24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ACOUSTICS IN BETWEEN: PERCEPTION OF SOUND IN ROOMS BEYOND STANDARD CRITERIA

Tor Halmrast

Statsbygg, Oslo, Norway and University of Oslo
email: th@statsbygg.no

Most room acoustic parameters are Energy/Time based: "How much sound energy is received during a certain period of time?" These established criteria are founded on a huge amount of knowledge, but fulfilling every criterion will not necessarily secure an excellent concert hall. The Energy/Time approach might lead to strange trains of thought like: "Anything happening before 50 ms is good for the clarity of speech." With modern light weight measuring equipment and "apps," we should take a closer look into cognitive and psychoacoustic aspects of sound in rooms and spaces: "Acoustics Between Times" (how the sound events are distributed within the different time limits) and "Acoustics Between Criteria" (how the standardised criteria might "mask" important information on the perception of room acoustics by putting too much weight on reproducibility, mean values, and standard deviations; demands that might lead to somewhat uninteresting measurements). In this article, we will look into how the reflections are distributed in time; how an astute listener or musician can use handclaps and other personal, impulsive sounds to investigate important details of room acoustics, how a slow attack gives reduced perceived high frequency; and what we can learn from the real experts of "seeing a room with the ears," namely persons who are blind.

Keywords: room acoustics, non-standardized measurements, comb filter, echolocation, attack

1. Introduction

This article is based on the author´s paper Acoustics in Between: Perception of Sound in Rooms Beyond Standard Criteria" in *Psychomusicology: Music, Mind, and Brain*, special edition in honour of Leo Beranek 100 years [1].

2. Between Time

When a signal is combined with a reflection of itself, the result depends on the excess distance the reflection travels: the time delay. If a single reflecting surface is far away, we perceive the reflection as a distinct echo. When the time delay is shorter than some 50 ms, we might not perceive the reflection as an echo for speech, but: the frequency content, (timbre/"klangfarbe") will change. A short reflection gives a very broad comb filter with large spacing between the "teeth" of the comb, (a large Comb-Between-Teeth Bandwidth, CBTB). A long delay between the direct sound and the reflected sound gives very narrow spacing between comb teeth, no noticeable coloration, and the reflection is perceived as a distinct echo in the time domain. The most interesting is somewhere in between: If the extra sound path for the reflected sound is 3.43 m, this gives a delay of 10 ms and a CBTB of 1000/10=100 Hz. This could be the reflection coming back to a person standing 1.715 m in front of a wall in a medium sized room.

If you were near a broadband sound source such as a waterfall and had a group of really fast, strong helpers who could move a big, sound-reflecting wall abruptly in a calculated, rhythmic pattern, you would perceive a musical theme. The melody would follow the pitches f = 1/Δt, where Δt is the time delay between the arrival of the direct sound and the reflection from the wall.

Different halls with the same Reverberation Time (RT) might sound very differently (even if the frequency distributions of the RTs are the same), because RT gives no information about the number, arrival time and strength of the reflections arriving within this period of time. We should investigate the distribution of reflections within the first part of the sound decay (impulse response) in greater detail. Beranek´s Initial Time Delay Gap (IDTG) is in fact such an investigation, see Beranek [2]. We should also investigate if there is a single, discrete reflection dominating, as this might give coloration (changes in timbre/"klangfarbe"), and how lack of early reflections might give longer attacks, perceived as less brilliance.

3. Between Musicians

Comparisons for the stages in several halls, [3] show that what we have called "Box-Klangfarbe" is perceived when a distinct reflection gives a comb filter with spacing between the dips (CBTB) in the order of the critical bandwidth, indicated as the Box-Klang-Zone in figure 1.

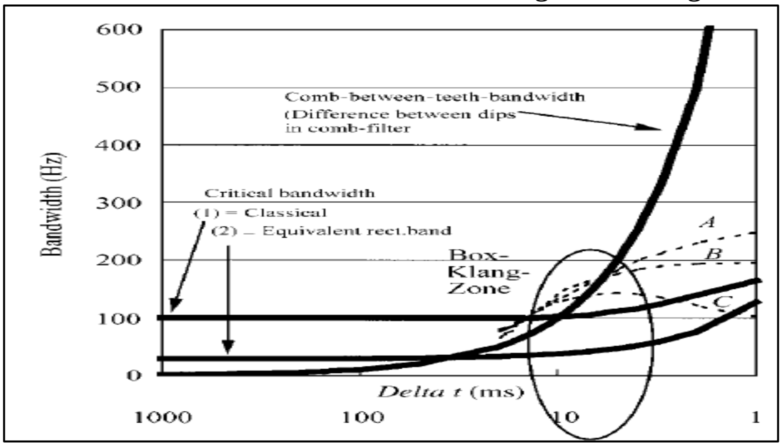

Fig.1. Comb-between-teeth-bandwidth (CBTB) compared to critical bandwidth.

The Box-Klang-Zone indicates that a strong, discrete reflection arriving some 5-20 ms after the direct sound will give a Box-Klangfarbe [3]. Adding more "diffuse"/scattered reflections within the 5-20 ms will reduce the perception of Box-Klangfarbe. The conclusion is that we need to look "between times" in order to find the really interesting details in an impulse response. Our measurements in Musikverein with musicians on stage shows a smooth reflection pattern in the impulse response, and no sign of comb filter coloration compared to several other halls (often with big reflectors over the orchestra, [3]). This is because Musikverein has a high ceiling over the stage and nice balconies, overhangs, and scattering statues and ornaments that supply many early, (not rhythmical) reflections before the arrival of the (rather late and relatively weak) ceiling reflection, 67 ms after the direct sound. Our investigations, [3], also showed that there is limited value in measuring podium acoustics in a concert hall on an empty stage. On an occupied stage, the direct sound through the orchestra is often weaker than the first reflection from a reflector over the stage. Our impulse responses are taken almost diagonally though the orchestra, from the rearmost *violin1* row to *bassoon/double bass* position, with both loudspeaker (somewhat directive) and microphone at height 1.2 m (typical for seated musicians), what might be called TOR (Through Orchestra Response. This and other examples from our measurements show that podium acoustics analysis such as Support/ST and the otherwise highly interesting approach of the Loudspeaker Orchestra (see Lokki [4]) performed on empty stages might be of somewhat limited value when discussing podium acoustics.

4. Between Our Mouth and Ears

The mouth is an excellent, somewhat directive source for acoustic test signals. How do people who are blind “see” rooms with their ears (echolocation)? If we first hear an echo with a 100 ms delay and were told that the reflecting surface is some 17 m away, and then hear an echo with a delay of 200 ms, all of us would likely be able to guess approximately the distance to this new reflecting surface. What if the delay is so short that it does not give any clear, distinct echo, but the direct sound and the reflection more or less smear together, as they do inside dwellings (of reasonable size)? If the sound itself is long, perhaps continuous, one would probably not perceive the reflection in the time domain at all, because its arrival would be masked by the direct/original sound itself. Adding a delayed reflection to long, broadband sounds like running water, constant noise from HVAC systems and walking with hard-soled shoes on gravel, give clear comb filter effects.

For somewhat shorter sounds like speech, the limit for an echo to be disturbing is given in many textbooks on acoustics as 50 ms. That might be a generally good estimation for Western slow speech, but clicking for echolocation uses much shorter sounds. For click sounds, Blauert [5] gives an “echo threshold” of “less than 2 ms”. Our tests indicate that one can hear a difference when adding a reflection with a delay of 1.5 ms to a click. Is this a distinct echo? Or does it sound like just a somewhat longer click? Or is the timbre changed?

Typical types of click used for echolocation with the mouth are short bursts, with a high centre frequency. To an untrained person all clicks sound alike, but fig. 2 shows that they are actually quite different from one “clicker” to another.

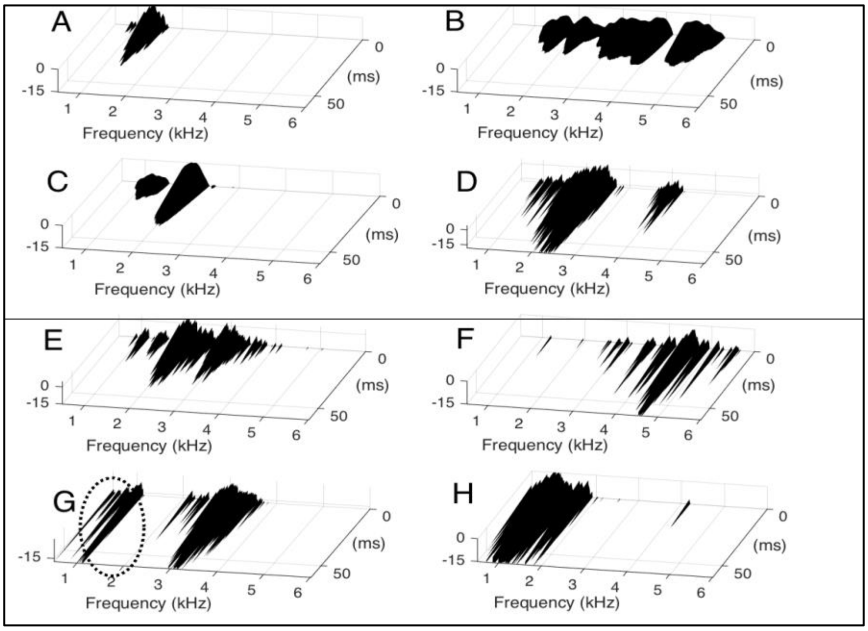

Fig.2. Waterfall-curve of personal clicks for echolocation. A, B, C= Professionals; D,E,F=Novice; G=D.Kish [6] (background noise marked with circle). H= Expert with lower frequency content

One instruction tutorial for echolocation used by the institutions for the blind in Norway is to form your mouth so as to pronounce a “K”, but: “Which K?” meaning “which vowel should we imagine/anticipate when producing the K-sound?” From tables of formants [7], we find that the main formant for the vowel ***“e(eh)”,*** and perhaps ***“i(ee)”*** gives approximately the same frequency content as the clicks of our most experienced performers of echolocation. Do such clicks give perceived comb filter coloration when mixed with a reflection? Figure 3 shows a click with and without a delay of 10 ms. In the frequency analysis we clearly see the comb filter with a CBTB of 1/10ms = 100 Hz. A duration of 10 ms means that the reflected sound has travelled 3.43 m longer than the direct sound, which could mean a wall 1.715 m in front of the clicker. This distance is quite representative for everyday life, and it was chosen for the tests because a 100 Hz CBTB is easy to spot in the zoom-in of the frequency response (using a linear frequency scale).

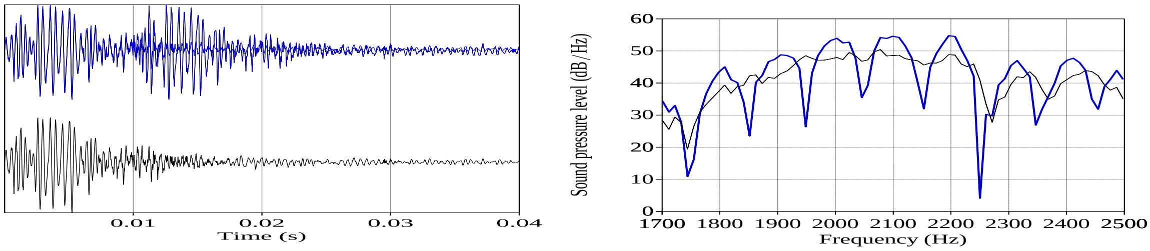

Fig. 3. Click for echolocation with/without 10 ms delayed reflection (1.715 m from a reflecting surface). Comb filter CBTB=100 Hz. Blue (upper): with reflection. Black (lower): without.

From practical tests for echolocation in a normal conference/music room we noticed that people who are blind love corners, because corners always give reflections back to the sender. A plane, long (high) plank is clearly detected when the clicker faces the plank ($\pm$ 5-10^o). If not, a cylinder is more easily detected than a plank, due to diffraction. All objects were more easily detected when the blind person could move the head (and the body) while clicking, and thus make use of the directivity of the mouth and the ears (and the body). We found that persons who are blind often turn their head and direct the click and the ears in order to "zoom-in" on objects and surfaces. This gives that it was harder to judge the presence of a reflecting plate when listening to "in-ear"-recordings of the clicks played back in headphones as a blindfold test, compared to performing the clicking themselves in real time, even for recordings of their own clicks.

Our preliminary tests on echolocation show that for short sounds we use our perception both in time and frequency domain.. The mean frequency of the "pass band" click is well suited in a frequency band where there is not much daily background noise. The knowledge gained from these studies of how blind persons exploit self-produced clicks should be important also for acousticians (and musicians) when listening and clapping/shouting in order to get a sense of the acoustic properties of a performance space.

5. Between Measurements

5.1 Self-perceived reverberation

Standardised demands for the sound source, the measurement uncertainty, repeatability and limits for maximum standard deviations etc. according to ISO 3382-1 are important (for instance in court), but using just these standardised parameters (and not looking into more details in the impulse response) may "mask" interesting detailed acoustic observations. We should perform more measurements, also non-standardised, using balloons, handclaps and our mouth as sources. For ordinary "long distances" between sources and receivers, the accuracy of some of these simplified sources is described in [8]. But when musicians (and low budget acousticians) judge the acoustics of a venue by clapping, shouting or making other kinds of more or less impulse-like sounds, the source and the receiver are (almost) at the same position (as for the echolocator). The sonic experience must be totally different compared to standardised reverberation measurements, but many musicians trust their clapping or Tongue Drops[1] and their ears when they enter a new venue. It seems like we are able to "recalculate", because reverberation times judged "by one's own ear", are often reasonably correct (see [9]). Figure 4 shows a typical impulse response and Schroeder curve of a handclap and a tongue drop in a room, recorded in the ear of the person who made the signal. How can we judge the reverberation time from such a curve, which falls abruptly after the direct sound? We measured several rooms both with claps/tongue drops recorded "in-ear" and with standardised, "long distance" measurements, [9]. We found that after the abrupt fall after the direct

[1] What we call a Tongue Drop" is somewhat lower in frequency than a typical click for echolocation used by persons who are blind, more in the "melodic frequency range".

sound, the decays are parallel. Thus we introduced what might be called "Self-Perceived Reverberation Time", T15*user*, taken between -20 dB and -35 dB of the decay, as shown in figure 4.

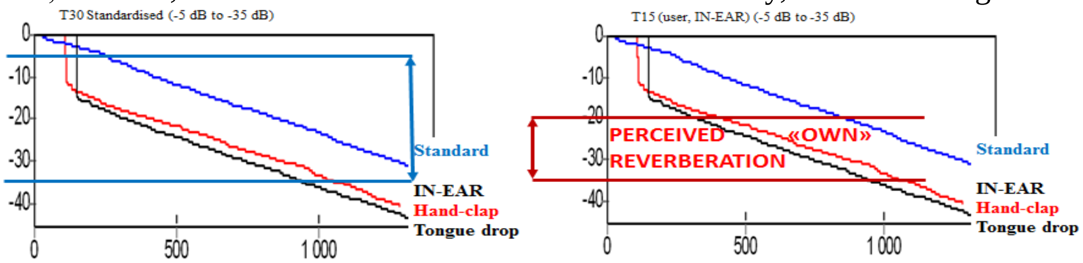

Fig.4. Definition of "Perceived Reverberation," from Schroeder curves.
T15*user* (right panel) compared to standardised T30 (left panel).

Comparisons for several halls/rooms show that the "self-perceived" T15 is quite close to the standardised "long distance" T30 following ISO 3382-1 (see [9]). In addition, measurements of "in ear" Perceived Reverberation are very easily performed in different directions, using the natural directivity of the ears and the body (and the mouth, for the tongue clicks). Directional measurements with more common equipment would require a dummy head, and a standardised directive loudspeaker would need to be invented. Our measurement of Directive Self-Perceived Reverberation on stage in Stavanger Concert House, comparing "in-ear"-recordings of tongue drops in different directions, shows that the self-perceived reverberation time for mid-frequencies is longer towards the audience than towards the rear of the stage (organ), which is as preferred.

5.2 Attack. Perceived timbre

In room acoustics we traditionally put most effort in the decay, analysing reverberation time. Standardised T30 measurements, however, start at -5 dB (cfr. fig. 4), and even EDT does not take the attack time into consideration. For the perceived timbre of music we should also consider attack, see [10]. The following figures show an introductional analysis on how convolution in general "smoothens" the attack.

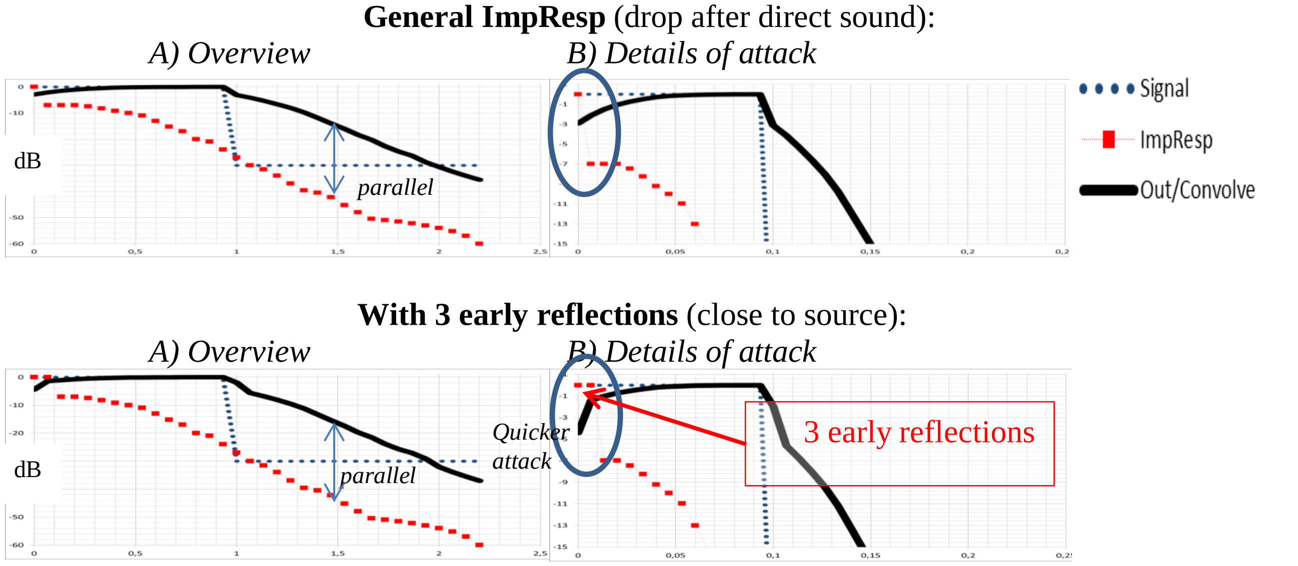

Fig.5. Calculated output of same signal, with impulse response including one (left) or several (right) early reflections. The effect on the attack time is clearly shown

We see that with more reflections close to the source, the attack is faster, even if the decay is the same[2]. We must, however, secure that the early reflections do not arrive "rhythmically", so to

[2] PS! The reason why the direct sound seems to be lower for the B figures, is that the absolute mean is increased, and these figures show normalised output.

produce clear comb filter coloration as discussed in 3.4 (see [3]. Further investigations should include autocorrelation of impulse responses, because a sudden reduction of correlation after the onset (as for allpass-filters used for electronic reverberation), might give extra "smoothing" of the signal attack.

6. Between Echoes

In several halls an echo-like situation is perceived even if there is no distinct single reflection that gives echo after common echo criteria, cfr.[11]. We found this situation in Folketeatret in Oslo, the former site of the Oslo Opera before the new opera building was erected in 2008, see [12].The measured ETC back to the stage is shown in figure 6. If we imagine a march tempo of MM = 120, the echo is a bit shorter than an 8th note.

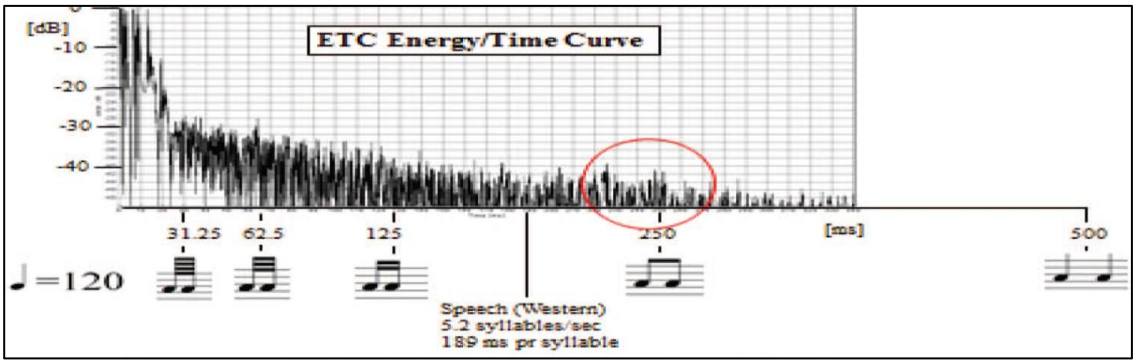

Fig.6. Folketeatret, Oslo. ETC curve back to the stage.
Rhythmic notation of reflections. The cluster at 220–250 ms is perceived as an echo.
The reflections arrive from corners between rear wall and ceiling in the hall.

Several reflections from the rear, upper "corners" behind the audience, and back to the stage are "integrated" and are clearly perceived as an echo when clapping on stage, and of course for a solo marimba. The musicians in the orchestra pit complained, but some singers of romantic opera loved the acoustics, because this reflection back to the stage arrived just before they had finished their phrase. Their own strong and long sounds (longer than an 8th note in MM = 120) masked the echo, and they felt as if a fresh delivery of air really helped them "fill the hall".

In this hall the echo was clearly too dominant, but a gentle "almost an echo" from the balcony fronts or from the rear wall is perceived in many high ranked concert halls. Such delayed reflections from the wall behind the audience might be the only possible support for singers and musicians of "long phrased, romantic music", but very disturbing for impulsive, and/or amplified music.

We can conclude that "An Echo is not an Echo" with clear limits in time and sound pressure level, but highly dependent on the type/length of the signal, and on the amount of masking (both by background noise including background music and by the signal itself). When judging the effect of echo back to the actor/singer/musician him or herself, close inspection/listening to the details in the "in-ear" measured impulse response is useful.

7. Between Walls

Musicians often rehearse in (too) small rooms. This might create problems regarding both sound pressure level and timbre. Acoustic criteria for music rehearsal rooms are often given as reverberation times (or more correctly: decay times). A study [13] showed that resonances in the bass and "shrill" for the higher frequencies might be more important issues. A small rehearsal room was investigated by detailed analysis of recordings of a special test composition, room acoustic measurements/ calculations/Odeon model and a Boundary Element Method (BEM) analysis of room resonances, for different settings of absorbers, see [13]. Detailed analysis beyond common building acoustics measurements indicated that two issues are more important than common sound pressure level and reverberation time criteria: Room resonances in the bass (tuba and trombone), and “Shrill”/”Shimmering” for high pitched instruments (clarinet).

The small rehearsal room was modified from “non-dampened” (just a curtain on one sidewall) to “dampened” (curtain, corner/bass absorber and four wall absorbers). The measured reduction of “in ear” self-perceived reverberation time was greater than the mean reduction measured with standardised methods. To see the actual reduction when playing in the room, we performed detailed analysis of “in-ear”-recordings of a special 60-sec short Test Composition played solo in strict rhythm (MM = 120). The test composition was recorded and analysed both with an omni microphone at a specific distance and angle from the instrument, and “in-ear” microphones. As a mean for these instruments, the measured reduction in sound pressure level at the musician´s ear when introducing the extra absorbers is 1-2 dB less than the 3-5 dB reduction measured with a constant loudspeaker source. This means that even if the sound pressure levels are (too) high in the small rehearsal room, most musicians compensate for the reduced “answer” from the room by playing 1-2 dB stronger. We also found that the clarinet “over-adjusts” to the damped acoustics, and plays even stronger. Her L_{eq} was actually higher in the dampened room than in the more reverberant. Interviews with the musicians indicate that this was in order to reduce the “shrill” for the higher notes, most problematic for the high notes of the clarinet.

Dampening the room was reported as a clear improvement by all the musicians, even though the decrease in sound pressure level is not that large, and the spectrograms of the whole 60-second composition do not show the differences very clearly. We need to study the perceived differences between dampened and non-dampened rooms more in detail, looking into Bass resonance and “Shrill” for high frequencies. All room resonances were calculated, and inspected in the Boundary Element Method (BEM) model. In the “dampened” situation, a flexible bass-/corner absorber was added. This gives a nice reduction of the resonances. Zooming-in on the spectrogram and a “loudness” analysis of the recordings non-dampened/dampened, we found that the note corresponding to the 111 Hz resonance (actually 1 Hz lower, the A2 is 110 Hz) gives a rise in the non-dampened room. The rise for this tone is shown both in the SPL histogram and in the spectrogram. We also found that this note has a somewhat longer duration compared to other pitches in the non-dampened situation.

In order to measure the perceived “shrill” for the high pitched tones of the clarinet in the “non-dampened” situation, we measured the “in-ear” Spectral Centroid. The mean value of the Spectral Centroid for the clarinet was reduced by 200 Hz when damping the rehearsal room with extra wall absorbers. “By ear” this high frequency “shrill”/“shimmering” was more annoying for the clarinet than for the trumpet. This is of course mainly because the clarinet plays this section one octave higher, C6, than, as an example, the trumpet, C5. The unpleasant sharpness around 2-4 kHz[3] is in the frequency region where the human ear is most sensitive and thus important regarding the possibility of hearing loss. The dBA and RT criteria commonly used in legislation for working environment do not detect such important information.

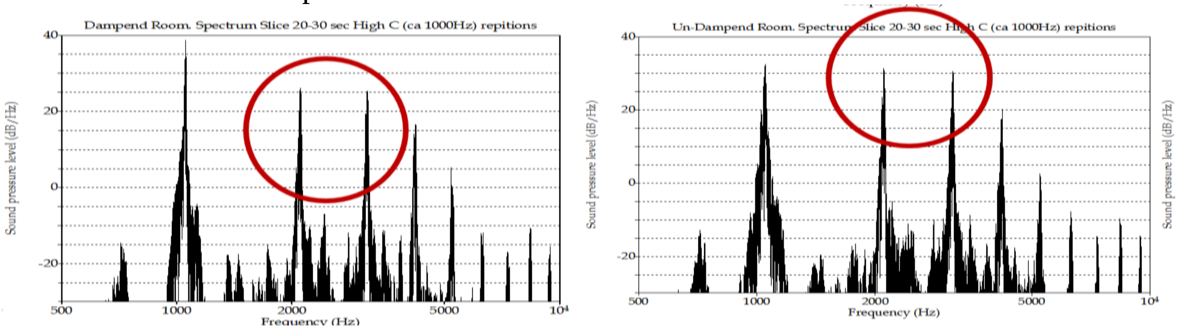

Fig. 7. Freq. analysis of 10 s of clarinet playing high C *natura* (app. 1kHz).
Right: Dampened Room. Left: Non-Dampened. Logarithmic frequency scale.

[3] As shown in fig.7: In high register, the clarinet produces not only the odd harmonics, but all harmonics, and this gives addition to the «shrilling» timbre.

The conclusions from this investigation in a (too) small rehearsal room, is that we need non-standardised measurements to find the many perceived **details that are more important than the standardised RT and sound pressure level criteria.**

8. Between Conclusion and the Future

This paper has presented several examples showing that standard room acoustic parameters do not reveal all interesting elements of the perceived acoustics in rooms. But the rapid growth of inexpensive and handy measuring equipment, such as “in-ear” microphones, handheld wav-recorders, laptops, smart phones and apps, makes it possible to measure impulse responses and spectrograms in many acoustic situations “on the fly” and investigate specific, non-standardised parts of the impulse responses. Inexpensive software makes alternative analysis possible, both regarding timbre and the distribution in time and direction of the room reflections. Such not (yet) standardised analysis will, by nature, not necessary fulfil the accuracy regarding standard deviations etc. earlier believed to be necessary, but offers the possibility for many comparative measurements, so that we can learn more about fascinating details of human auditory cognition. Performing the measurements in practical situations on stage and in the musician’s ear would also decrease the common gap in terminology between musicians and acousticians. The more “in between” we measure, the more efficiently we could learn what additional measurement quantities might be useful. Especially we need to look into how a long attack time is perceived as a reduction of high frequencies, even if the overall spectrum is unchanged, see [10].

REFERENCES

1 Halmrast, T (2015). Acoustics in Between: Perception of Sound in Rooms Beyond Standard Criteria. Psychomusicology: Music, Mind, and Brain © 2015 American Psychological Association 2015, Vol. 25, No. 3, 256–271 0275-3987/15/ http://dx.doi.org/10.1037/pmu0000077

2 Beranek, L. (2004). *Concert halls and opera houses: Music, acoustics and architecture*.2nd ed. New York, NY: Springer.

3 Halmrast, T. Orchestral Timbre: Comb-Filter Coloration From Reflections. Journal of Sound and Vibration (2000) 232(1), 53}69 doi:10.1006/jsvi.1999.2700

4 Lokki, T. Throw away that standard and listen”. *Building Acoustics, 20, No.4.* p.283-293

5 Blauert, J. (1997). *Spatial hearing.* Rev. ed. Cambridge, MA: The MIT Press

6 Kish, D. (2013). World Access for the Blind http://www.worldaccessfortheblind.org/

7 Thienhaus, E., (1934). Neuere Versuche zur Klangfarbe und Lautstärke von Vokalen. [More recent attempts to control volume and timbre of vowels]. Zeitschrift für Technische Physik, 15, 637- 641.

8 Halmrast, T. & Buen, A. (2008): Simplified room acoustic measurements. *Proceedings of the Institute of Acoustics, Vol.30 Pt.3,* .

9 Halmrast, T. (2013). When source is also receiver. *Building Acoustics Vol.20, Number 4 2013* p. 403-422, and updated for BNAM, Tallin. 2014.

10 Halmrast, T. Analysis of Music Performed in Different Acoustic Settings in Stavanger Concert House. ICSV 2017, London

11 Dietsch, L. & Kraak, W.(1986). Ein objektives Kriterium zur Erfassung von Echostörungen bei Musik- und Sprachdarbietungen, (An objective criterion for the measurement of echo disturbances during presentation of music and speech). *Acustica* ***60,*** (3), 205-216

12 Halmrast, T., Buen, A. & Ihlen, G. (2002). The influence of a large reflector over the orchestra pit in an opera house. Institute of Acoustics (UK, London 2002)

13 Halmrast, T. (2014). Musician´s timbre and strength in (too) small rooms. EEA Acoustics in Practice, Issue 3. http://euracoustics.org/AiP/2_1_2014/ updated for EAA Forum Acusticum, Krakow 2014