legends: Orange (Medium Priority), Red (Critical/High Priority), Blue (Info), Green (Good work!)

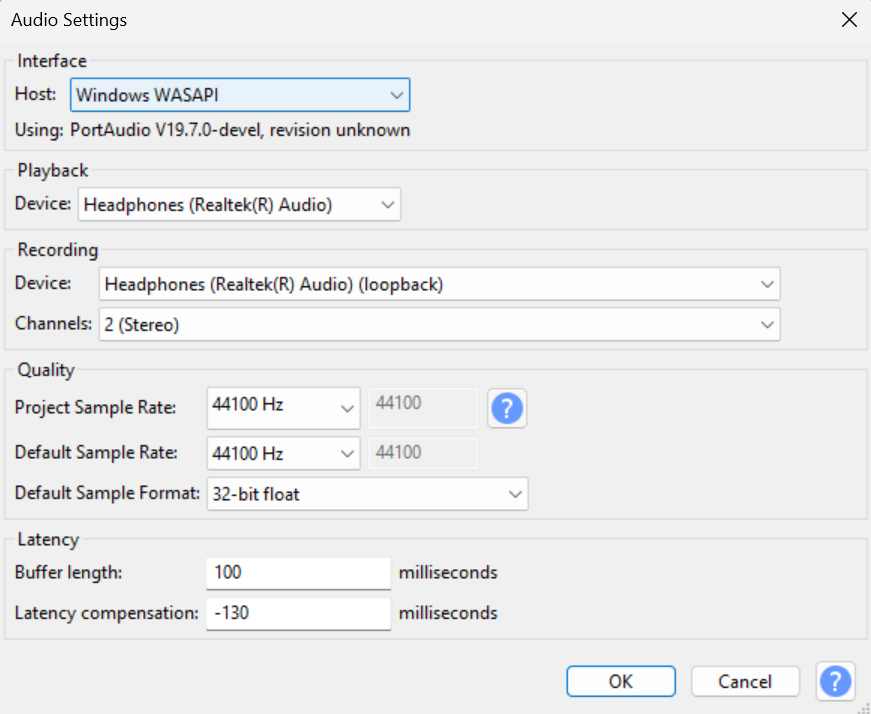
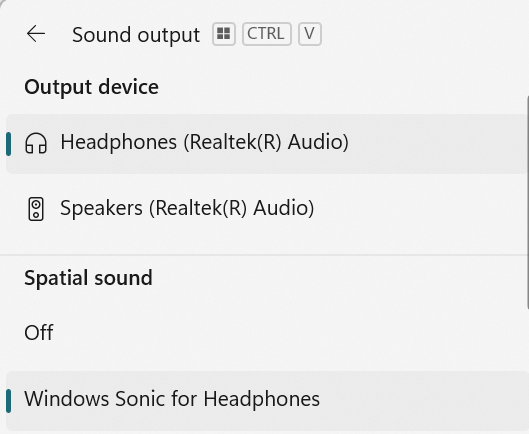
Done, Important

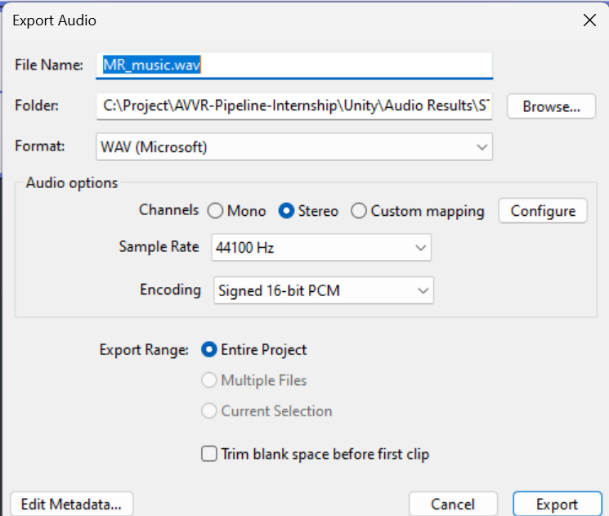
# Monday, 22 July 2024

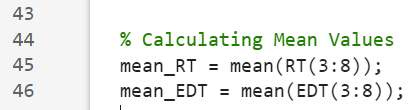
Morning started with Midway Development Day talks etc. Thus, real work started afternoon after lunch lol.

## **UPDATED TODO:**

1. Confirm steam audio setting and standardize/clean project directory to be less convoluted and organised (Mona gave her project as reference)
   1. automate baking with 1 button to streamline
   2. Add lighting indicator for when steam audio is working
2. Generate all audio done
3. Do analysis using provided matlab, have proper report and findings for meeting.
4. Create unity scenes for all scenes and its category (VRDemo, nonVRDemo)
5. Add Rachel and Joao for Demo for more immersive VR demo , and probably have more audio cue. otw
6. Continue monodepth optimisation using enhance360debug.py
7. Remove docker clutter automatically
8. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

For Eval scenes, record at least 30 sec after sound stop to be safe (to make sure get all reverb/reflections). Thus, disable loop and start record earlier (before pressing play, make sure no other sound). Recording using Audacity with following setup on Sony WH1000-XM4, at 50% Windows volume with Spatial Audio, make sure it is stereo. There is concern regarding audio quality and compression of using external software instead of directly through Unity but unfortunately there’s no simple way with Steam Audio compared to Google Resonance. (Need to confirm with Dr. Hansung if this is acceptable)

Audio export with following setting.

Matlab changes on octave band from 2:6 to 3:8 to follow previous paper methodology (as advised by Mona). To analyse the sound, just change directory definition and run the Main file (Main\_Immersive\_EDT\_RT60.m)

To get the RIR file for analysis, run the .ipynb coded by Mona in respective scene folders.

Honestly my brain kinda fried alr today so will continue tomorrow instead.

# Tuesday, 23 July 2024

Interesting to note, one of the most important factor for room reverb properties is its size, and the material properties inside it, imo the furniture shape etc is not as they only serve purpose to make room smaller (less echo) and depending on type of furniture/objects.

Another interesting thing to note, is that sometimes the reverb effect is not applied for first few second when game is play (for example, speech on ST), especially apparent when there supposed to be strong reverb. So using loop and having very long silence at the end might be needed.

Generated all audio!

## **UPDATED TODO:**

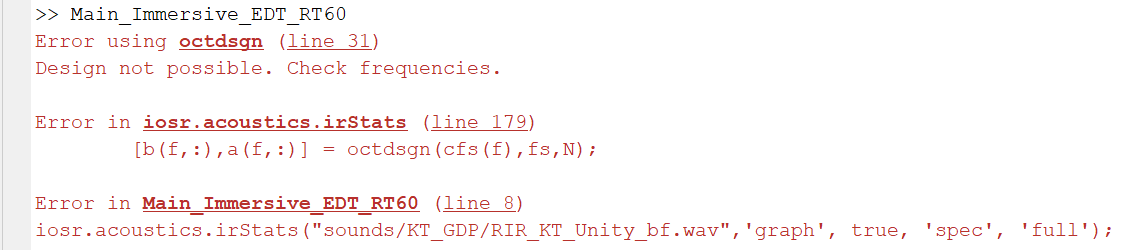
1. Do analysis using provided matlab, have proper report and findings for meeting.
2. Confirm steam audio setting and standardize/clean project directory to be less convoluted and organised (Mona gave her project as reference)
   1. automate baking with 1 button to streamline
   2. Add lighting indicator for when steam audio is working
3. Create unity scenes for all scenes and its category (VRDemo, nonVRDemo)
4. Add Rachel and Joao for Demo for more immersive VR demo , and probably have more audio cue. otw
5. Continue monodepth optimisation using enhance360debug.py
6. Remove docker clutter automatically
7. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

Copied the RIR analysis folder one by one for commit and organisation to remove clutter (unneeded files), and also to make sure the original form Mona is unchanged for reference. It is also easier to have everything related to this internship in one repo instead of two.  
Changed absolute paths to relative paths in Kitchen .ipynb for RIR generation form sine sweep deconvolve using Mona provided .ipynb. Using Python version 3.8.19 after asking Mona (she uses 3.8.9, major version is what matters the most imo so no less compatibility/discrepancy issue).

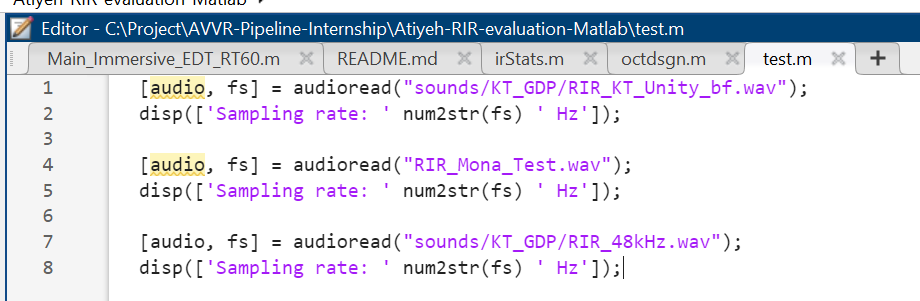
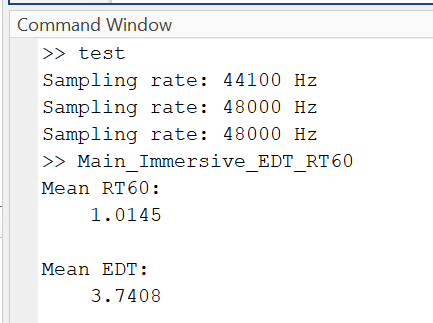
Added Atiyeh-matlab folder for sounds output path. Added Main matlab file for RT60 and EDT analysis. Now modifying it to work with new folder in sounds folder (KT\_GDP etc)

Copied dependencies folder (IoSR Toolbox, RIRs, and octave) from original repo to working repo

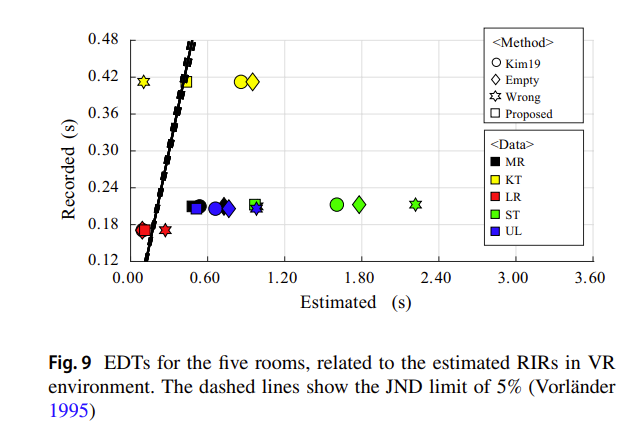
Got this error…

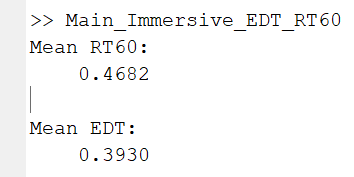


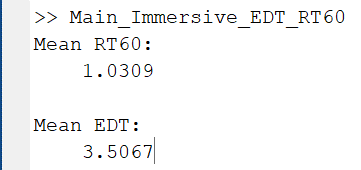
Found out the culprit, it is the sampling rate. It should be 48kHz instead of 44.1kHz. This is confirmed by looking into sampling rate of Mona audio files (cus they work), and after converting mine to 48kHz using Audacity (right click track, rate to 48kHz, then change export setting to 48kHz), it also works

 The problem is whether or not this affect the result in meaningful way, should I just rerecord everything at 48kHz in the first place?

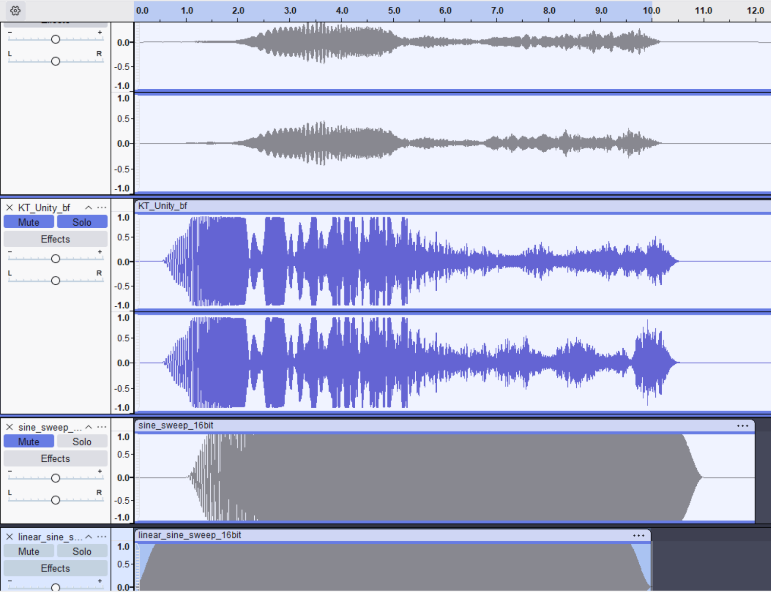
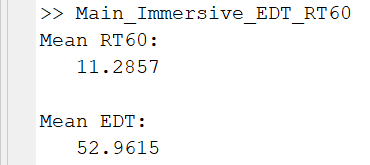
Will need to do some testing on KT first to see if I should rerecord or if its okay to just convert after.

It is most likely not alright, as seen on the right, the EDT is way off…

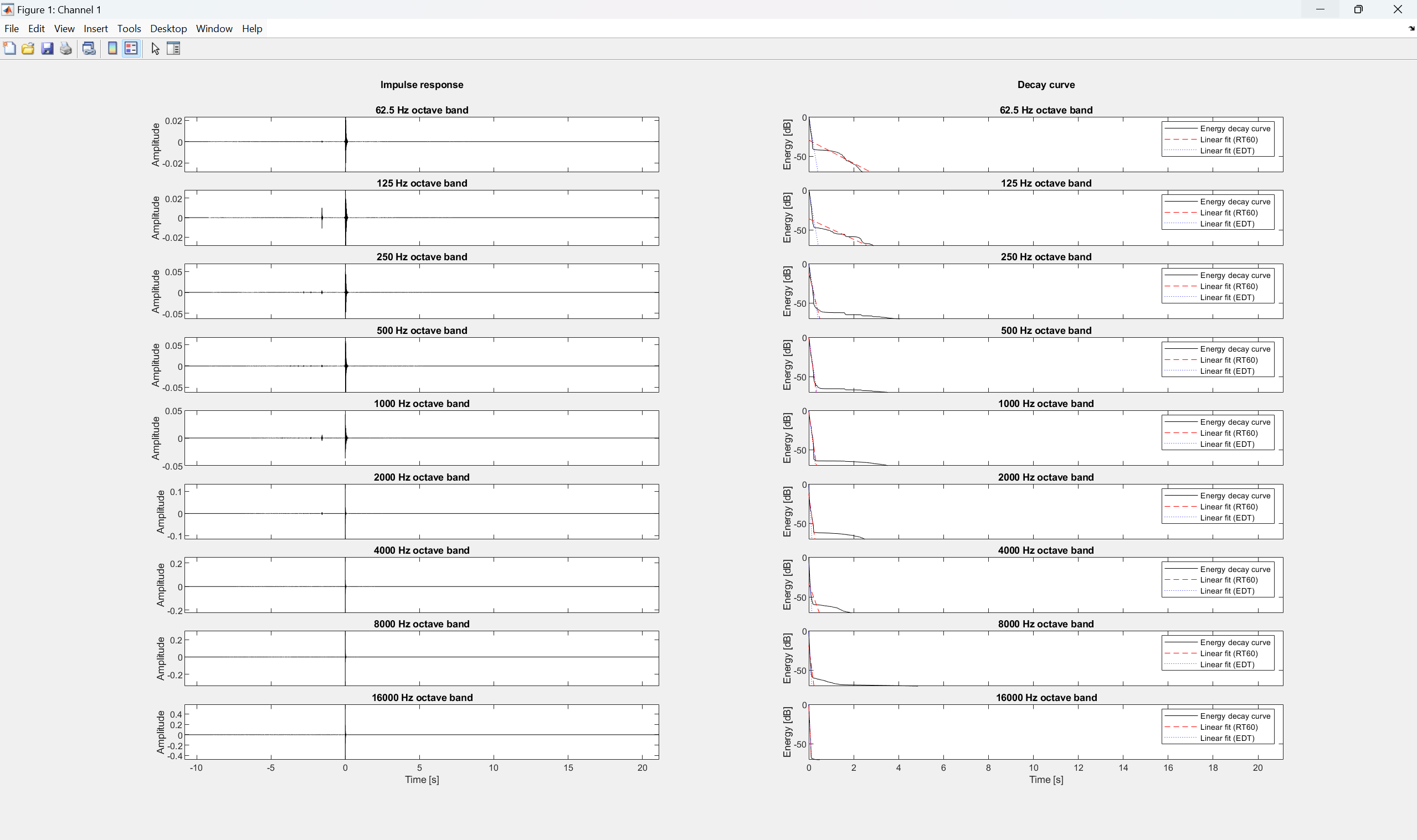
For reference this is EDT and RT60 Mona got for KT…

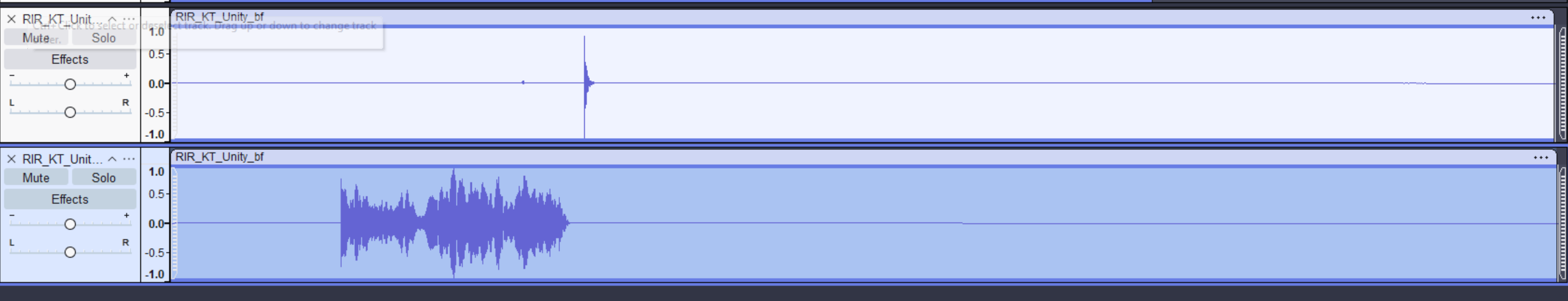
No difference…

After looking back into the waveform of generated audio and input audio etc. I think the problem is that I used wrong sine sweep. Should’ve used linear sine sweep audio file instead, as seen, the second track (Mona) starts earlier, which coincide with linear.

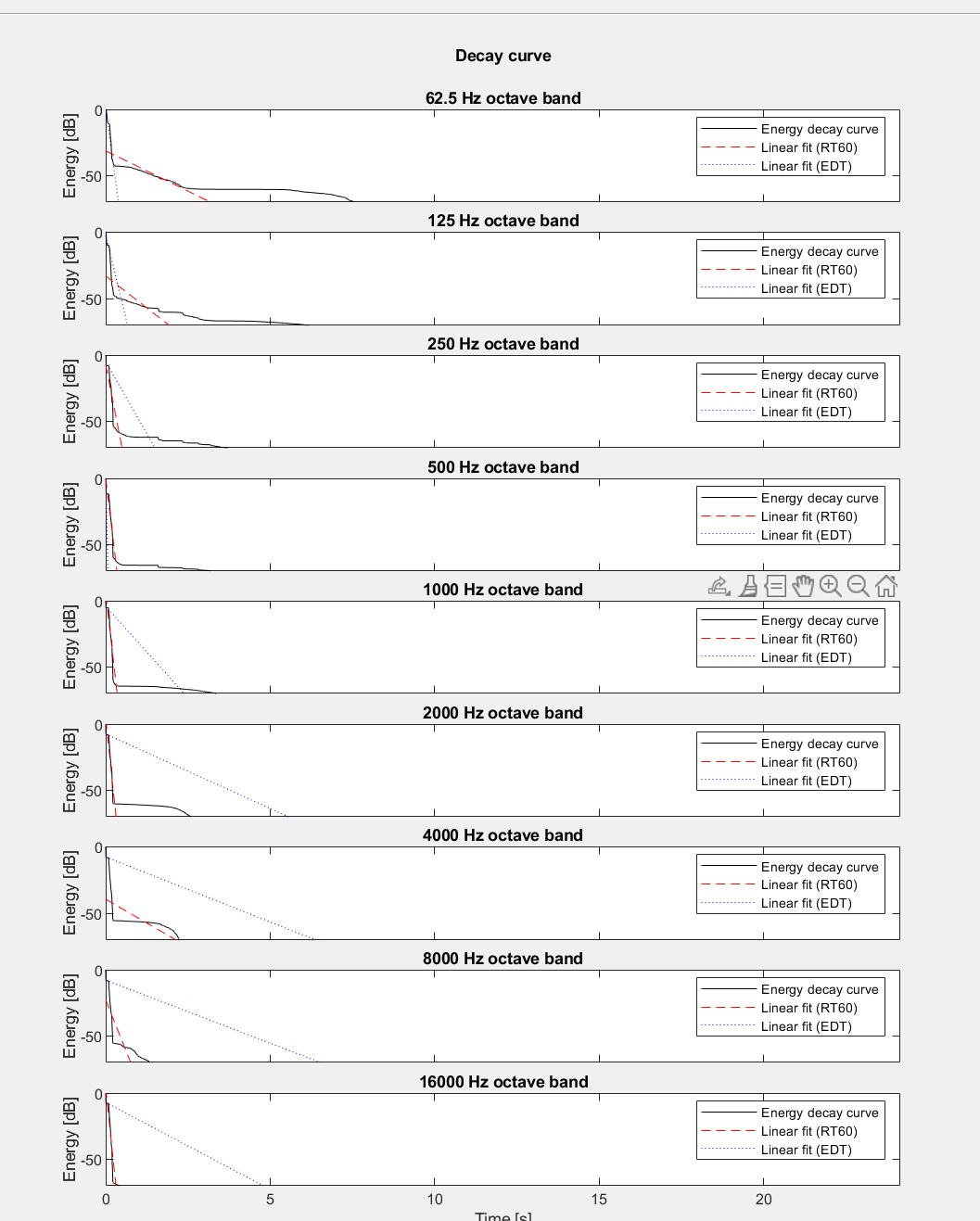
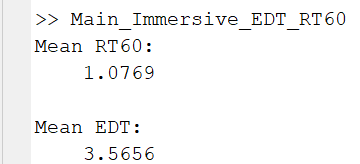
Ok nvm, this give even worse result…

Ok after looking at the generated matlab charts and also RIR on audacity.

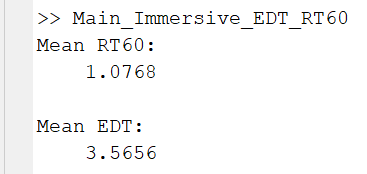




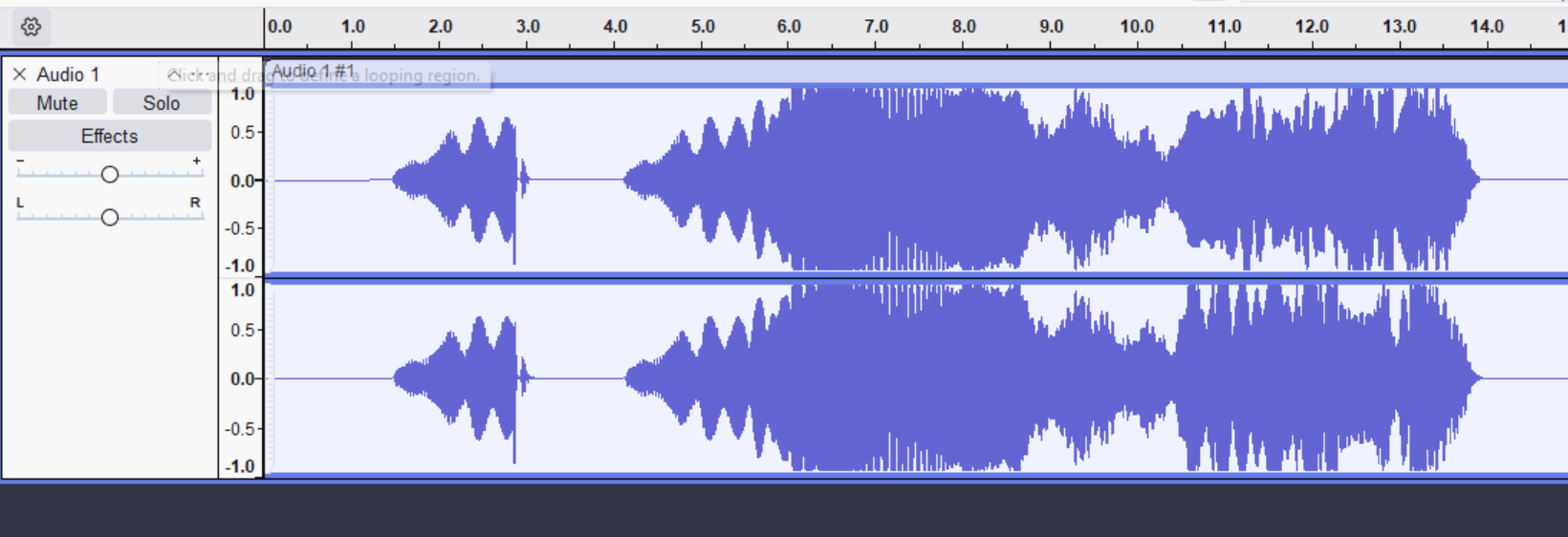
Top is correct (Mona’s), bottom is what I got, its obvious that’s the wrong file as deconvolve obviously not working to get RIR. So linear sweep is wrong.   
Lets try increasing probe number (decrease horizontal spacing in probe batch box to see if accuracy improves)

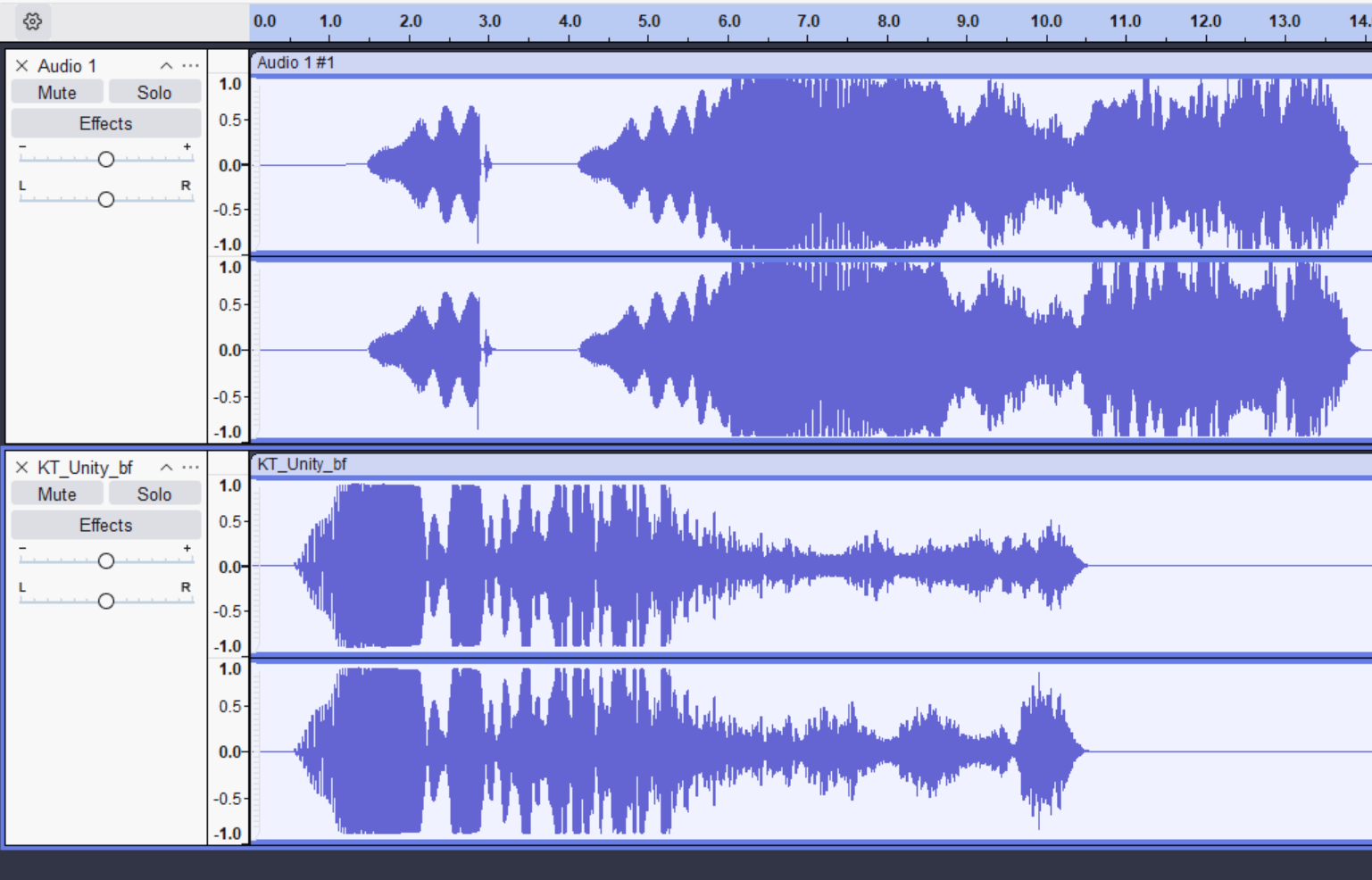
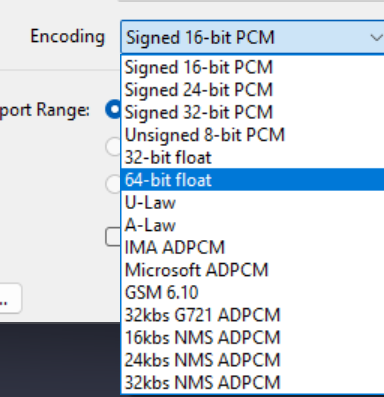
OK after looking into my generated audio EDT graph more, im confident this high EDT value is due to the bug mentioned before (where the reverb doesn’t take effect until few millisecond into the game play mode, which causes very bad reverb effect, in turn bad EDT/RT60 (theory), interesting enough, this only prevalent for higher octave band (frequencies)

Ig the fix is to delay the audio playback a bit by either adding more silence in front (I don’t want to), or not play on awake, instead call it using custom script after 1 sec of awake or something (this can also be used to internally record audio via unity scripting later instead of external software)

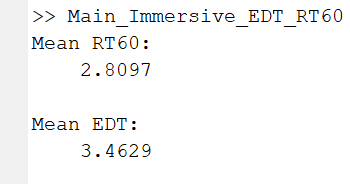
Unfortunately, even after audio delay playback, still got same value, and same EDT graphs as before…

Maybe I need to trim the start? Wavepad have this option easily and that’s what it looks to get Mona’s audio file. Also, maybe I should use higher volume (unlikely the culprit but worth checking)

Ok trimming start did nothing, but interesting enough, changing volume on audio source from 0.13 to 1 causes some frequency to cutoff… 

And when comparing to Mona’s recorded audio, notice how its shorter.. Did she remove the first part??

Another thought, maybe the encoding also matters, it was at signed 16 bit PCM all this time, but I think wavepad default/highest is at 32bit PCM, so im changing to that.

Ok even after removing the early part, and trimming, and at volume 1.0 instead of 0.13, I still get very wrong, in fact it is now worse…

There must be something wrong somewhere but im tired so off for now…

# Wednesday, 24 July 2024

Emailed Mona for advice, she suggested using her settings in Unity instead of S3A.

## Prep for Meeting with Dr Hansung Kim:

### What I did:

1. Continued refining and debugging the pipeline, focusing on depth map enhancement and mesh generation.
2. Set up and debugged Steam Audio components in Unity (from Mona, its better to use baked instead of realtime, also what S3A uses, the only option for Steam Audio atp)
3. Created helper unity scripts for model dimension calculation and positioning.
4. Standardized scene scaling ratio and probe positioning and alignment for 5 evaluation scenes, table and instruction from week 4 for reference.

## Dimension (x, y, z) of scene meshes from original enhance360.py and 360monodepth to calculate scale ratio needed as follows:

1. KT – (4.16, 2.08, 5.44) enter scale (0.83, 1.28, 1.22) , align corner 2,
2. MR – (4.64, 1.84, 5.28), first rotate -90, enter scale (1.06, 1.27, 0.92), align corner 3
3. LR – (5.12, 2.08, 5.60), enter scale in table, align corner 0, swap X and Z for coords place.
4. ST – (6.08, 2.08, 6.40), first rotate -90, enter scale (2.67, 3.125, 2.39)
5. UL – (4.48, 1.92, 4.24), enter scale in table, align corner 0

|  |  |  |  |
| --- | --- | --- | --- |
| Scene and ratio (x, y, z) | Cam/Listener coords | Audio/Source coords | Ground Truth Dimen |
| KT (0.83, 1.28, 1.22) | (4.100, 1.705, 1.585) | (1.953, 1.676, 1.982) | (3.46, 2.67, 6.64) |
| MR (0.92, 1.27, 1.06) | (2.12, 1.00, 0.33) | (2.12, 1.00, 3.00) | (4.28, 2.33, 5.61) |
| LR (0.99, 1.39, 1.01) | (2.55, 1.08, 2.79) | (0.51, 1.20, 2.80) | (5.05, 2.90, 5.64) |
| ST (2.39, 3.125, 2.67) | (6.94, 1.50, 5.00) | (4.94, 1.50, 5.00) | (14.55, 6.50, 17.08) |
| UL (1.16, 1.52, 1.31) | (3.58, 1.07, 2.27) | (1.32, 1.07, 2.24) | (5.20, 2.91, 5.57) |



From <https://doi.org/10.1007/s10055-021-00594-3>, in Z, X, Y ( Length, Width, Height).

1. Generated audio for all scenes and set up the audio recording process (using external software: Audacity, Mona uses Wavepad)… concerns regarding audio compression?
2. Organized RIR analysis folders and modified MATLAB scripts for compatibility (absolute path to relative).
3. Reorganised GDP unity project folders to be more clear and easier to navigate. (scene folders, \_scripts etc WIP)
4. Began troubleshooting EDT/RT60 analysis issues.

### Problems encountered (unsolved):

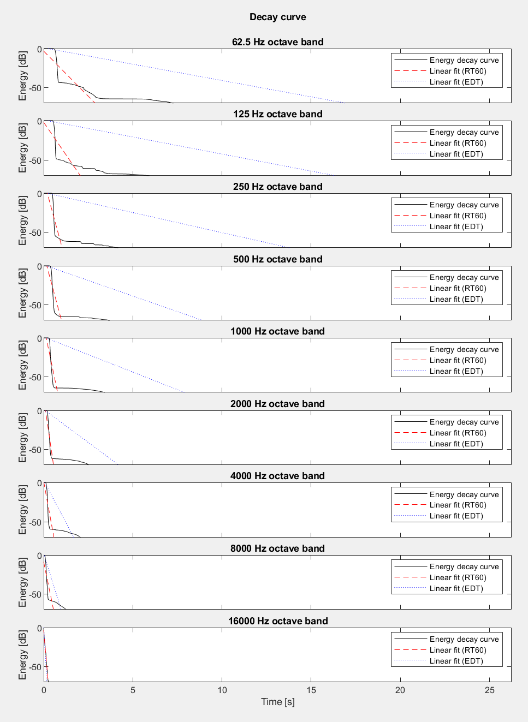
1. Issues with mono depth enhancement and mesh quality. -> fix\_limits function in preproc is the culprit, it causes the issue on close limits, but needed for far limits (for mirrors and proper room scale)
2. ModelDimension calculation bugs in Unity. (it shows dimension depending on the axes, but the model can be on wrong rotation initially, giving wrong width and length reversed.)
3. Persistent problems with audio analysis results, despite trying various audio settings. (tried, trimmed, higher volume, different software), EDT and RT60 is too high.

### Plan for following week:

1. Continue investigating and resolving EDT/RT60 analysis issues. (test other scenes, could be isolated KT scene issue? It seems EDT is most problematic, find out how EDT calculation works, is it due to audio reverb delay bug? )
2. Further optimize the audio recording and analysis process (if needed)
3. Refine Unity project organization and Steam Audio settings. (standardisation/constant for all scenes? Or have proper table/list to refer.
4. Explore additional methods for improving mono depth map quality. (delve deeper into fix\_limits function and modify for mono, monodepth map looks good, just incompatible with edgenet as noted by GDP, just CV problem atm)
5. Implement automation for audio baking in Unity.
6. Create separate Unity scenes for VR and non-VR demos.
7. Add more immersive elements (e.g., Rachel and Joao) to the VR demo. (following S3A)
8. Continue working on pipeline optimization, including Docker clutter removal and LiDAR file management.

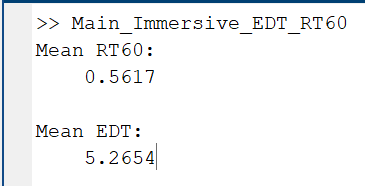
## Meeting with Dr. Hansung notes:

* Continue with plan
* No meeting next week. (next meeting maybe around 6/7th August)
* Finish audio evaluation and analysis perfectly first (with graph etc), Mona’ SSC part a bit different right now so optional instead

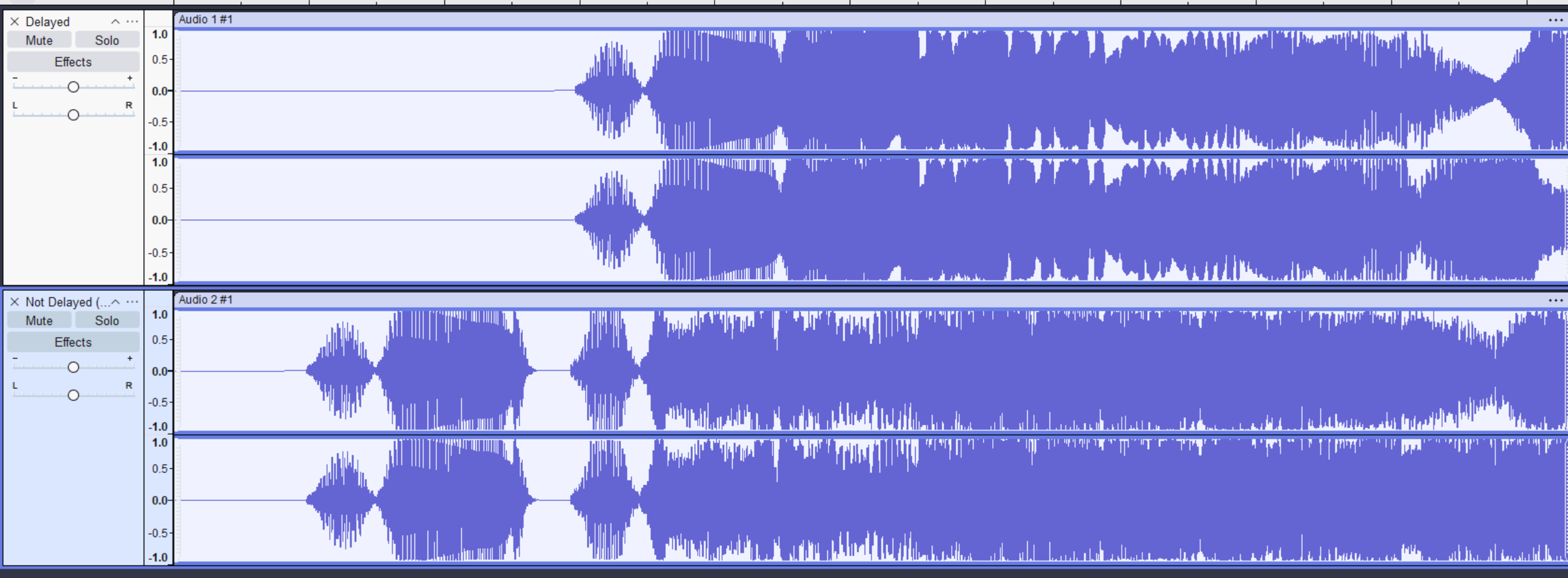
Verdict: figure out the EDT weird bug, must be something wrong with deconvolve or my recorded audio somewhere. Lets try delayed audio again on normal setting (S3A), then Mona’s setting. Also, rerecord audio for other scenes on 48kHz sampling rate.

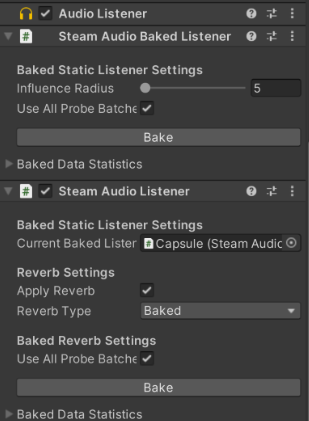
Also, maybe lets try record S3A and see if its probably my recording workflow is the problem.

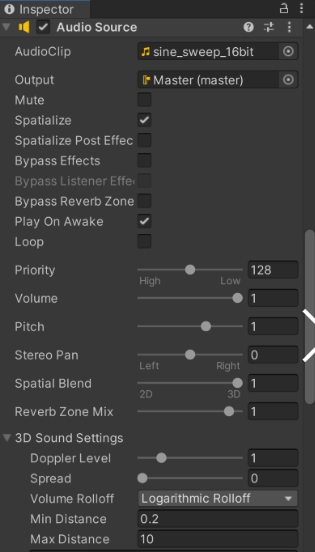
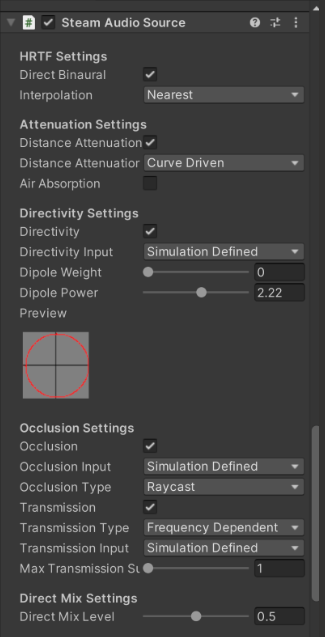
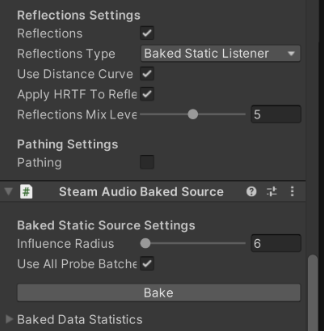
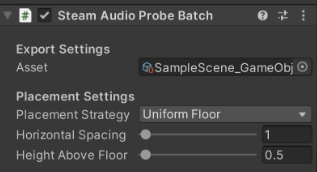
Interestingly enough, RT60 might be correct and only EDT is problematic, and when analysing MR scene audio, it proves that its not isolated issue to only KT scene. Here is the result for MR.

As seen, RT60 is fairly accurate and within range and from right, EDT is absolutely wrong line fit…

NEW BIG FINDING! The audio cutoff at low frequency (bottom waveform) is due to audio reverb delay bug mentioned before.. Compare it to the delayed version!

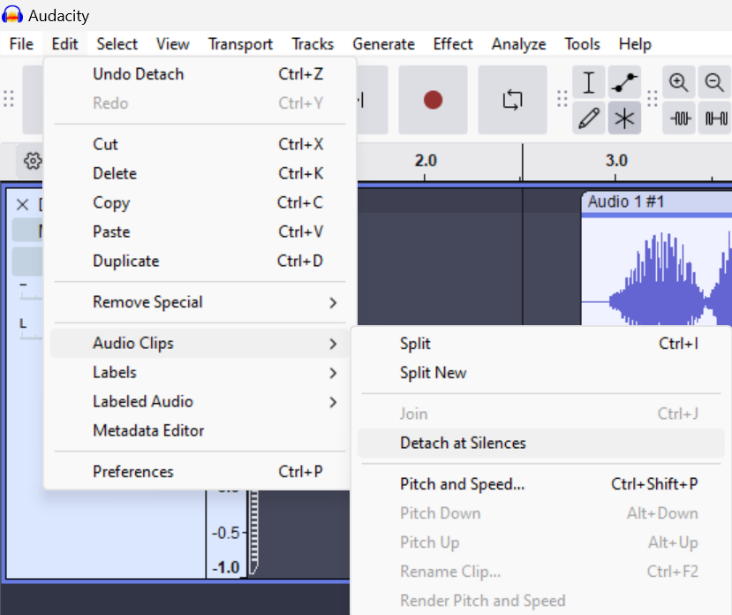
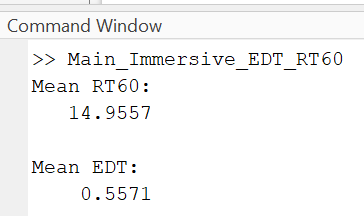


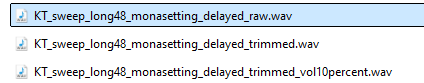
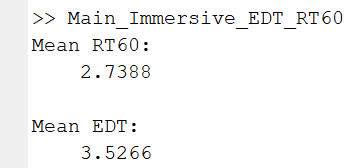
This is using Mona’s setting (top waveform used my delayedaudiodelay.cs script and turn of play on awake) instead of S3A settings (which is unreliable/not standardised/tuned to latest steam audio and might have changed from actual used for evaluation):

This might also be the noise source on the RIR… This is obviously a bug due to the way Steam audio initialise and for play on awake audio source..

Let’s analyse it to make sure.

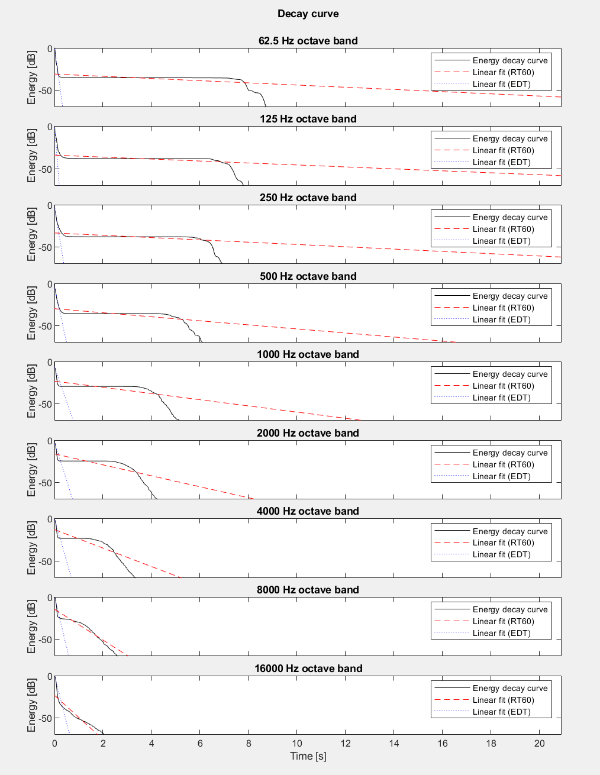
I’m trimming the start as well to follow Mona’s waveform shape.

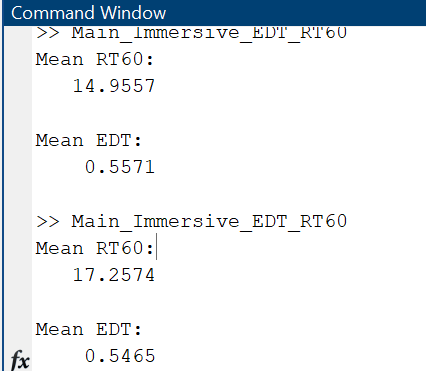
On Audacity, I used Detach at silences and when exporting, trim blank space. This is needed or else would get above RT60. EDT NOW FIXED THO!!!

 Nvm, even trimmed version got same RT60… Could it be the volume its recorded? Trying on 10% windows Volume instead of 50%.. Ok idk if im tripping but I got exactly same result. All 3 variant of mona setting have exact same Mean RT60 and EDT as above… confirmed!

Lets try with custom setting, go back to previous 0.5 vol on listener using 1 min dist and 30 max distance, 1 direct mix, 1 indirect mix level and 50% windows volume. Lets call this customsetting1. Ok this bring us back to previous result ->

Im convinced this result mix is better but the EDT best fit line is buggy..

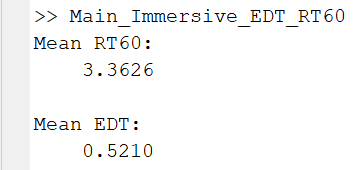
Ok NVM, I think Mona’s setting (stronger mix level for reflections is better in this case), because 1 and 1 mix level gives strong early reflections which what gives very high EDT and RT60 as the best line fit is absolutely broken due to strong early reflections.

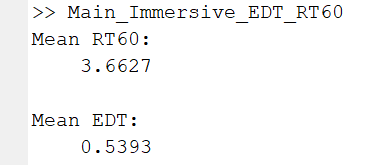
Now I think about it, volume does matter, atleast I shouldn’t let it clip, but not too low that some audio cant be heard. Also stupid windows 11, for some reason, my headphone volume don’t do anything, I need to change it on volume mixer. Unity volume mixer at 20% seems good to not clipping. Also, as proven before, no need trimming/detach at silence.   
🡨 Unfortunately, that still doesn’t fix the RT60 issue, idk why there’s the flat line around -30db on almost all frequency except highest frequency as seen 🡪

Can also confirm using wavepad vs audacity have no difference. Using max vol to clip some freq also doesn’t work, in fact, it increases EDT to higher value (1.5s)… idk man, continue tomorrow.

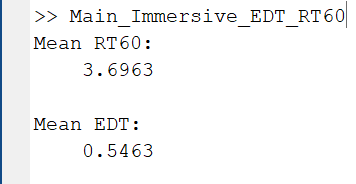
# Thursday, 25 July 2024

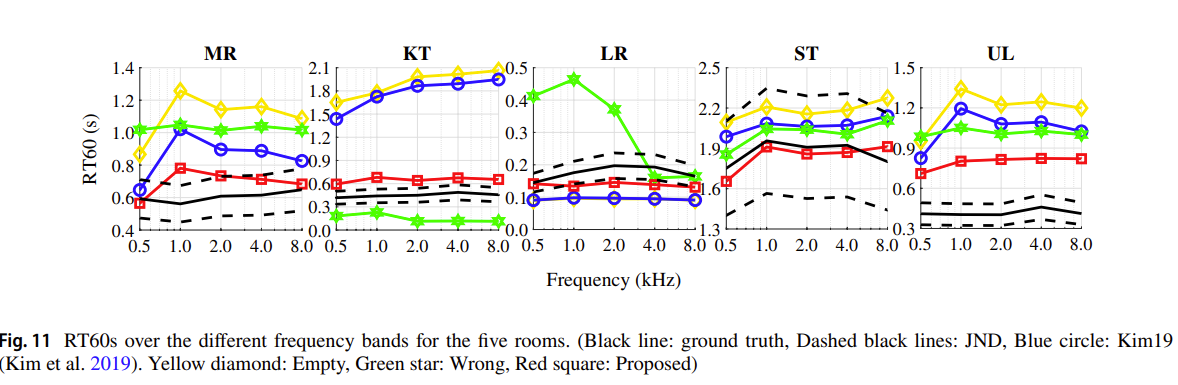
Ok, just realised I should probably try Mona’s setting on MR scene as well to see if the issue is isolated/only for KT. Same problem (too high RT60), but I just realised I haven’t tried Mona settings but only change mix ratio (Direct to reflections).  
Here is 1 to 1 Mix ratio instead of 0.5 to 5 in MR.

As seen, interestingly, it seems fairly accurate now (RT60 still so high, but atleast EDT is very accurate. Let’s try 0.75 to 2.5 next, and find sweet spot > This feels very bad but ig atleast all scenes should have same setting in the end to make this more ‘scientific’ or not too tuned, or else the resulting mesh would not really mean much. Maybe I should tune and calibrate on LR

But for now, lets test same 1 to 1 setting on KT scene and see what happens.  
Also to note, from last day testing and Mona’s email, it seems that volume mixer level (on windows at least, even clipping) don’t really affect much on the result, there’s some but not as prevalent as changing the ratio so far.

As seen, KT using 1to1 mix ratio and mona setting give very similar result to MR, which means now we just need to find the best mix ratio parameter (direct and reflection mix level). In hindsight, this is obvious because Mona’s mesh is still voxel but mine is not so there should be vast different somehow even when using latest Steam Audio plugins, the same can be said with S3A setting. Tho, interestingly Mona setting do fix way too high EDT (which means S3A settings have issue with EDT, due to strong early reflections messing up the line fit, while Mona settings atm gives too high RT60.)

Next, lets try 0.5 on direct mix level and 1 on reflections mix level to see the effect of lower direct mix.

🡨Result, it did virtually nothing. Next lets try 1 on direct mix and 0.5 on reflections mix level. OK, it does slightly decreases RT60 (by 1 sec, but I don’t think this should be the solution, but we already know increasing reflections mix will just increase RT60 as the noise floor level also increases… It feels like I should really just get RT30 and extrapolate to get RT60 at this point, but before that. Lets try LR on 1 to 1 ratio (Honestly, I feel like 1 to 1 should be the standard/realistic setting and should be for all scene.) Ok after looking at Mona’s ST octave bands decay curve, it seems that using RT60 might not be the best after all? Because the mean might beaccurate, but it looks like coincidencewhen looked further into each bang, as some have reverb lower than 1s, and other longer than 2/3s which average it out to be near what Kim21 got. For reference, this is the result from previous paper. For now, I’m thinking of lets just use 1 to 1 (I’m assuming previous early reflections is due to distance attenuation problem? Idk, surely befcause of strong early reflections but need to check more

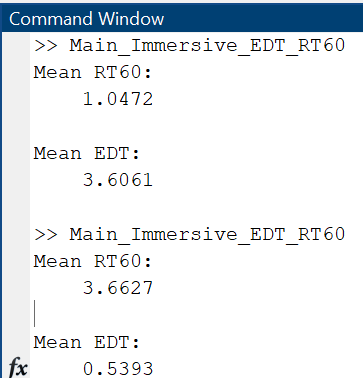
In that case, lets experiment/test more and see what exactly causes that strong early reflection (Which cause very high EDT). Then, lets use RT30 extrapolated to RT60 for RT60 values, and get all scenes value and compare with Kim21 first, because I have doubts on Mona’s result/recording setup/parameters atm. It is unfortunate that Mona didn’t have LR because seeing dead/anechoic room results would be helpful for calibration/testing imo.

## **FOCUSED TODO**

1. Experiment with settings/parameter to see what causes high EDT on S3A setting (strong early reflections, causing bug on line fitting)
2. Use RT30 for RT60 estimation via extrapolation, and generate all graph/value to compare with Kim21, especially for RT60 on each bands as seen on picture above.
3. Try maximise/increase global steam audio parameters to increase accuracy (hopefully?) to take advantage of using baked instead of realtime.

### High EDT on S3A setting investigation

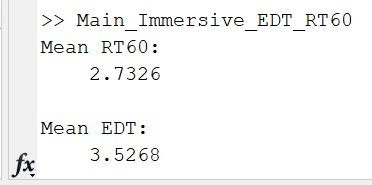
The main culprit right now is the volume and distance attenuation setting, 2nd is the lack of occlusion/transmission.

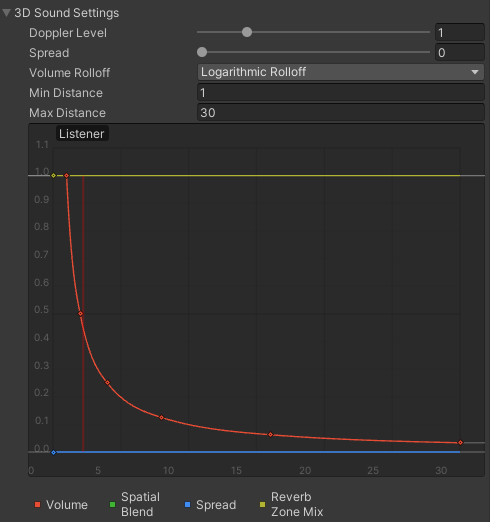
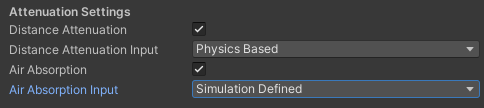
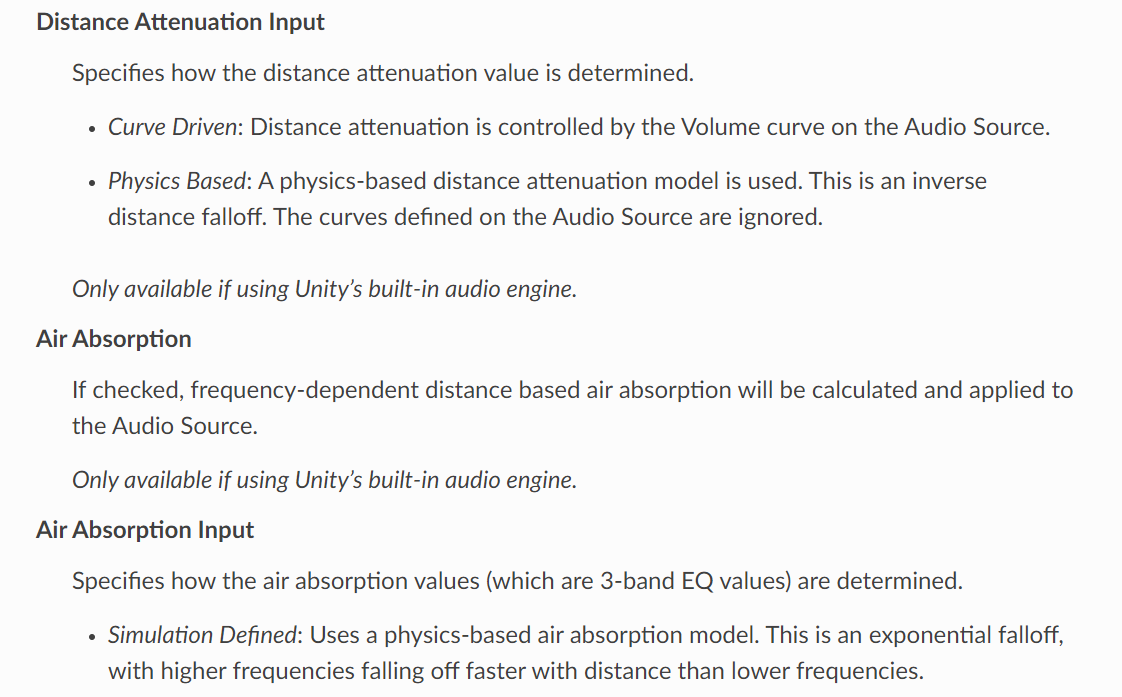
Lets try the 1st one, first set setting according to S3A KT audio source, then reproduce to see if still get same audio result as before. OK, got same result as before can also confirm from octave band decay curve graph that it is due to strong early reflection which causes best line fit bug(?). Top is default S3A, bottom is with Mona volume and attenuation 🡪

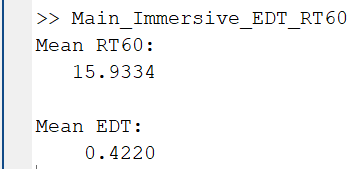
Next lets change the volume to 1 and min distance to 0.2, and max to 10 instead of 1 and 30 (default S3A in KT Eval). OK confirmed that volume in audio source does matter and possibly distance attenuation (checking) as its part of the Unity simulation. Windows/Driver level audio mixer level/system volume however doesn’t because its not simulation related I think as proven before.

Note: only need to bake audio source when changing only audio source parameter (not changing listener location etc), which makes sense, only bake gameobject where u change its parameter.

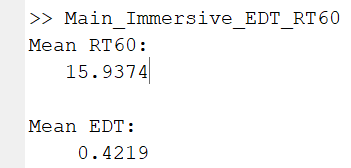
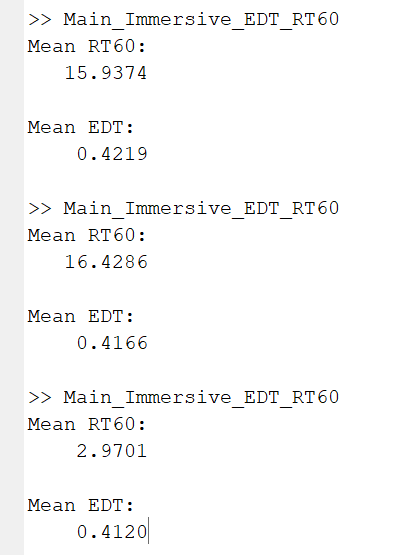
Now lets change only volume from 0.13 to 1, but keep distance at 1 and 30. To see if distance its distance attenuation or volume that matters.

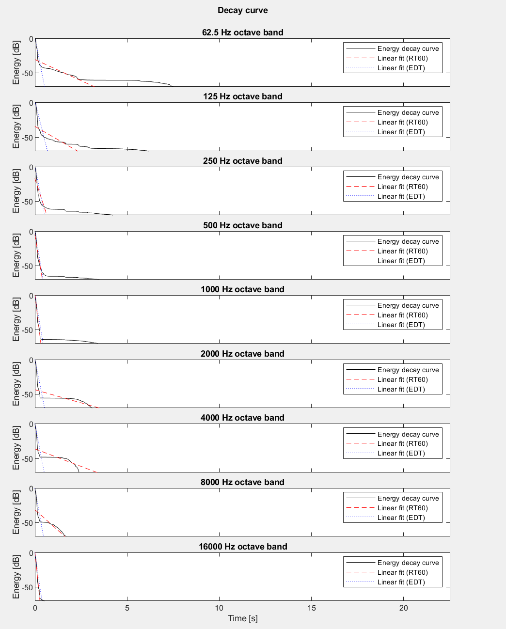
The results shows that distance attenuation is the culprit for EDT, while volume is for RT60 which makes sense. Due to higher volume level, the noise floor level is also higher. Due to longer max distance on distance attenuation (curve driven), the listener position hears stronger early reflection thus higher EDT due to best line fit anomalies.

Question is now, if realism is the goal, why not enable attenuation based on physics based instead of curve driven (which is more useful for directive/game experience with non simulation in mind). At the same regard, might as well enable simulation defined air absorption as well.

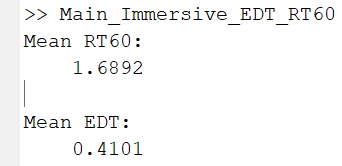
But first, just to see if its usable, lets only enable physics based attenuation instead of curve driven for now.

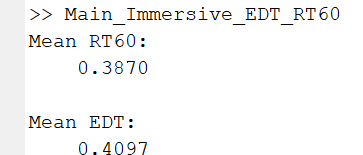
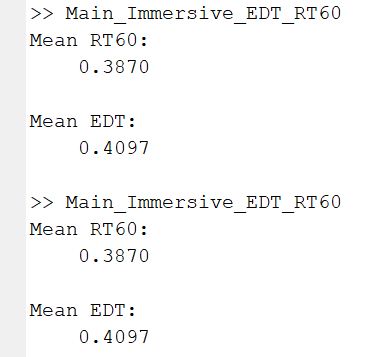
YES! Physics based attenuation works! This is great as saves the headache of thinking whats the best curve for each scene. As expected, because volume is still 1, RT60 still too high due to high floor noise level, lets try enabling air absorption (simulation defined) and see what changes.

Air absorption doesn’t do much, probably more effective on larger room like ST, but I’ll keep it enable for now. Next, lets lower the volume to 0.5 and see if this fixes RT60. Unfortunately, this doesn’t, so lets go back to 0.13 and see if this changes anything. If not, then I think I’m forced to use RT30/45 or something and extrapolate.

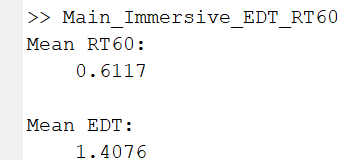
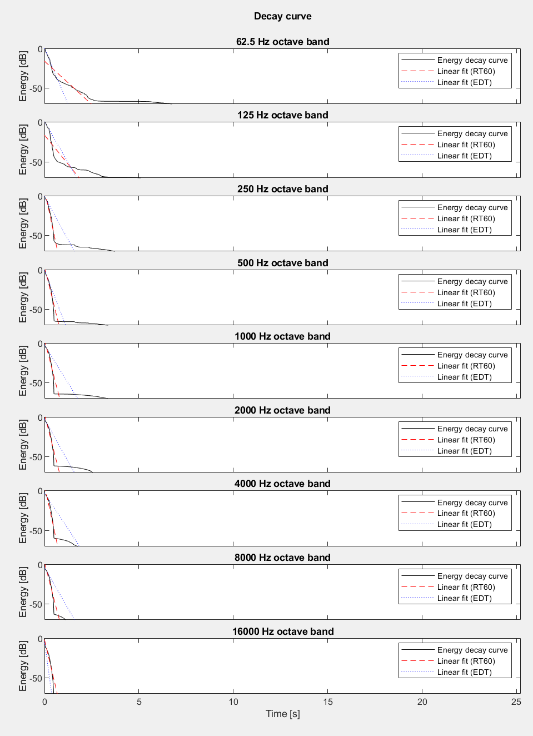
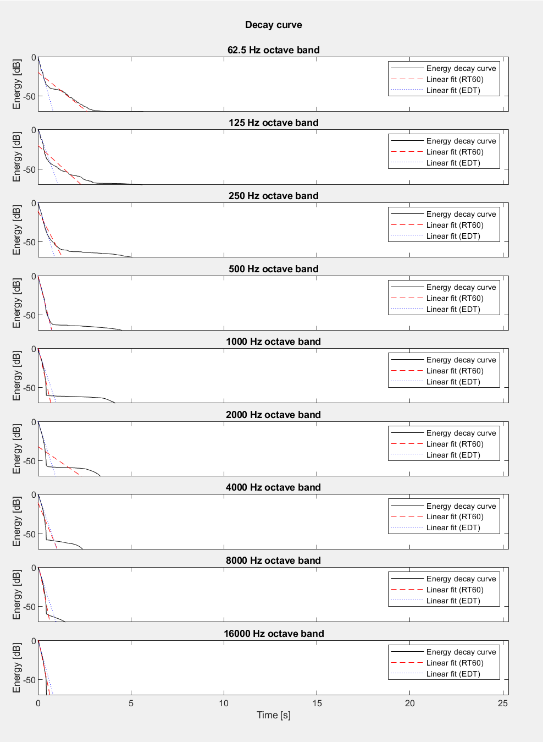
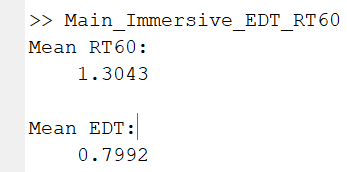
🡨 Top is vol at 0.5 and bottom is at 0.13, as seen it helps a lot but still not as accurate. This is most likely due to the noise floor level being at around -50 instead at less than -60, causing the best line fit problem again as seen in following graph 🡪

Lets try at 0.05 volume just in case. I also maxed (100%) System/Windows volume level/mixer just in case its too quiet.

Volume at 0.05 helps a lot but still not enough as the noise floor still present around -50 to -60db especially in lower frequency but still present in some frequency like 2k and 4k interestingly, so changing band wont really solve it either. Lets try 0.01 just in case, at max System/Windows volume mixer again.

LETSGO IT WORKS SOMEHOW, the result for volume 0.01 🡪 only concern is that whether or not this same setting will work for other scenes, especially like ST, would the low volume cause the sounds to dissipates even before reflecting… This requires more testing for sure. But before that, lets try same volume 0.01 but this time at 50% Windows/System volume mixer level to see if it affects the calculation/recording.

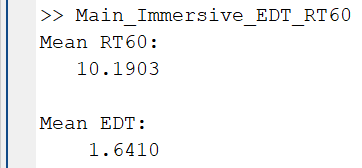
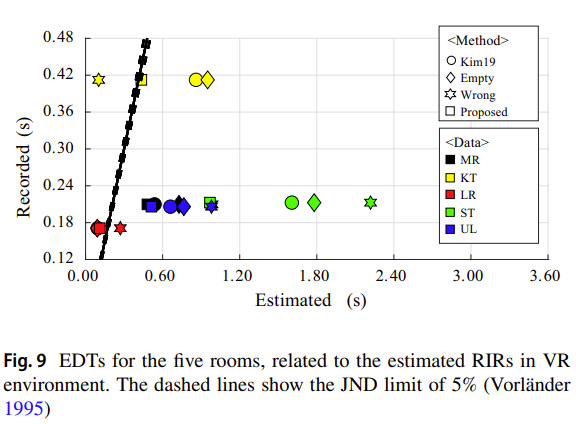
CONCLUSION: NOPE AS TESTED BEFORE, volume mixer/recording level does not matter as long as there’s enough sound/decibel ig. Top is at 100% and bottom is at 50% Windows mixer/recording volume level with audio source at 0.01 volume. Next, let’s setup ST ipynb and test it there to see if very small volume like this would affect it.

BIG BAD NEWS! This is what I got 🡪 for KT, obviously wrong as too small RT60, and EDT higher than RT60. There is strong early reverb somehow, and I think RT60 too low because as concerned, the sound energy dissipated even before reflecting due to too low volume. Honestly, could be because air absorption setting, but I still think it should be enabled.

From left to right: My decay curve, Mona’s decay curve and mean RT60, EDT. Tbh could also be because the way I record is different to Mona? A builtin steam audio recorder is vastly needed ngl.

Lets try changin volume from 0.01 to 1 first. Now both EDT and RT60 are wrong, but atleast RT60 is higher than EDT as expected. Verdict for now is its impossible to get rid of noise floor level and get accurate RT60 measurement, so might as well use RT30 and extrapolate it (idk why my setup is having problem while Mona’s doesn’t)

Next, let’s try changing back the mix level ratio to what Mona had set (0.5 to 5) and see if this fixes EDT value for ST. Now I think about it, I should also enable occlusion and transmissin if realism reproduction is the goal. Going to do it as well (tbf, its probably not going to change much in terms of static listener, as occlusion is more important for moving listener for positioning etc).

NVM, even using Mona setting (w/ physics based attenuation and air absorption), this is what I got. Tbf, even the proposed Kim21 ST got way higher EDT (around 0.95 sec) than actual recorded (around 0.20 sec), so this might just be limitation of Unity Steam Audio simulation and one of its quirk. Interestingly though, Mona managed to get very low EDT (around 0.8, but still not as accurate as recorded) EDT graph from Kim21 paper for ref 🡪  
So verdict so far is to use Mona setting, with physics based attenuation, air absorption, at 0.5 volume (to lower EDT a bit esp for ST), and get RT60 from extrapolated RT30 instead. That’s all for today.