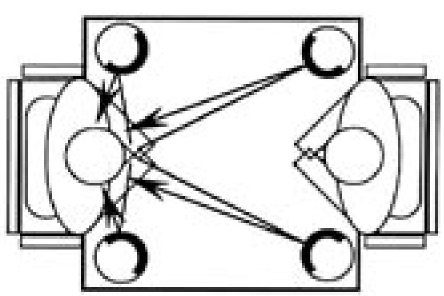


10 SPEECH TRAINING GOBO



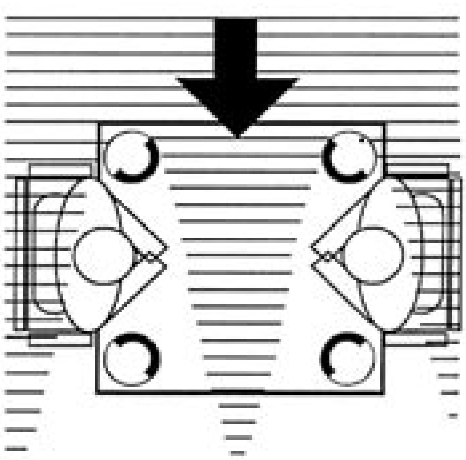
This experiment was performed with the speech therapy department at the University of California, San Diego. A speech training table is typically set out in a larger room. It is the size of a card table. Four Traps were added with the reflectors facing inward. Speech training ensued with the hearing impaired.

The single bounce off the table creates a comb filter effect. By adding 4 additional reflectors the comb coloration effects reduced. The listener tries to imitate sounds. To mimic comb filtered speech is inappropriate and would be judged incorrect by the teacher. The multiple reflected signal is a more honest and accurate signal to imitate.



When speaking, the student hears the sound of their own voice better due to backscattering off the reflectors. Acoustic feedback increases the rate of learning. The teacher and student both speak through the same “chamber”. The teacher can also better hear detail in the speech of the student due to the enhanced acoustic coupling by this set up.

Hearing impaired seem to hear better with one ear oriented directly toward the speaker. In general, they are also very susceptible to distraction by sounds coming in from the side. In this configuration, sound from the side is blocked and replaced with a reflection of the speaker’s voice. Traps to the left and right of the speaker also block room noise and help the speaker’s voice stand out more distinctly. Additionally, the traps near the speaker help to block directionally competing sound. Not only does the gobo signal enhance speech but it reduces distraction.



An interesting aside is that these Traps form somewhat of a “blinder” for the student. Hearing impaired are easily visually distracted, as are the learning disabled. This table top gobo provides a substantial degree of visual barrier effect allowing the student and teacher to be in better contact.

11 INTELLIGIBLE LISTENING

The dynamic level of listening for the hearing impaired is limited and compressed. For example at 50 dB,A, sound might really be at the threshold of audibility for some frequency range yet at 80 dB,A the weak frequency range would sound as loud as with any unimpaired ear. In addition to spectral threshold and dynamic range problems the hearing impaired very often lose the ability to discriminate one sound above others in a crowded, noisy sound field. Echolocation is the ability to corollate the sounds from some particular direction using signals into both ears. Current practice in hearing aids is limited only to bandwidth level controls. Phase and time alignments are not yet available.

The saturated sound fusion acoustic space provides remarkable listening benefit for the hearing impaired. This is most especially evident in contrast to the hearing aid and the mic/headset options presently in common use.

All three factors found in the QSF acoustic space are an aid to listening.

1. The RT-60 is fast, both external and internal room noise is rapidly attenuated so as to not be competitive with the signal.

2. The sound fusion window is saturated with diffuse ambience so the direct sound is enhanced not in level but over time.

3. The acoustic space is wide and smooth so the received signal remains consistent despite variations in listening and speaking position.

Because the hearing impaired are quite vulnerable to room noise masking of the intended signal, a quiet and non-reverberant room is always recommended for better listening results. Intelligibility tests show an inverse correlation between intelligibility scores and RT-60. This leads to the conclusion that an anechoic space is the best space for listening.

Although the tendency for better listening performance lies in the anechoic direction, it is agreed by researchers that some ambience is better for listening than none. This leads to the curious conclusion that a less than “perfect” intelligibility rating might actually be the more intelligible for listening. Not only has this come up in the field of speech intelligibility* but in the present work.

The reflections that could enhance intelligibility should not be echo effects, outside the time windows of 50 ms. They would have to be very early reflections that are correlated with the direct signal. Again, reflections used in correlation processes must be coherent, not random phase type reflections. In intelligibility testing, for example with Techron, % Alcons work, the direct/reverb ratio must be established. The D/R ratio varies depending on how many milliseconds after the direct signal that the cursor for calculation is located. It is generally agreed that +50 to 70 ms is a good location for D/R ratio calculations.

12 HIGH INTELLIGIBILITY LISTENING ROOM

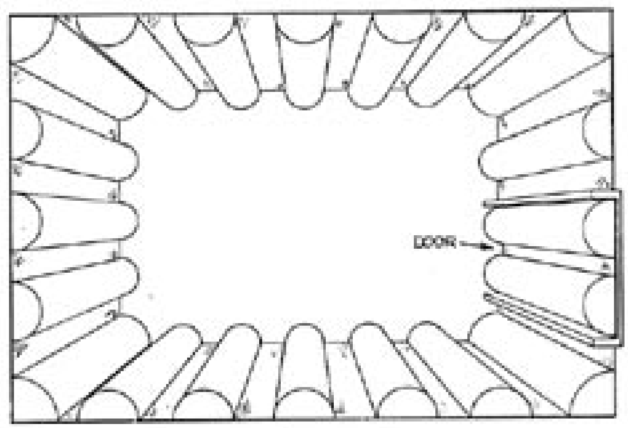
Beyond the technical aspect of intelligibility measurements is the autonomic response characteristics of the human listener. In the dry anechoic room, without reflections, sound levels vary almost violently. This is evidenced by the continual contraction of inner ear muscles. Loud sounds cause an autonomic flinch reaction by the muscles of the inner pap that protect the ear from damage due to further loud sounds. The Fletcher Munsion curves are ample evidence of this limiter action. However, the reaction time for this process is 1/10 second. The staccato of speech in a very dry room produces a rapid sequence of 70 dB level changes and the listener is plagued with a distracting and tiring flinch reaction.

* Peter Mapp, Private Communication, June APRS, London 90

The saturated sound fusion window of 50 to 40 ms provides just enough ambience to reduce the autonomic flinch responses to a minimal level. But even here the lack of echoic ambience results in some dynamic level flinch, a loudness suppression reaction. For non-recording purposes the diffusive ambience probably should be stretched to 100 ms~ lightly into the echo region.

The statistical sound field of a diffusive space allows the listener and/or speaker to move positions and the received sound to remain the same. An inexperienced speaker using a microphone that is hardwired or IR coupled into headsets injects a new problem into communication - mic position relative to the speaker. For the teacher of a group of hearing impaired to wear headphones inhibits the ability to listen to the students' response. The students cannot be mic'd and so the teacher does not wear the headset.

Headphones are promoted for learning speech because they reduce the intrusion of unwanted and distracting signal. But they exclude the airborne sound of the student's voice to oneself. Ultimately, this will be the primary feedback mechanism one has aside from bone conduction.



In the coherent diffusive sound field of the QSF space, each student can wear hearing aids and hear the teacher relatively equally, irrespective of student or teacher position. Conversely, the teacher hears a student's speech as best as possible, so does each student. This is a direct consequence of the wide, open statistically diffuse, ambient sound field of the QSF technique.

The student has to practice speech in a space that allows them to hear themselves on their own hearing aids. It is the only feedback system they will carry with them into the "real world". Hearing aids pick up a lot of room ambience in a regular room. There is no ambience in a dry room. The Quick Sound Field technique is an acoustic space that minimizes room noise and maximizes acoustic feedback. Not only in promise but in practice the QSF type sound field is a significant contribution towards improving the quality of life for the hearing impaired.

13 HISTORY

The very first QSF space was assembled spontaneously by one recording engineer in 1984 who happened to have a lot of modular acoustic units at hand in the studio. Since then many spontaneous QSF gobo applications have been reported by users in the field. The only common ingredient is that a variety of modular acoustic units were available and the engineer's ear leads the way.

The formal development of the QSF method took place in 1987 when a sampling booth design was needed. The traditional dry room gave unacceptable signal for this level of processing. The acoustic signature of the first booth was successful and a lighter weight version of the booth continues in its place. The original booth has found a home in a west coast recording school.

Within months after the initial QSF sampling booth was produced, a number of studio projects included this booth technique. Each time, the vocal talent in the studio discovered this new acoustic space, they insisted on recording in it. Often, a second smaller vocal booth had to be built to put the sampling room back on line with its intended use. The most notable QSF sampling room was picked up by musician Pete Townshend of the WHO. He was so impressed with this new acoustic space that he endorsed it, without remuneration to help encourage other engineers to try it. His room remains at Eel Pie Studios just outside of London.

Because of the early pioneering efforts of forward thinking, recording engineers, the Quick Sound Field has developed into a dependable recording technique.

14 CONCLUSION

The acoustic signature of colorless “ambience” has been presented. Acoustic systems to produce this effect have been developed, with five years of experience and testing in a wide variety of applications. The consensus is that this Quick Sound Field method is a major improvement in mic technique.

The Quick Sound Field method uses the Haas Effect - this time involving statistical diffuse reflections instead of discrete reflections. The QSF establishes a fundamental distinction between coherent and incoherent reflections. While random phase, incoherent reflections may be acceptable in the realm of echo control, they produce masking within the Haas sound fusion time period.

Historically, an inordinant amount of technical expertise and effort has been focused on the control room acoustic. Now, attention turns to the mic. The Quick Sound Field is the next logical step in comprehensive development of the studio acoustic.