Study of Phase Reconstruction Methods Employed at Room Acoustic Simulation

Bruno Masiero*, Sönke Pelzer

Institute of Technical Acoustics, RWTH-Aachen University, Email: {bma,spe}@akustik.rwth-aachen.de

Introduction

For an increase in truthfulness of room acoustic simulations, the directivity of the sound sources must be taken into account. But information of source directivity is usually stored in directivity balloon format, only containing 1/3 octave averaged sound pressure level information, thus neglecting all phase information from the source's radiation pattern. Now, a spectrum with inadequate phase component might lead to disastrous results when simulating impulse responses.

Phase reconstruction methods with varying degree of complexity can be applied to guaranty a compact and even causal impulse response (IR). This work deals with the question if humans can differentiate IRs with different phase contents. This question will be answered based on comparative listening tests, which presents various types of signal convolved with simulated IRs of two rooms with varying simulation parameters. For each simulation, three different IRs are computed, each based on a different phase reconstruction method.

The intention is to verify if the difference between IRs calculated with increasing degree of computational complexity is audible, giving an indication on how IR calculation should be implemented on a real-time room acoustics simulation software.

Directivity Database

The real-time room acoustics simulation software Raven [1] is able to account for source directivity in its calculations. To do so, it extracts the directivity information from a source directivity file. These files store the source's far-field radiated energy in an angle resolution of 5° and averaged in 1/3-octave bands (with center frequencies from 20 Hz to 20 kHz) [2]. The values are normalized in the range from 0 to 1.

To extract the far-field transfer function from this source in a given direction, Raven reads the energy values of the point in the grid nearest to the desired point. These 30 energy values are then interpolated to 257 linearly distributed frequency bins. The vector is then used to produce a IR with 512 samples length.

The drawback of this method is that no phase information is recorded, meaning that the resulting IR is wrapped non-causal (NP). When such an IR is extended with zeros, its frequency content is altered and a comb-filter effect occurs.

*Scholarship holder from CNPq–Brazil.

Phase Contents

To avoid the comb-filter effect, two new IRs with artificially generated phase contents are analyzed. The first solution is to unwrap the IR, i.e., to shift the second half of the IR in front of the first half and to eliminate its DC component (allowing zeros to be smoothly appended to the IR). This computationally cheap operation results in an IR that is still non-causal (LP), with a symmetrical pre-ringing, what would clearly not be expected from a real sound source. The second suggestion is to create an IR with minimum phase [3, Chap. 10.3], that results in a causal IR whose energy is concentrated in its early part as would be expected for a natural sound source (MP). This operation is considerably more cost intensive to compute than the previous operation and if both IRs present no audible difference, the first method should be preferred in a real-time simulation environment.

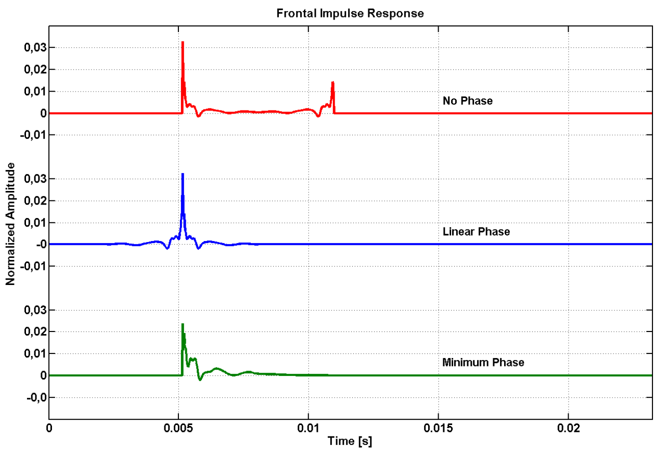

Figure 1: Comparison of IRs with same magnitude spectrum but different phase components.

Listening Tests

To verify if the difference between the IRs generated with the three phase reconstruction methods is perceivable, a listening test was performed. The test was set-up as a "three-alternative forced choice" (3-AFC) test. Three stimuli – two of them being identical – are presented in random order to the test subjects, whose task is to choose the odd stimulus among the three presented options. The hypothesis assumed in this test is that if subjects correctly identify the odd stimulus with a high degree of confidence, than the difference between the two presented stimulus is audible.

The presented stimuli were three anechoic recordings of voice and solo musical instruments (mallets and brass) convolved with the IR simulated for two rooms: a small cubic chamber simulated once with highly absorbent walls and once with highly reflective walls, and a large concert hall simulated once with only direct sound and reverberant decay and once with the complete IR (direct sound, early reflections and reverberant decay). Each IR was calculated with the three phase reconstruction methods as described in the previous section.

For each combination of signal and room type, three comparison pairs were presented:

Table 1: Comparison pairs presented to test subjects.

A	Minimum Phase Vs No Phase
B	No Phase Vs Linear Phase
C	Linear Phase Vs Minimum Phase

Results

The following factors were analyzed from the listening test results: audibility of IR difference, influence of signal type and influence of room size. The result of a 3AFC test is a dichotomous data set and so it cannot have a normal statistic distribution. Even though the ANOVA test is based on the assumption that the data under analysis has normal probability distribution, Knoke [4] showed that if a dichotomous data set is large enough, the ANOVA method can be used to test the overall hypothesis of no difference among groups of dichotomous responses.

The one dimensional ANOVA analysis on the subjects responses ($F = 1.43, p > 0.01$) could not reject the null hypothesis (H0) that subjects answer do not differ in accuracy, suggesting that test subjects performed the test equally well. Thus the response from all subjects was used on the following analysis.

Overall IR difference

Here the formulated null hypothesis was: difference between pair elements was equally audible in all three comparison pairs. This H0 hypothesis was rejected by the one-way ANOVA ($F = 89.06, p < 0.01$). Doing a pairwise comparison of the result it was possible to verify that the discrimination rate of pair C was significantly lower than the discrimination rate of pairs A and B (which were not significantly different between themselves). This result can be interpreted as follows: The pairs A and B have a high discrimination rate, allowing the interpretation that stimuli LP and MP are distant to NP in a given perception space. The pair C has low discrimination rate, allowing the interpretation that stimuli LP and MP are close to each other in this same perception space. That suggests that the stimulus NP is audibly different from the other two stimuli and that the difference between the stimuli LP and MP is not audible.

Influence of Room Size

The null hypothesis that room size have no influence on the discriminability of the IRs was rejected ($F = 41.65, p < 0.01$). A multiple comparison test shows that even though for both rooms the discriminability of pair C is significantly lower than that of pairs A and B, for the larger concert hall the discriminability of pair C was significantly higher than that in a small room.

Influence of Signal Type

The null hypothesis that signal type have no influence on the discriminability of the IRs was also rejected ($F = 44.56, p < 0.01$). A multiple comparison test shows that even though for all three signals the discriminability of pair C is significantly lower than that of pairs A and B, for the voice signal the discrimination rate of pairs A and B is significantly lower than for the other two signal types. And for mallets the discrimination rate of pair C is significantly higher that of the other two signal types.

Summary and Discussion

Even though the overall analysis showed that stimuli convolved with an IR with zero phase are significantly different than stimuli LP and MP, an individual analysis showed that some combinations of simulated room and convolved signal show a behavior that is significantly different from the overall behavior.

Two important distortions to the overall result could be verified. The first occurs when a mallets signal is played in a small reverberant room or in a large room with no early reflections. In this case all comparison pairs present high discriminability rates and the null hypothesis that all comparison pairs have the same discrimination rate cannot be rejected ($F = 0.21, p > 0.01$). This indicates that in this situation all three stimuli are audibly different. This is probably caused by the transient and broadband characteristics of the percussive content which easily discloses differences of the prominences in the IR. The second occurs when a voice signal is played on a large room. Here, the discrimination rate drops and the null hypothesis that all comparison pairs have the same discrimination rate cannot be rejected ($F = 2.68, p > 0.01$). This effect is probably due to the fact that the speech signal is continuous and thus masking particular anomalies of the IR.

References

[1] Tobias Lentz, Dirk Schröder, Michael Vorländer, and Ingo Assenmacher. Virtual reality system with integrated sound field simulation and reproduction. *EURASIP Journal on Advances in Signal Processing*, 2007:19, 2007.

[2] D. Schröder, S. Reuter, G. Behler, and I. Assenmacher. Multimodale Darstellung natürliches Schallquellen in virtuellen Umgebungen. In *Fortschritte der Akustik: DAGA 2008*, 2008.

[3] Alan Oppehnheim and Ronald Schafer. *Discrete-Time Signal Processing*, volume Signal Processing Series. Prentice Hall, 1st edition, 1989.

[4] James D. Knoke. Multiple comparisons with dichotomous data. *Journal of the American Statistical Association*, 71(356):849–853, 1976.