Whitepaper: GA-Prediction vs Measurements how close can they be?

v3, January 13, 2022

Bengt-Inge Dalenbäck, CATT

Introduction

The last decade, and especially the last 5 years, a quite dramatic change in use of GA-based software has been observed and it is necessary to revisit the many factors that affect how close a GA-based prediction can come to a measurement, factors that previously used to be well understood and considered. It is also seen by that the last years more and more warnings have had to be added to the software.

The factors will be divided into:

· Limitations of Geometrical Acoustics (GA) itself

· The particular algorithm a software employs

· The use/user: model, input data, prediction settings, checking, listening, and asking

· Measurements compared to

It is important to realize that the factors affecting how good a prediction can be are multiplicative rather than an average. Let us assume that the factors are in the range [0..1] where the one depending on GA is *a* say = 0.9 (i.e. a good case for GA prediction), and the rest are: algorithm *b*, the use/user part: *c*, *d*, *e*, *f*, *g*, *h*, then the overall prediction quality will be:

$$q_{mul} = a \cdot b \cdot c \cdot d \cdot e \cdot f \cdot g \cdot h$$

and not an average

$$q_{avg} = (a + b + c + d + e + f + g + h)/8.$$

An example a = 0.9, b = 0.8, c = 0.95, d = 0.7, e = 0.5, f = 0.7, g = 0.4, h = 0.2 then:

$$q_{avg} = (0.9 + 0.8 + 0.95 + 0.7 + 0.5 + 0.7 + 0.4 + 0.2)/8 = 0.64$$

that does not seem so bad, but for the multiplicative effect:

$$q_{mul} = 0.9 \cdot 0.8 \cdot 0.95 \cdot 0.7 \cdot 0.5 \cdot 0.7 \cdot 0.4 \cdot 0.2 = 0.013$$

In reality some factors will have weights or functions for their importance in a given case but the important conclusion is that it matters little if e.g. a model is well done, and the absorption is well set, unless the other factors also are high. A typical example can be that all factors are high except the prediction settings (number of rays typically that are set very low or left as the initial suggestion) and no checking and it topples the whole prediction.

Limitations of Geometrical Acoustics (GA) itself

Here is since some years back already a separate whitepaper “What is Geometrical Acoustics?” [1]. The last decade there have been many instances where a user had not heard of GA and e.g. assumed it to meant “making a geometrical room model”, “a way to analyze the acoustics of a room” or “first you make points then you make planes”.

The particular algorithm a software employs

Let us assume that the case at hand is a good candidate for GA, the next factor is the GA-based prediction algorithm (software) where there still are quite big and fundamental differences especially for non-diffuse rooms. Fundamental requirements are fully frequency dependent scattering (i.e. in all algorithm parts), an algorithm that can predict flutter echoes and double-sloped decays (which can be due to flutter, coupled spaces or a non-mixing geometry and unevenly distributed absorption), in short is able to handle non-diffuse cases. In addition there will be differences in e.g. loudspeaker handling, array modeling, early diffraction if any, and more, and features, but this short whitepaper will focus on pure room acoustics prediction with an ideal omni-directional source. There may also be misleading use of marketing statements like *highly accurate*, *precision* and even actually claiming *exact prediction* [2] which risks that especially new users, that now more and more often lack a proper background, will assume that a good prediction is somehow automatic but it is very far from the case.

The software use/user: model, input data, prediction settings, checking, listening and asking

Let us assume that the case at hand is a good candidate for GA and the software used have sound and well tested algorithms, is that not enough? Unfortunately not since the software does not make the model, does not set absorption, does not set scattering, does not choose prediction settings (at best it can give initial suggestions), and does not evaluate (look at) or listen to (if available) the results. Even a very good GA-based software can only give a potential for a good prediction, it is not automatic. It has several times been shown that when the same case/room is given to several software users (even of the same software) the predicted results can vary significantly, like e.g. T_{30} values between 0.3 and 1.0 sec (measured was 0.5 sec) and even SPL varied up to 12 dB [3] and where most differences simply were due to very varying input data. The first time this became apparent was in the first software round-robin in 1994 [4] where in a first phase the participants (a mix of developers of the many software at that time and users, 17 participants) were to set their own input data (absorption, scattering up to each participant if their software used it) and make their own models based on the same type of information as was typical for a consulting project and the results varied dramatically. Then in a second phase all used the same input data and the differences became much smaller but still big due to at that time the software varied more and had more limitations. Unfortunately this has not changed in a fundamental way since then e.g. due to a more common assumption that with a software no questions have to be asked and not much checking and thinking is needed. The quote below summarizes it well, a sort of long editorial or keynote that was about software and prediction in general and development of software, also outside of room acoustics, FEM for cars etc. [5]:

> “*First of all, and unlike the technology enthusiast, the mainstream user will assume the technology that is implemented as established - it works, no (or not too many) questions asked*”.

That is were we are since several years back but GA remains GA and all parts in this section remain as important.

Model: each software has its rules or recommendations on how a model should be constructed but common should be warp-free surfaces, reflecting side faced correctly, only surfaces that can be reached by sound modeled and no holes (if it is to be a closed room modeled).

Input data: naturally absorption coefficients play a big role but may be unknown, especially in existing rooms whose acoustics is to be improved, or have limited accuracy such as high absorbing porous absorbers. Surfaces with fine details are to be replaced by larger ones where the omitted details are compensated for by frequency dependent scattering and, if for model-construction reasons, some small surfaces have to be used apply software dependent manual or more automatic scattering depending on surface size vs wavelength. However, as discussed in [1] there is no way, within GA, to make small surfaces reflect correctly, using scattering that drops with frequency (as the surface size / wavelength ratio decreases) is only a way to avoid strong specular reflections from such surfaces.

Prediction settings: items such as number of rays and echogram/IR length are naturally case dependent and need to be set and verified to be sufficient. Non-diffuse rooms naturally need more rays than diffuse rooms, and echograms need to be long enough for the longest T_{30} to be evaluated (down to -45 dB but for auralization down to -60 dB). Typically, recommended values can only be initial suggestions and e.g. regarding echogram length can initially only assume Sabine to be valid (for most rooms modeled now Sabine is not valid) perhaps with some initial lengthening if the room has an uneven absorption distribution. It may be argued that number of rays should be automatic so that rays are sent out until some measures have stabilized but I, being old school, consider the “getting to know the room” stage to be a very important part even crucial, for GA use, and if a room requires many rays it is generally also a room harder to predict well such as it being a very non-diffuse room. The more that is performed automatically the less the user will think about the case and the room becomes a black box. It will also mean that very little is learned with each room so an ability to predict better over time may not develop.

Checking: OK, let us now assume that we have a prediction result. Before jumping directly to look at T_{30} numbers, which clearly is typical now, echograms and decays must be studied to see if they look plausible or the Prediction settings have to be changed (more rays typically). And to know what IRs and decays look like it is naturally necessary to have seen many measured room IRs.

Listening: as an extremely useful way of checking, if a software has realistic auralization, is to listen to the predicted room impulse responses (IRs), if they do not sound natural it is likely a case of too few rays (or more dangerously if it is an algorithm that can give a plausible-sounding IR even with very few rays). To know if it sounds natural means to also have listened to many real measured room IRs.

Asking: if something looks or sounds strange, and can’t be figured out, send the model files and prediction settings to the developer in question and ask for feedback. Naturally it is extra important if it is one of a user’s first projects, or a complicated or new type of project, or if it is a prestigious case. But as can be realized due to the many factors involved a very good prediction can never be guaranteed, these software are tools to be used with understanding and experience and are not absolute predictors. One of the clear changes the last 5 or so years is that very few questions have been asked.

Measurements compared to

Let us assume that the case at hand is a good candidate for GA, and that a software with sound algorithms has been used well by a person with a good background and experience in acoustics and GA, and that the model, the input data, and prediction settings have been well considered, and a sanity check performed on the results. What remains then is that the measurements have not been good enough. Generally it is the least common reason for bad prediction/measurement comparisons but it is still common. The most common for T_{30} is that the measured IR contains noise that has to be compensated for, even in one of the seven round-robin CATT-Acoustic has participated in T_{30} was due to background noise evaluated too long. Also Lateral Fraction can vary considerably depending on who measured, in [6] for round-robin III big variations were reported (the measured value for one position varied between 8-35% at 125 Hz and 19-37% at 500 Hz depending on the measurement group). Since about 20 years, if model and input data essentially looks OK, I no longer look for potential algorithm problems, or dig deeper into the case, until I have seen some measured IRs and all too often the problem has been the measurement. I have no proof but am suspecting that comparisons to measurements are now much less common than they used to be, but it is really the only way for an individual software user to ensure that his/her predictions improve over time and that the GA-based modeling is not just a more or less synthetic blind and risky exercise.

A common mistake for small absorbing rooms like classrooms, or receiver positions close to a source, and a dodecahedron has been used for the measurement but compared to a prediction with an ideal omni. If it is not considered, or known, that a dodecahedron is allowed to vary with angle up to +/-6 dB at 4k [7] (it used to be up to +/-8 dB from 1kHz and up in a previous standard). Even the direct sound can be wrong and a first order reflection in a hard wall can become stronger than the direct sound, and there is little chance for a decent comparison. Such a mistake has even appeared in a reviewed article where it then was blamed on the software used.

A critical approach and repeated “self-accreditation”

As a summary a critical approach has to be used to have a chance to achieve good enough predictions. Not only being critical about the quality and capabilities of the particular algorithm used, overall and for each given case, which in deed is very important (it is then also a factor that may be common to all or most prediction projects). Many times new users should actually test/verify themselves, such as e.g. creating a simple room that ought to create a flutter echo and see if one is predicted, or to test simple but non-diffuse rooms and see if T_{30} will still be (too) close to Sabine. To predict some old already measured cases is a very good way to start with GA prediction. Not only that but also to be self-critical and not just accept prediction results without careful thinking but think about the room and look at the model. There have e.g. been many cases where a room is highly likely to give severe flutter echoes and even if the prediction has shown it, and auralization makes it apparent, it has still been missed, or prediction settings, such as number of rays, have been insufficient to show it (but then the echogram often looks, and sounds, like a train-wreck and should anyway have been a clear warning-signal). In many of theses cases the flutter could/should have been realized even without any prediction at all but simply by looking at the model. Since room acoustic consulting is growing rapidly it is also necessary for users to regularly repeat a “self-accreditation” since if someone say 2005 had a good background and experience and had become good at GA prediction it does not mean that when a new employee starts at that company in 2015 automatically and immediately will be equally good at it. Why not have a number of “house benchmarks” where new employees model and set their own input data and re-predict and it can be compared to the measured data and previous predictions?

References

[1] http://www.catt.se/What_is_Geometrical_Acoustics_Generic.pdf
This is the generic version that does not refer to CATT-Acoustic documentation details.

[2] One much used GA-based software has publically claimed *exact prediction* since many years, Dec 12 2021 it was still claimed and the first time it was noticed by me was in Dec 2012.

[3] This official user round-robin can be requested from bid@catt.se, this whitepaper will not discuss individual software.

[4] “International Round Robin on Room Acoustical Computer Simulations”, M. Vorländer, ICA Trondheim, 1995 (presentation of results)

[5] EuroNoise ’95. Unfortunately I had not copied the title page with the author name and have not been able to find out who it was.

[6] “Report on the 3rd Round Robin on Room Acoustical Computer Simulation – Part I: Measurements”, I. Bork, July 2005 Acta Acustica united with Acustica 91(4):740-752

[7] “The Source Directivity of a Dodecahedron Sound Source Determined by Stepwise Rotation”, C. Hak, R. Wenmaekers, J. Hak, R. van Luxemburg, Forum Acusticum Aalborg 2011