# Tor Halmrast: Sam Phillips´ Slap Back Echo;

## Luckily in Mono

**Abstract**

'Slap back echo' was created by Sam Phillips for Elvis Presley´s Sun recordings and later simulated by RCA. Using cepstrum and autocorrelation, we find that the tape delay used in Sun Studios was 134 - 137 ms, which is so long that the echo is perceived mainly as a distinct echo in the time domain, more than a coloration of timbre in the frequency domain. Even though the delay time is long, the echo is still perceived as rather "intimate", because the echo is in mono. Panned in stereo, the feeling of being inside a quite small room would disappear. The simulations by RCA, using a hallway instead of tape echo, shows somewhat shorter and less pronounced delays.

## *Introduction*

Elvis Presley´s first recordings were done by Sam Philips in Sun Studios in Memphis. He created a tape echo by using two Ampex 350 tape recorders. This method was used for, amongst others, *Tryin´ to Get to You,* and *Baby Let´s Play House* from 1955. The story goes that when Elvis changed to the bigger company RCA Victor´s studio on McGavock St., Nashville, *"RCA was anxious to recreate the 'slapback' echo effect that Sam Phillips had created at Sun. To add them to Elvis' vocals Chet [Atkins] and engineer Bob Farris created a pseudo 'echo chamber' by setting up a speaker at one end of a long hallway and a microphone at the other end and recording the echo live"*[1]. This was done for the recording of *Heartbreak Hotel* (Jan. 1956). A third type of echo/reverb analysed, was for the RCA Hollywood recordings of amongst others *Long Tall Sally* (Sept. 1956).

The delay time of the slap back echo used by Sam Philips has been discussed, mostly on web sites. Steven Trent[2] argues that *"The delay time should be between 60 and 120 milliseconds",* and that: *"At 60 Milliseconds, you don't hear much doubling, but everything sounds a little thicker and*

---

[1] http://www.history-of-rock.com/sam_phillips_sun_records.htm
[2] https://stevenrtrent.com/omblog/blog/what-is-slapback-echo

*fatter"*, and that *"At 120 Milliseconds you are really starting to hear the distinct doubling of the sound"*. Elvispresleymusic.com[3] discusses a suggestion of 530 ms, but find that to be *"too long"*, and argues that 163 ms *"seemed to sound about right"*. The Wikipedia page on Slapback Echo mentions *"75-250 ms, with little or no feedback"*.

Pictures and info from the studios[4] indicate that Sun Studios has acoustic tiles on most walls, probably high-/mid-frequency absorbers, and the room is described as so dry that *"When you speak, you can feel the air pressure in the room"*. Some bass resonance problems have been reported, (as such tiles seldom absorb lower frequencies), but this should not have influenced the frequency range of Elvis´ voice too much.[5] Pictures from RCA Nashville, from the web site[6] show thick curtains and what seems to be mid-/high frequency wall absorbers, so the conclusion must be that there should be no possibility of getting a natural, distinct, separate, late echo for a vocal inside these rooms.

## Overview

Searching for recordings with Elvis unaccompanied 'a cappella' sections, we found two examples from Sun Records *(Baby Let´s Play House*, and *Tryin´ to Get to You*), and two from the early RCA period *(Heartbreak Hotel* and *Long Tall Sally)*. For details about the recordings, see References. We investigated the presence of echo in the recordings by analysing Power Cepstrum and Autocorrelation. Some additional analyses included Linear Predictive Coding and reverberation times (or more correctly: decay times, as the decay is far from linear). The analyses were performed in Praat, Pure Data, MIR Toolbox (MatLab), Librosa (Python) and Arta.

The paper starts with the investigations of the unaccompanied 'a cappella' parts of Elvis´ Sun recordings, and here we find that they have one distinct, single reflection with a delay time of 134 - 137 ms. For the RCA recordings analysed, the decay is more smooth (not giving such distinct echo) and the delay times for the strongest reflections are somewhat shorter; 82 ms in Nashville and 117 ms in Hollywood. More technical information regarding the analysis methods; Comb Filters, Cepstrum and Autocorrelation is given in a special section.

Even if the delay times are very long, the recordings are still perceived as being quite 'intimate', so, in the next section we investigate if the measured time delays could provide what is defined as comb filter coloration, and we

---

[3] http://www.elvispresleymusic.com.au/...is_at_sun.html
[4] http://www.history-of-rock.com/sam_phillips_sun_records.htm
[5] http://www.scottymoore.net/studios.html
[6] http://www.history-of-rock.com/sam_phillips_sun_records.htm

find that the delay for Sun Studios is so long that the 'boxy' sounding timbre 'Box-Klangfarbe' as defined by Halmrast (1999 and 2000) is not perceived. (*Klangfarbe* is German for *Timbre*). Also the delay times for RCA are too long to provide such Box-Klangfarbe, but for these recordings, the overall reverberation decay somewhat masks the measured delayed 'reflection', so that a clear, distinct echo is not as easily perceived as for the Sun recordings. Even if the delay times are so long that we do not get the kind of 'close-ness'/'intimacy' typical for Box-Klangfarbe comb filter coloration, Elvis´ voice is clearly perceived as being rather close, or at least inside a small room. In the final part we discuss that the reason for this is that the recordings are in mono, not stereo, which has been verified in a simulation/listening test.

## Sun recordings

Two recordings from Sun Studio with audible slap back echoes were analysed: *Tryin´ To Get To You* and *Baby Let´s Play House*.

### Tryin´ to Get to You

The Power Cepstrum of the 'a cappella' opening of *Tryin´ to Get to You* is shown in figure 1. We see that the delay time for the echo is 0.137 s = 137 ms, which is confirmed also in the Autocorrelation analysis in the lower pane of figure. 1. The x-axis of the Power Cepstrum is shown in seconds, but as explained later, quefrency along the x-axis in figure 1, upper pane, is a 'sort of time axis'. In the theoretical part we explain that also pitch provides results in cepstrum analysis, so we have to check that the pitch is separated from our detection of echo in figure 1. In figure 2 we see that the
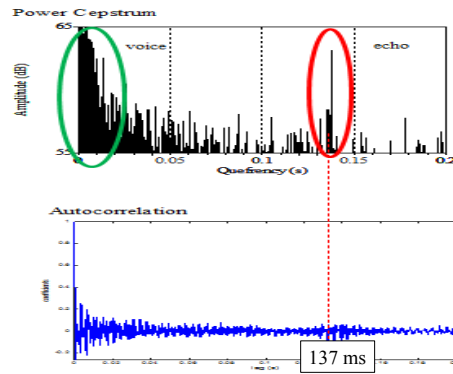


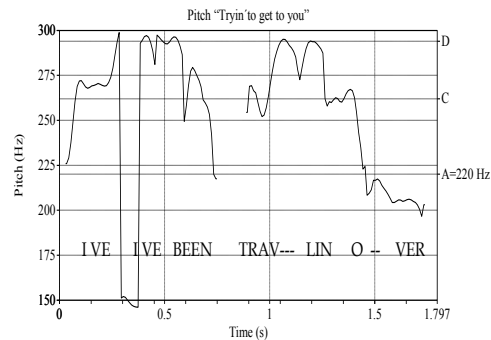Figure 1: Power Cepstrogram, Power Cepstrum and Autocorrelation for "Tryin´.



Figure 2: Pitch, for "Tryin´ To Get To You".

fundamental pitch is mainly in the region $f_o$ = 250-300 Hz, corresponding to quefrencies ($1/f_o$,) between 1/250 = 0.004 s = 4 ms and 1/300 = 0.0033 s =

3.3 ms. From the upper pane in figure 1 we see that the 'voice' thus occupies the lower part of the Power Cepstrum.

A delay time of 137 ms corresponds to an excess travelling path for the reflected sound of 343 x 0.137 = 47 m, assuming a speed of sound of 343 m/s. If you imagine standing in front of a wall, listening to the echo back to yourself, such a delay time means that the wall must be 47/2 = 23.5 m away.

## Baby Let´s Play House

The pitch analysis to the left in figure 3 is inaccurate, due to the very short syllables in the "a capella" opening of *Baby Let´s Play House*. We could have performed additional analysis with a shorter window for the Fast Fourier Transform, but then the uncertainty in frequency would increase, and, as long as we know the main pitches of Elvis voice in this part, we have enough information to know that we have separation of the voice and reflections in the cepstrum analysis. From the panes to the right we see that the delay time for the slap back echo is 134 m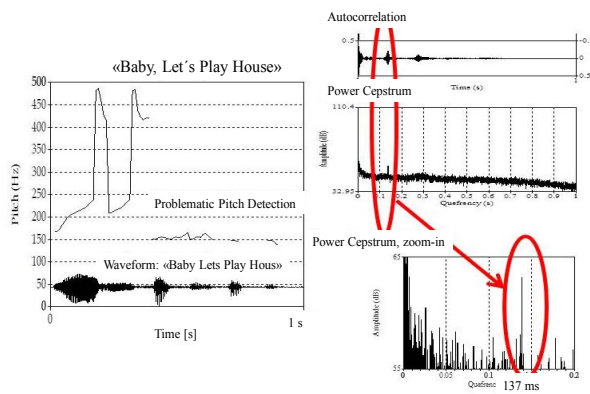s. We find indications of 134 ms also by looking at the Acoustical Energy Decay for the 250 Hz octave, which is shown in figure 4, together with the corresponding Schroeder curve, (backwards integration of the decay), but this is not a clear method for analysing echoes. Reverberation time for the decay of a voice burst including an echo is not really defined,



Figure 3: Pitch and waveform, Autocorrelation, Power Cepstrum and Power Cepstrum Zoom-In of "Baby Let´s Play House".

but for this 1/1 octave band around 250 Hz (which includes the fundamental frequency of this sound file), T30 is app. 0.4 seconds. (If such a reverberation time was measured in a real room, it would be categorised as something between a well-furnished living room and a dampened studio). The small difference between 134 ms for this recording and 137 ms for *Tryin' to Get to You is* might be due to measurements uncertainty, possible small stretch of tape etc.
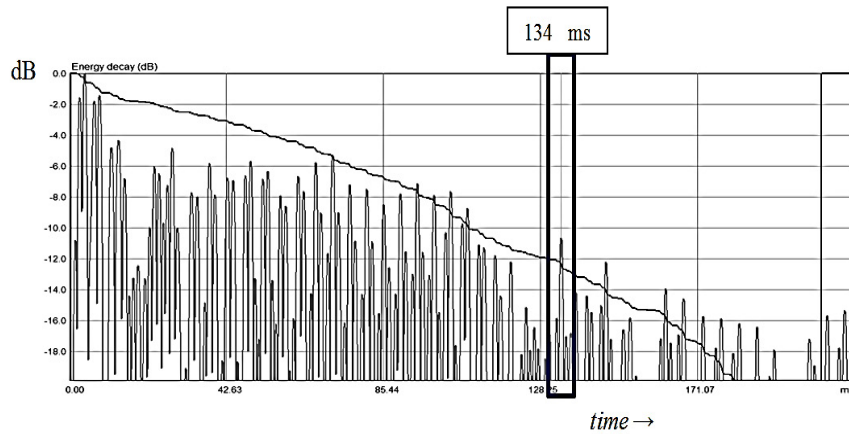
*Figure 4: Acoustical Energy Decay, 1/1 oct. 250 Hzfrom short voice utter-ing followed by "almost silence" Also shown is the corresponding Schroed-er curve. "Baby Let´s Play House"*

## RCA recordings

### RCA Nashville, 'Heartbreak Hotel'

From figure 5, upper pane, we see that the fundamental pitch is about 330 Hz (the note E, a fourth below the Chamber Tone A=440 Hz). In the lower pane, we see a "reflection" at 82 ms for *Heartbreak Hotel*, which is shorter than the measured tape delay from Sun Studios. If we assume that the infor-mation about the 'pseudo echo chamber' mentioned in Part 1 is correct, this measured delay time could indicate a reflection path in the hallway of 343 x 0.082 = 28 meters, assuming a speed of sound of 343 m/s. (This might ap-pear to be somewhat long for a common hallway, but the actual set-up in the hallway is not known. It could, theoretically, also be that both the microphone and the loudspeaker were positioned at the same end of a hallway with half of this length). The perceived reverberation for this recording pro-vides less impression of
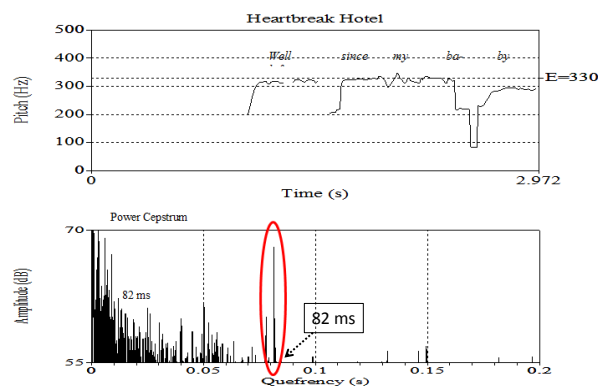


*Figure 5: Pitch and Power Cepstrum (Praat), for 'Heartbreak Hotel'.*

141

a distinct echo than the slap back echo from Sun Studios. This is probably due to many diffuse reflections also from the side walls of the hallway that arrive earlier than the main 'echo' from the end of the hallway.
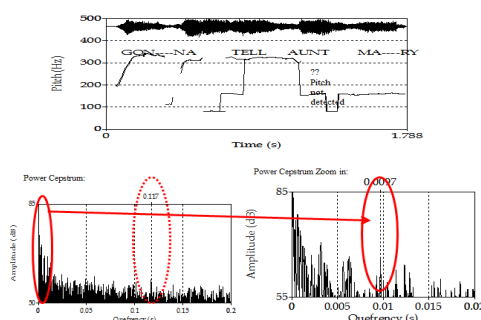
## RCA Hollywood, "Long Tall Sally"



*Figure 6: Pitch (and wave form), Power Cepstrum, and Zoom-in of Power Cepstrum for "Long Tall Sally".*

From the middle pane of figure 6 we find that for *Long Tall Sally*, there is no strong indication of a distinct echo, but we see a weaker reflection at 117 ms. From the zoom-in of the Power Cepstrum in the lower pane, we could speculate if there is a short reflection at 9.7 ms. Such a short delay could have been due to a floor reflection, but, as discussed in the Appendix, this is interesting, but somewhat questionable.

## Analysis Method: Comb filters; Cepstrum; Autocorrelation

### Comb Filters

General information and theory on comb filtering due to an added, delayed reflection (echo, delay) is given in Halmrast (1999 and 2000). A short summary is that when a delayed reflection with the same strength is added to a signal, some frequencies will sum up in-phase, giving + 6 dB. Exactly in-between these frequencies, the delayed reflection and the original signal will be out of phase, and there will be sharp cancellation to (theoretically) minus infinity dB. Using a linear frequency scale, the resulting spectrum looks like a comb, and is called a comb filter. In Halmrast (1999 and 2000) CBTB (Comb-Between-Teeth-Bandwidth) is defined as the distance (in Hz) between the succeeding teeth (and also between the succeeding peaks, but they are often not as easily detected). A short reflection provides very broad combs, with very large CBTB, and could be looked upon as a simple bass boost. A very long delay for the reflection/echo provides very small CBTB, so the teeth in the comb filter will be so close that we do not perceive the changing in timbre, but the reflection is perceived as a distinct echo in the time domain.

We will start with discussing a click signal with an added a delayed reflection/echo. Figure 7 shows the signal (or what we could regard as an Impulse Response). The delay is chosen to be 137 ms (as for Elvis´s Sun recording in 3.1), and, for simplicity, the reflection is equally strong as the direct sound.
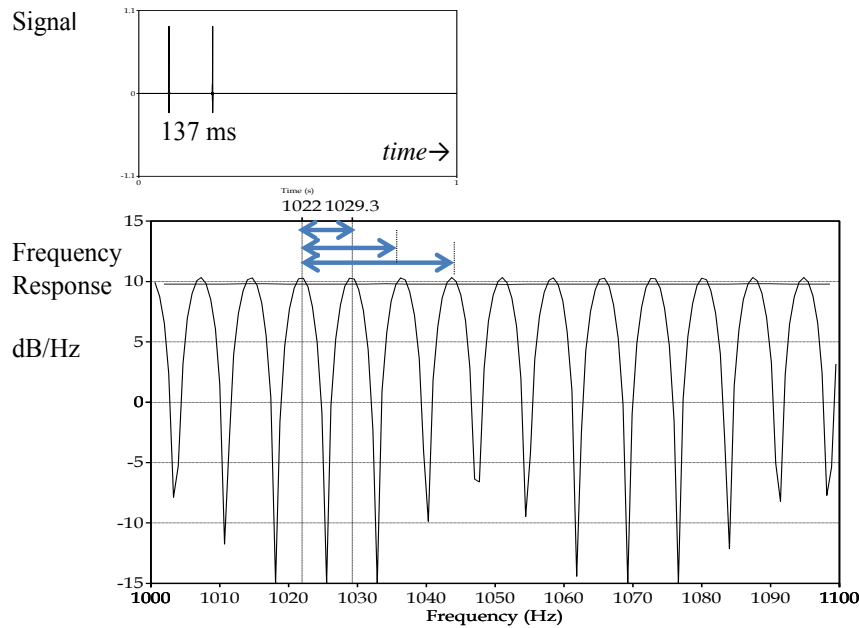


*Figure 7: Upper pane: A pulse (dirac pulse) with a repetition (echo, delay) of 137 ms. Lower pane: Zoom in of the frequency response of the pulse with reflection, showing the comb filter .*

In the lower pane of figure 9 we see the comb filter shape. The figure is zoomed in to frequencies between 1000 Hz and 1100 Hz, but the shape of this curve is repeated through the whole frequency range, from 0 dB. We see that the distance between the dips is CBTB = 129.3-122.0 = 7.3 ms. (The 'distance between two peaks' is also 7.3 ms, but the peaks are not that easily detected). The 'distance' is CBTB = $1/\Delta t$ [Hz], where $\Delta t$ is the difference in time (in seconds) between the delayed echo and the direct sound.

The CBTB is dependent only on the time delay, but the amplitude of the peak and also the depth of the dips depend on how strong the reflection is. For the general understanding of comb filters as a result of adding reflection(s), it could be mentioned that if a comb filter has some kind of repetition/feedback, the peaks gradually becomes sharper and the dips wider in frequency so that the comb shape will be gradually be transformed into upside-down. The feedback could be electronic or natural. Blowing a flute is such a feedback system of repetitive reflections between the blowing point and the flute´s end/tone hole. The fundamental and harmonic partials are

143

shown like regular, sharp peaks in frequency and this shape is actually the comb shape from figure 7 upside-down, (often with gradually falling amplitude).

## Cepstrum

The word Cepstrum comes from reversing the first four letters of 'spectrum', and is chosen because we analyse the 'rhythmic behaviour' of the spectrum, or, the somewhat un-academic: 'spectrum of the spectrum'. As we see from the lower pane of figure 7, the shape of the frequency spectrum repeats itself every 1029.3-1022= 7.3 Hz. This is what we find by cepstrum analysis. Cepstrum is the inverse (Discrete) Fourier transform (IDFT) of the log magnitude spectrum, and a float diagram looks like figure 8.
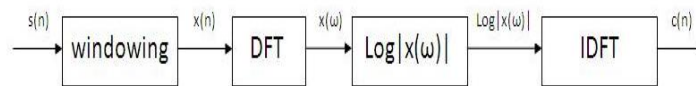


*Figure 8: Float diagram of Cepstrum analysis.*

There are several versions of cepstrum: complex cepstrum, real cepstrum, phase cepstrum and power cepstrum, and the result can be shown as a plot of cepstrum over time, example: Power Cepstrogram. The version we used for our analysis is the Power Cepstrum, which is the squared magnitude of the inverse Fourier transform of the logarithm of the squared magnitude of the Fourier transform of a signal, which can be written as equation 1.

$$Power\ Cepstrum = |\mathcal{F}^{-1}\{\log(|\mathcal{F}\{f(t)\}|^2)\}|^2 \qquad (1)$$

where $\mathcal{F}$ is Fourier the transform (DFT), and $\mathcal{F}^{-1}$ is the inverse Fourier transform (IDFT). The x-axis in cepstrum analysis is called Quefrency. It is shown in seconds, but is a sort of "illusionary" time scale. Power Cepstrum of the pulse with added delayed reflection (echo) at 137 ms from figure 8 is shown in figure 9.

We see the peak at 137 ms, which we know (because we introduced this delay), so this peak provides the time delay correctly for this example. But
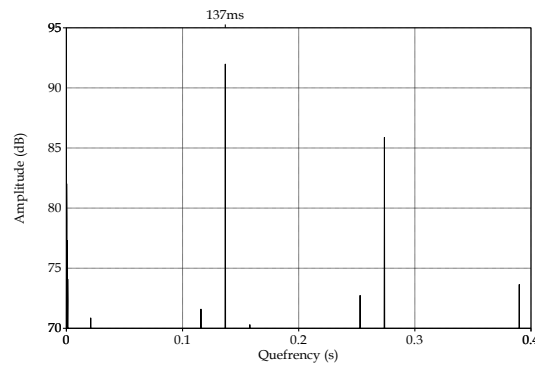


*Figure 9: Power Cepstrum of a pulse with a reflection at 137 ms*

we also see a (smaller) peak at the double value, and we could find smaller and smaller peaks also at triple value etc. This is because we do not really measure 'time', but how often the spectrum (in fig.7) 'repeats itself', and from figure 9 we can see also 'repetitions' between every other or every third etc. peak (or dip) in the spectrum in figure 7.

For this test, with a just pulse as signal, the result from the cepstrum analysis was quite clear, but for a more common (musical) signal, the quefrency value found by cepstrum analysis provides information about every kind of 'repeating behaviour' in the spectrum of the signal, not only comb filters due to echoes etc. Another type of signal that provides 'repetitions' in the spectrum is a spoken or sung voice that provides the fundamental pitch and the overtones as multiples of the fundamental. This will also appear in the cepstrum analysis of the voice, but luckily, in a somewhat separate region along the quefrency axis, as pointed out in figure 1. Figure 10 shows the Power Cepstrogram (The evolution of Power Cepstrum over time) for Elvis´ Sun recording, zoomed in to the range where one can use Cepstrum for analysis of pitch (in order to provide results like in fig. 2).

Room resonances, formants and flutter echo will also appear in the cepstrum. Figure 11 shows a typical overview, with a "zoom-in" in the lower pane.

We have marked the lower limit for echo as 50 ms, which is a typical limit for echo for speech given in the literature (possibly mostly for western lan-
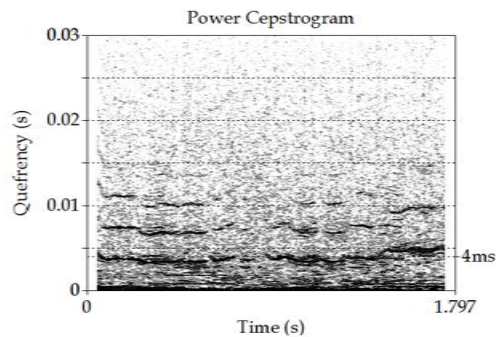


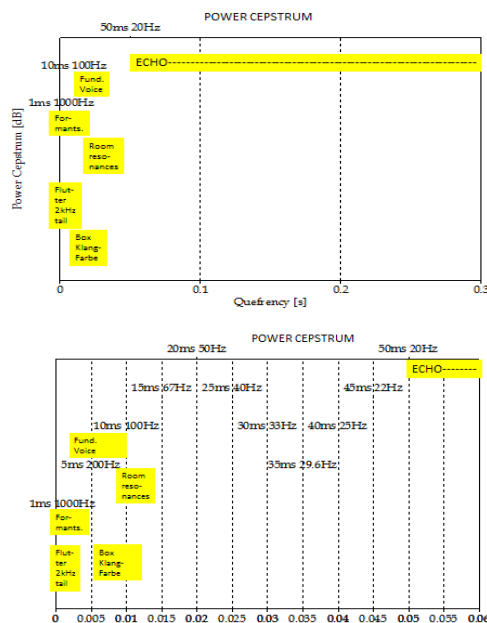*Figure 10: Power Cepstrogram for "Tryin´ To Get To You"*



*Figure 11: Upper pane: Overview of typical results from cepstrum analysis; Lower pane: zoom in on "early part".*

guages). The limit for detection of distinct echo depends on type of signal, and clicks or transient instruments (marimba etc.) have shorter limits. (see Halmrast 2000). As an example of how the human ear works, we can notice that a repetition of 50 ms corresponds to a frequency of 1/0.050=20 Hz, which is known to be the lower limit for humans to hear sounds as "tones". That means that quefrencies shorter than 0.05 is (might be) perceived as pitch, and that the smaller the quefrency is, the higher the pitch.

## Autocorrelation

Correlation between two signals provides information about "how alike" they are (cross-correlation). Autocorrelation shows to what degree at signal 'looks like itself' after different times. Both Autocorrelation and Cepstrum can be used for analysing pitch of complex, harmonic signals, and both also provides information regarding echoes. The differences in analysing methods and results are not discussed further here. Technically, Autocorrelation provides values also for negative time, but for sound signals, negative time is, of course, skipped.

## Timbre/Comb Filter coloration

Investigations in rooms show that if a distinct, separate reflection provides a comb filter with a CBTB in the order of critical bandwidth along the basilar membrane in the cochlea, the added reflection is perceived as if the source was placed in a small 'boxy' room, and such influence of the timbre is called 'Bok-Klangfarbe'-coloration in Halmrast (1999 and 2000). This is indicated in figure 14 as Box-Klang-Zone.

On the x-axis in figure 12 we have the time delay, Δt in milliseconds, (shown in negative direction for a logarithmic scale along the x-axis) and below that, the corresponding excess pathway for the delayed echo, shown in the drawings below as typical sound pathways from a source, reflected back to sender from surface at different typical distances. The bottom x-axis shows the corresponding centre frequencies of the Comb Between Teeth Band (CBTB = 1/Δt in Hz, also in a logarithmic scale, but since it is the inverse of the time axis, this axis is in the positive direction). The y-axis shows bandwidths in
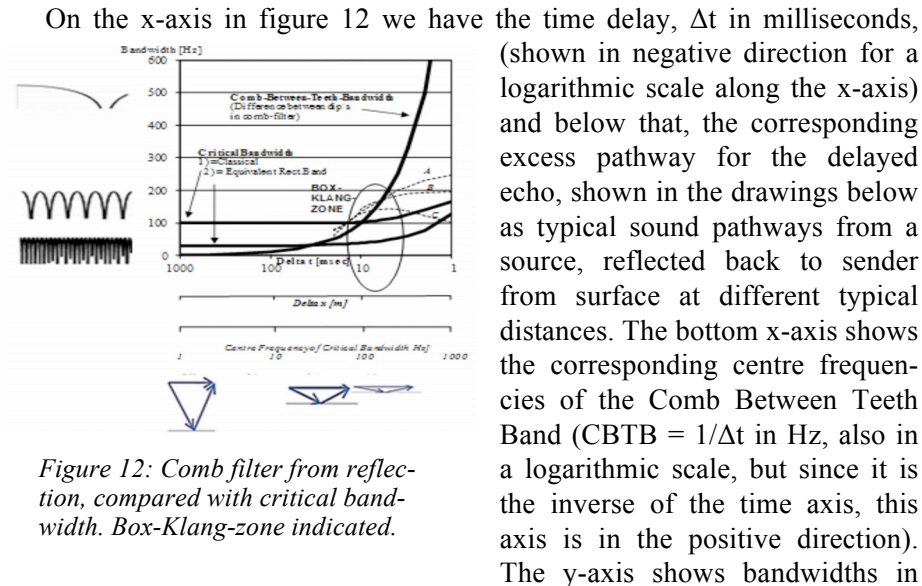


*Figure 12: Comb filter from reflection, compared with critical bandwidth. Box-Klang-zone indicated.*

Hz. Two curves show the Critical Bandwidth along the basilar membrane in the cochlea for different centre frequencies, one curve from measurements in acoustic literature, and the other for the more 'mathematical' ERB; Equivalent Rect. Band. The difference between these data for Critical Bandwidth of our hearing is not very important for our study. Both these two 'psychoacoustic' curves for our Critical Bandwidth are compared with the curve for CBTB (mirroring the lower x-axis, but remember that the y-axis is linear). The drawings to the left of the y-axis shows typical shapes of the comb filter corresponding to the time delays. (Such shapes of comb filters due to reflections with different time delays will be shown more in detail in fig. 13).

In Halmrast (1999 and 2002) it was found that when listening in 'real rooms', a 'boxy' timbre was perceived when a distinct, separate reflection (delay) with a CBTB comparable in size with the Critical Bandwidth is added to a broad banded signal (like a singing voice). This region where the curves cross in figure 8 is called the 'Box-Klang-Zone'. (The curves marked A, B and C are results from a psychoacoustic investigation for noise signals with added reflections, see Salomons (1995), and will not be discussed further).

Do the echoes provide "closeness" due to comb filter coloration? Figure 13 shows the typical, simplified comb filters created by the different time delays measured from the recordings. (Ca. 130 ms for Sun and 80 ms for RCA). The figure shows spectrum (linear frequency) from 0 to 200 Hz, and this pattern is of course repeated all the way up in frequency.

In general, long delays provide very small CBTB, so that the "combs" are not perceived in the frequency domain, and the reflection is perceived as a distinct echo in the time domain. For very short delays, the reflection provides a rise for the bass (as a low pass filter). For a medium delay time (example: the 9.7 ms shown in the lower pane of fig. 13), CBTB is in the order of critical bandwidth. Figure 14 shows the delays measured from the recordings, compared with the results for Box-Klangfarbe Coloration from figure 12, cfr. Halmrast (1999 and 2000).
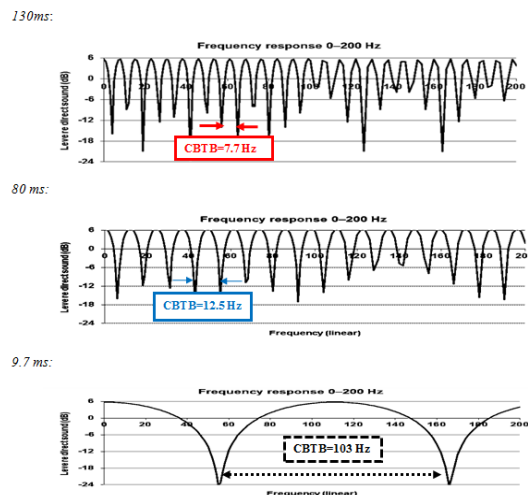


*Figure 13: Typical comb filters for different delay times. Upper: ca. 130 ms (Sun Studios), Middle: 80 ms (RCA, Nashville) Lower: 9ms (as for a floor reflection).*
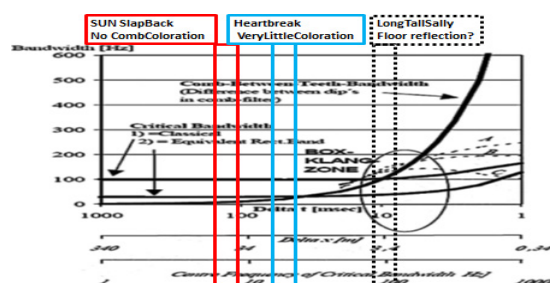
*Figure 16: Analysis of possibility for Box-Klangfarbe. Measured delays for echoes and the corresponding CBTB of their comb filters compared with critical bandwidth*

We see that we must have shorter delay times than measured for Sun Studios, (and also for RCA), in order to get a clear Box-Klangfarbe coloration due to comb filtering. As an example, an imaginary, strong, distinct 'floor reflection' (9.7 ms, shown in the black dotted rectangle in fig. 14) would be in the 'Box-Klang-Zone'). As shown in Appendix, such a floor reflection is interesting, but questionable, and in any case not strong enough to provide 'intimacy', but it is included here, in order to show what kind of short delays would be necessary in order to get Box-Klangfarbe.

## Tape recorders and mono/stereo

### Creative use of tape recorders

The modern, magnetic tape recorder was introduced by BASF/IG Farben in Germany in the 1930´s, and was intended to be just a machine for recording sounds. But creative people soon started to experiment using these machines for musical composition and sound effects. Pierre Schaeffer (1910-1995) made his *Etude aux chemins de Fer* in Radio France 1948, one of a series of short 'studies', which used the effect of playing recorded material at different speeds. Later, at about same the time as Sam Philips´ recordings of Elvis, Karlheinz Stockhausen (1928-2007) experimented in Germany, using tape loops played back at different speeds and also re-recorded in a very reverberant room. However, his most important composition, *Gesang der Jünglinge* (1955–56) was composed after Elvis´ Sun recordings. We cannot suspect that Sam Philips was aware of these European 'contemporary' music experiments. It is more likely that Sam Philips could have heard the creative use of tape recorders in more 'popular' styles of music, like how Les Paul (1915-2009) and others used the tape machine for 'sound on sound', for instance *Lover (When You´re Near Me)* from 1949.

## Mono-Stereo

All these examples of creative use of the tape recorder were in mono. It is true that Alan Dower Blumlein at EMI/Abbey Road patented 'stereo' already in 1931, but at that time it was only used in a few very special cinemas etc. For commercial recordings, stereo was not much used before Abbey Road/EMI 'rediscovered' it in 1956, so mono was the 'only choice' at the time of the early Elvis' recordings.

## Adding more comb filters

The slap back echo is often described as giving Elvis more 'power' (as for a 'King'), but on the other side, the echo still provides some kind of 'intimacy', as being inside a room. As long as a mono signal is played back thru just one loudspeaker, the only dimension possible to increase is the depth, by adding an echo 'behind' the signal/singer. Such an echo arriving from the same angle as the direct sound is not very common in real life, other than in fact; in a corridor, behind the talker/singer, just like RCA tried. Such echoes are not much discussed in the literature, but a corridor could, at same time, be considered to be both small/intimate and long/big.

The perception of a comb filter (a delay) depends mostly on the time delay: Short reflections give coloration in the frequency domain and long reflections give echo in the time domain. The strength of the comb filter effect depends on the relative strength of the reflection compared to the direct sound. Comb filtering is the result of 'mixing' a signal with a reflection. A comb filtering occurs no matter where this 'mixing' appears: a) on the site of the recording, b) in the mixing console/tape machine or c) in the listening room. This means that a listener might hear several comb filter effects overlapping, depending on the loudspeaker set up etc. When several comb filters overlap, they might: 1) increase the comb filter effect (only in special situations where the delay times are multiples of each other, for instance a flutter echo), or, more common: 2) somewhat cancel each other out (if there are several reflections with more 'randomly' distributed delay times). For the playback of a mono signal with an added reflection (like Elvis´ Sun records) thru one central positioned, single loudspeaker, the comb filtering/echo on the record will of course be the present for all listening positions in the room. If this mono record was played back thru two (or more) loudspeakers, there will be added comb filtering(s) due to the time difference between the arrival of the signal from speaker 1 and speaker 2, (if the listener is not sitting 'dead centre'). This might, in some positions, lead to a decrease of the perception of the slap back echo on the record. Moving the second loudspeakers very far away (137 ms equals 47 m assuming a speed of sound of 343 m/s) and to the side (while keeping the level at the listener constant!), the delay will be perceived more and more like a distinct slap back echo in the time domain,

but arriving from a totally different direction than the direct sound from the first loudspeaker. We can imagine that this set up will not give any kind of intimacy.

In addition to such comb filtering (echoes) created by the listening room/loudspeaker positions, we have some additional effect due to anatomy of our head. When the signal from one loudspeaker in a room reaches the listener, the arrival time will be slightly different at the two ears, typically between 0 ms when the source is right in the front and 0.5 ms directly from the side. We will not go into details about spatial/binaural hearing, but this, and other effects provide what is called *'binaural decorrelation'* (see Salomons 1995).

The conclusion is that the perceived effect of comb filter coloration and probably also slap back echo is somewhat reduced when listening to a mono signal played back through several loudspeakers and especially for the situation when the set up resembles stereo.

## *'Artificial stereo'*

After stereo had been introduced, some companies wanted to 'renew' their old mono recordings. Such methods of creating 'artificial stereo' were of course introduced later than the Slap Back Echo recordings by Sam Philips, but the method is included here in order to show that the perceived effect of a delayed reflection panned in stereo is highly dependent on the time delay. Very short delay comb filtering, typically <15 ms, was a common method for *'Mono Electronically Remixed to Stereo'* or *'Duphonics'* introduced by Capitol Records in1961:
   - splitting the mono signal into two channels,
   - delaying the left and the right signals
   - desynchronizing the two channels by fractions of a second,
   - cutting the bass frequencies in one channel with a high-pass filter,
   - cutting the treble frequencies in the other channel with a low-pass filter.
   - In some cases, the effect was enhanced with reverberation and other technical tricks".[7]

This explanation about the filtering is probably not quite correct. Based on what we now know from the theory about comb filters, the resulting comb filtering is somewhat more complex than the discussed high-pass/low-pass filtering a common method was to mix a short delayed 'reflection' with the direct sound in-phase for left channel and out-of-phase for the right channel. The result will be that the comb filter for the out-of-phase channel will be shifted ½ CBTB. This is probably the method introduced by RCA as Elec-

---

[7] https://en.wikipedia.org/wiki/Duophonic

tronic Stereo Reprocessing System (ESP) in 1961, and used for the processed stereo versions of Elvis first albums. They sound quite weird, and almost all 'intimacy' is lost. For some tracks there is almost a high/low filtering of Left/Right, as indicated for Capitol above, but that is probably due to setting the time delay of the comb filter too low, so that the first teeth in the comb filters moves towards higher frequencies, and the result is almost like a low pass filter in the left channel and a high pass filter in the right, but in the high treble, the next peaks of the comb are audible also in the left channel. The main idea is that comb filters are the result both for artificial stereo and Slap Back Echo in mono. The main difference regarding the comb filtering is that the time delay is much longer for Slap Back Echo.

## Research on echoes

Echoes in rooms have been investigated by numerous researchers, especially in order to avoid them in concert halls. The most common echo criterion was introduced by Dietsch (1986) but this does not give full answer regarding the perception of an echo, and includes very little discussion about the angle of arrival of the echo. A strong echo with a delay time as long as 134-137 ms is clearly perceived as a distinct echo, so the question about the possible difference in perception for different angles of arrival is not discussed in common literature. For the design of concert halls, Barron (2010) states the importance of lateral (sideways) reflections in order to give 'envelopment' for acoustic music. The results are mainly given for lateral reflection arriving from sidewalls and thus from a different angle than the direct sound, but a delay time of 134-137 ms is way out of the region of time delay discussed for beneficial lateral reflections in concert halls.

## Simulation of slap back echo panned in 'real' stereo

All the recordings of Elvis in Sun Studios were in mono. To be absolutely sure that no *'Electronically Re-channeled for stereo'*-like effect had been added later, this was checked by comparing the autocorrelation of one channel (Left) and the cross correlation of both channels (Left and Right), see figure 15 as an example (for *Heartbreak Hotel*), and we see that the curves are identical.
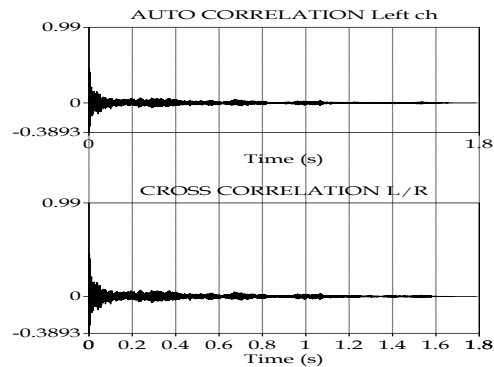


*Figure 15: Check that the recordings are mono. (Heartbreak Hotel) Upper: Autocorrelation Left ch. Lower: Cross Correlation Left/Right.*

As there is not much in literature about the difference in perception of long echoes in mono/stereo it was necessary to perform our own simulation test. We gave a short extract of Elvis´s voice (without any clear slap back echo) a single, distinct echo hard panned in stereo. From listening tests and the audience´s reaction when these files were played at the ARP conference in 2017, it is clear that a simulated time delay of 134-137 ms in the other channel makes the whole 'room' disappear, sounding like a distinct open air echo from one single building positioned to the side of the listener.

Summarised we find that adding a late delayed reflection was a most convenient way to "fatten" the sound around 1955, and that Sam Philips Slap Back Echo only works in mono.

## Conclusions

Our measurements of **Slap Back Echo** from Sun Studios *(Baby Let´s Play House*, *Tryin' to Get to You*) show that this is a single tape echo in mono with a time delay of 134-137 milliseconds. The delay for the RCA, Nashville recording (*Heartbreak Hotel)* is less distinct and shorter, app. 82 ms, which might correspond to a reflection path in a hallway. In RCA Hollywood there is a reflection after 111 ms, but this is even more masked by some general reverb.

Because the slap back echo from Sun Studios is in mono, the delayed reflection is perceived as somewhat 'intimate', even if the delay is so long that they do not provide comb filter coloration (Box-Klangfarbe) in the sense defined in Halmrast (1999 and 2000). If a slap back echo was panned in stereo, the intimate feeling of being in a room would have disappeared.

## References

Dietsch, L. (1983): *"Objektive raumakustische Kriterien zur Erfassung von Echostörungen und Lautstärke bei Musik und Sprachdarbietungen"*, Dissertation, Technische Universität Dresden

Barron, M. (2010): *"Auditorium Acoustics and Architectural Design"* Spoon Press, London

Halmrast, T. (1999): Musical Timbre; Combfilter-Coloration from Reflections, *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)* http://www.dafx.de/

Halmrast, T. (2000): Orchestral Timbre: Comb-Filter Coloration from Reflections. *Journal of Sound and Vibration* (2000) 232(1), 53}69. doi:10.1006/jsvi.1999.2700, (available online at http://www.idealibrary.com and www.tor.halmrast.no)

Salomons, A.M (1995) "Coloration and Binaural Decorrelation of Sound due to Reflections" Dr. thesis Technical University, Delft, Netherlands

http://www.history-of-rock.com/sam_phillips_sun_records.htm (last visited 20180312)

https://stevenrtrent.com/omblog/blog/what-is-slapback-echo (last visited 20180312)

https://stevenrtrent.com/omblog/blog/what-is-slapback-echo (last visited 20180312)

http://www.history-of-rock.com/sam_phillips_sun_records.htm (last visited 20180312)

http://www.history-of-rock.com/sam_phillips_sun_records.htm (last visited 20180312)

https://en.wikipedia.org/wiki/Duophonic  (last visited 20180312)
https://en.wikipedia.org/wiki/Les_Paul (last visited 20180312)

*Discography*

Presley, E. (1955) *Tryin´ to Get to You* and *Baby Let´s Play House* Sun Studios, Memphis
Presley, E. (1956) *Heartbreak Hotel,* RCA Victor´s studio,  McGavock St., Nashville
Presley, E. (1956) *Long, Tall Sally,* RCA studio, Hollywood
    For recording dates, see: http://www.keithflynn.com/recording-sessions/50_index.html
    All recordings taken from The Complete Elvis Presley Masters 30 CD set

*APPENDIX*

## IS THE SPIKE AT 9.7 MS IN FIGURE 7 A WEAK FLOOR REFLECTION?

We have found that the slap back echoes have a too long decay time to pro-vide comb filter coloration (Box-Klangfarbe). In the search of any shorter delays in the cepstrum curves: Could the small spike at 9.7 ms in figure 6 be a floor reflection?  If so, the distance from the singer´s mouth down to the floor and up to the microphone must have been 343 x 0.0097 = 3.33 m. (as-suming a speed of sound of 343 m/s). This is (theoretically) possible, as Elv-is was 1.82 m tall, so that his mouth (and the microphone) could be posi-tioned 1.67 m above the floor. Such a delay is indicated with a dotted black rectangle in figure 9. A 9.7 ms reflection might of course also have arrived from the ceiling or a wall positioned 1.67 m from the singer´s mouth/microphone, but as mentioned in Part 1, this is unlikely in the actual rooms, as the walls are absorbent in the photos.
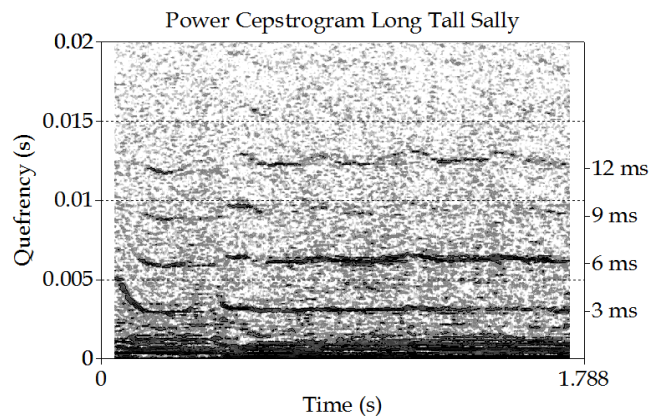


*Figure 16: Power Cepstrogram of "Long Tall Sally". Que-frencies corresponding to the fundamental pitch indicated as the lowest number to the right.*

If we investigate this further, we see that such a short delay time is almost in the order of the period of the fundamental frequencies of the male voice and we need to distinguish between the 'voice' and the 'reflection' in the Cepstrum analysis. From figure 16 we see that a mean value for the fundamental pitch in this short clip of *Long Tall Sally* is around 330 Hz, which corresponds to a period of 1/330 = 3 ms. This is also detected in the Power Cepstrogram (Power Cepstrum over time) shown in figure 16. The lowest number on the right hand side of figure 16 indicates the same result, a fundamental pitch of 1/0.003 = 330 Hz.

*Figure 17: Autocorrelation of Linear Predictive Coding of "Long Tall Sally" with white noise as excitation.*

The fact that we do not see an additional, strong spike at 9.7 ms is a first indication that there might be no strong, distinct floor reflection with a delay of 9.7 ms. (also because 9.7 is close to 3 x 3 = 9 from the voice). In order to try to eliminate both the fundamental and overtones of the voice, Linear Predictive Coding (LPC) was performed; exchanging Elvis´ voice chords with white noise as excitation. The result sounds like Elvis whispering *"Gon´na Tell Aunt Mary"*. Power Cepstrum of this sound file (unfortunately) gave no clear indication of a strong, distinct reflection after 9.7 ms because the LPC transform smooth the attacks (and thus the echo). Therefore we instead performed autocorrelation of this is sound file. From figure 17 we might see that there is a somewhat higher value for app. 9.7 ms, but this is not very strong and also 'smeared out' in time. The result is by no way clear, but indicates that if this was a floor reflection, the singer must have moved his head somewhat during the recording, so that the distance from the floor changed over time,….a highly reasonable assumption for Elvis!