THE EFFECT OF EARLY REFLECTIONS ON PERCEIVED TIMBRE – ANALYZED WITH AN AUDITORY MODEL

Tapio Lokki[1], *Ville Pulkki*[2], *and Lauri Savioja*[1]

Helsinki University of Technology
[1]Telecommunications Software and Multimedia Laboratory
P.O.Box 5400, FIN-02015 HUT, Finland
[2] Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000, FIN-02015 HUT, Finland
{Tapio.Lokki,Ville.Pulkki,Lauri.Savioja}@hut.fi

ABSTRACT

In this paper the effect of early reflections to perceived timbre is studied. We apply room acoustic modeling to obtain simulated impulse responses which are explored. The timbre of modeled impulse responses is predicted with an analysis method motivated by auditory perception. Such analyses are utilized to determine guidelines for the required set of early reflections to be rendered for a high quality auralization. The results suggest that many orders of reflections have to be searched to guarantee that all possible reflection paths before a certain time stamp are found.

1. INTRODUCTION

Virtual acoustic environments can be created either with physics-based [1] or with perception-based [2] auralization. In the latter one, the auralization is not based on exact room acoustic modeling, instead perceptually relevant parameters are used to create a natural sounding virtual auditory environment. In physics-based auralization the rendering is based on room acoustic modeling and the aim is to render a 3D model of a space audible.

In auralization systems the rendering process can be divided into two parts: early sound (encompassing the direct sound and early reflections) and late reverberation. This division originates from human auditory perception since early sound influence more to the perceived auditory environment than late reverberation. Due to this reason, late reverberation is often modeled with an algorithm which produce diffuse reverberation, regardless of the geometry. In contrary, early sound is modeled as accurately as possible by simulating sound propagation inside the 3D model under study.

In room acoustic modeling early reflections can be conceptually further divided into specular, diffracted and diffuse reflections. This division can be seen as decomposition of early sound field [3]. In this decomposition concept each reflection can be modeled with an “equivalent source” which emits a wavefront. Actually, the well known image-source method [4, 5] is based on such decomposition since in the image-source method each specular reflection is modeled with an image source. Correspondingly, diffraction from edges of the surfaces can be modeled with edge sources and diffuse reflections with surface sources [3].

In this paper we utilize the DIVA auralization system [1] to calculate impulse response for a case study. The DIVA system models specular early reflections with the image-source method.

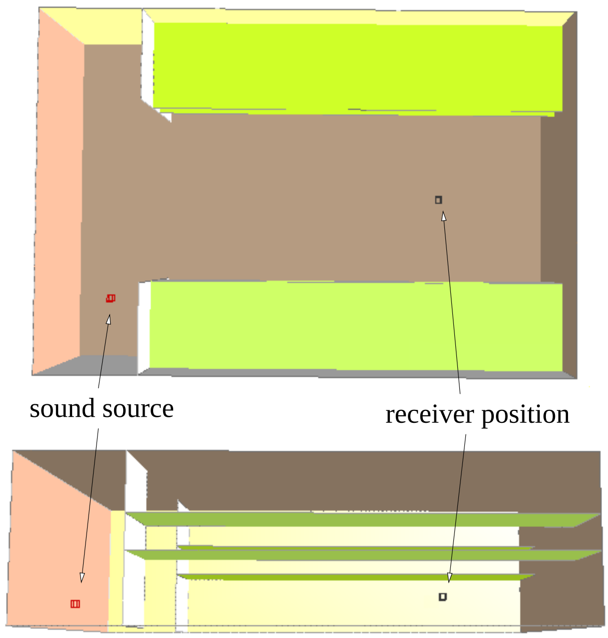

Figure 1: *Simple room geometry utilized in simulations. Sound source was positioned so that it is occluded by a surface from the receiver viewpoint.*

In addition, edge diffraction [6] is modeled and added to the image-source method as diffracted image sources [7, 8]. Diffracted image source denotes an image source that contains diffraction from at least one edge and any number of specular reflections. In this study we have allowed only one diffraction per a diffracted image source. Furthermore, call image sources (corresponding specular reflections) with *specular IS* and diffracted image sources with *diffracted IS*.

The geometry for this case study is depicted in Fig. 1. It is a simplified rectangular room model containing in addition two walls and four shelves inside the room. The dimensions of the room are 16 m x 26 m x 7 m corresponding to dimensions of a small concert hall or a large lecture room. The source and receiver positions are indicated in Fig. 1 and they are positioned so that

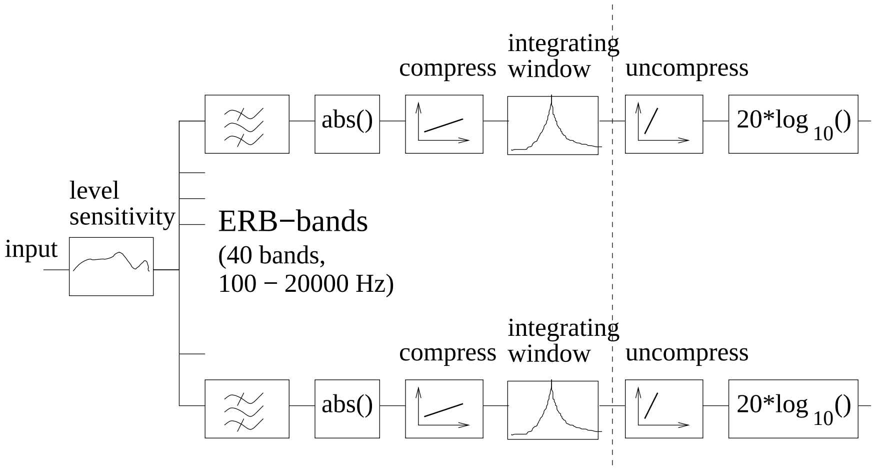

Figure 2: *Block diagram of the analysis method utilized to study timbre of simulated responses.*

the direct sound is occluded by a surface from the viewpoint of receiver.

With this example geometry simulations with different order of specular and diffracted ISs were calculated to be able to study the effect of early reflections to the perceived timbre. In these simulations all specular and diffracted ISs were processed with distance delay and attenuation, and air absorption. For simplicity, the sound source and receiver had omnidirectional characteristics. All surfaces were considered as hard walls, i.e., no material absorption was applied. In addition, surfaces were assumed to be rigid and flat resulting that diffuse reflections could be neglected. Naturally, such simulations do not correspond to any real world case, but these simplification were made to emphasize the effect of early reflections to the perceived timbre. In other words, these simulations show the worst possible case.

2. ANALYSIS METHOD

The simulated impulse responses have been analyzed with an analysis method motivated by auditory perception [9]. The resolution of this analysis method is adapted to be the resolution of human hearing.

A block diagram of the applied analysis method is presented in Fig. 2. First block models the level sensitivity of human auditory system with a frequency weighting filter, fitted to the inverse of the 60 dB equal loudness curve [10]. Then the signal is fed into a gammatone filterbank [11] which divides it into 40 equivalent rectangular bandwidth (ERB) scale bands [12, 13], simulating the frequency resolution of human ear. After the division into ERB-scale bands absolute signal values are taken. For implementation reasons absolute values are used instead of the half-wave rectification which happens in the hair cells of human ear. The next stage in the analysis is formed by a compression and a sliding window which together roughly simulate the time resolution of the ear [14]. The final step of the analysis is to use a proper mapping for visualization purposes. By uncompressing and taking the logarithm of the rectified and temporally processed signal in each frequency band the decibel values can be depicted in a time-frequency plot. Auditory loudness level scale measured as phons could be utilized as well, but in this study we used decibel scale.

2.1. Timbre

Timbre has been defined by the American Standards Association [15] as “*that attribute of auditory sensation in terms of which a listener can judge two sounds similarly presented and having the same loudness and pitch are dissimilar*”. Furthermore, a revised definition [16] is “*Timbre is that attribute of auditory sensation whereby a listener can judge two similarly presented sounds are dissimilar using any criteria than pitch, loudness or duration*”. Therefore, it can be assumed that timbre is related to auditory spectrum and its changes with time of a sound object. We estimate the timbre by simulating and monitoring the auditory spectrum with time utilizing the auditory model.

3. SIMULATIONS

Typically, image sources considered in auralization are searched according to their order. For example, all image sources may be calculated up to third order, corresponding one, two, or three bounces from surfaces before reaching the receiver. The reason for this common practice is that the number of image sources grows exponentially in function of reflection order, thus leading to laborious simulations with higher orders of image sources. However, the order of image sources does not imply to which time stamp they appear to impulse responses. Since the perceptual relevance of early reflections is mainly related to their appearing time and level, the impulse responses might not be valid.

In section 3.1 we monitor the responses generated by IS of different orders. However, these responses do not imply which IS contributes to the perceived sound, i.e., which of them are audible. Thus, in section 3.2 we study predicted timbres of these responses to find which IS orders generate audible timbre changes.

3.1. Energy time curves of ISs of different orders

In Fig. 3 we plot energy time curves (ETC) (squared impulse responses with logarithmic magnitude scale) of specular and diffracted ISs up to the 6th order. ETCs of specular ISs are considered first. The ETCs overlap in time prominently. The starting time t_{beg} of impulse response seems to be related roughly linearly to IS order N as $t_{\mathrm{beg}} = N * 10$ ms. Thus, at least in this case, to compute all specular ISs arriving before 50 ms time stamp, ISs should be computed up to 5th order. To compute all specular ISs up to 100 ms, computations should be reached up to 10th order, if the assumption of ETC starting point is valid also with higher orders of ISs.

Sound signals from first diffracted ISs of each IS order seem to arrive before responses from specular IS appear. To compute all diffracted ISs arriving before 50 ms time stamp, the order of ISs should be larger than 6th, it can be estimated that 8th order IS should be enough.

3.2. Effect of IS order to timbre

We have studied the effect of IS order and inclusion of diffraction to perceived timbre. A reference timbre was selected to be all specular and diffracted ISs computed up to 6th order. The timbres for different cases are computed and plotted, and the difference from the reference timbre is taken and visualized in each case. Therefore, we can monitor how the timbre evolves with IS order, and how it approaches the reference case.

In Figs. 4 and 5 the predictions of perceived timbre are visualized for simulated impulse responses with different orders of ISs. In upmost row the timbres are shown, that would be perceived if only specular ISs were auralized. In second row the difference between reference and specular ISs is shown. Third row presents the timbre when diffracted and specular ISs are considered in modeling. Fourth row presents the difference between the reference timbre and timbre computed with diffracted and specular ISs. Orders if ISs from one to six are presented in different columns.

Due to the geometry of the room, there are only two first-order reflections in this case. In addition, only one diffraction is present, which is shown very faintly at low frequencies before first reflections (the direct sound is occluded by a surface). In difference plots it can be seen that simulation of neither specular nor specular and diffracted first-order ISs provides us feasible auralized timbre. Prominent differences occur in both cases from 0 ms.

Second-order specular ISs do not change the situation prominently. The response continues up to 80 ms, however the difference to reference timbre is large at all frequencies and all time stamps. When diffraction is taken into account, prominently better results are obtained roughly to 15 ms time stamp, although there exists some differences at frequencies near 3 kHz. When this is compared with plotted ETCs in Fig. 3, it can be seen that third-order ISs appear near 15 ms time stamp, which corresponds to differences in timbre. The differences near 3 kHz may be due to third-order diffracted ISs that arrive at around 15 ms time stamp.

Third-order specular ISs make the difference to the reference smaller between 20 and 40 ms time stamps. It seems that 0-20 ms time region can not be modeled at all with specular ISs. Third-order specular and diffracted ISs provide perfect result up to 20ms, where there are some deviations at frequencies above 2 kHz for a short time period. The timbre is fairly good up to 30 ms stamp, after which prominent differences occur. The differences near 20 ms time stamp are evidently caused by fourth-order diffracted ISs arriving at that time stamp, which can be seen in Fig. 3.

Specular ISs from fourth to sixth order provide gradually better responses especially between 30 ms and 90 ms. However, there are prominent differences at all frequencies to the reference. When diffraction is included, the timbre seems to be perfect to 40 ms with fourth-order ISs, and to 50 ms with fifth-order ISs. However, first fifth-order diffracted ISs appear already at 30 ms time stamp, and first sixth-order diffracted ISs at 35 ms time stamp which should be seen as deviations in difference plots. This simulation thus suggests that their influence would be inaudible in auralization.

Similar analysis was repeated to another listening position, in which the sound source was visible to the listener. The results were similar to the presented ones, although while the direct sound was visible, there was prominently less artifacts in timbre of low order specular ISs.

4. CONCLUSION

Physics-based auralization systems often render early sound and late reverberation separately. The starting time of the first outputs from the late reverberation is typically fixed, e.g., being 50 ms after direct sound. In such a case early sound before 50 ms time stamp is usually modeled with the image-source method. Typically, applied modeling scheme for early sound is to search image sources up to a certain order. However, the presented simulation results showed that with such modeling scheme a great number of high order reflections, appearing in same time with lower order reflections, are missed and this would change the perceived timbre of auralization.

Based on the presented analysis it can be suggested that for high-quality physics-based auralization many orders of reflections should be searched to guarantee that all possible reflection paths before a certain time stamp are found. Nevertheless, only those reflections arriving before this time stamp to the receiving position need to be auralized if separate late reverberation algorithm is applied. This finding is useful in designing and optimizing the room acoustic modeling and auralization systems.

5. ACKNOWLEDGMENTS

This work has been financed by the Helsinki Graduate School in Computer Science and the Graduate School of Electronics, Telecommunication, and Automation, of the Academy of Finland.

6. REFERENCES

[1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, “Creating interactive virtual acoustic environments,” *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[2] J.-M. Jot, “Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces,” *Multimedia Systems, Special Issue on Audio and Multimedia*, vol. 7, no. 1, pp. 55–69, 1999.

[3] U.P. Svensson and U.R. Kristiansen, “Computational modeling and simulation of acoustic spaces,” in *AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 15-17 2002, Accepted for publication.

[4] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[5] J. Borish, “Extension of the image model to arbitrary polyhedra,” *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.

[6] U.P. Svensson, R.I. Fred, and J. Vanderkooy, “Analytic secondary source model of edge diffraction impulse responses,” *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2331–2344, 1999.

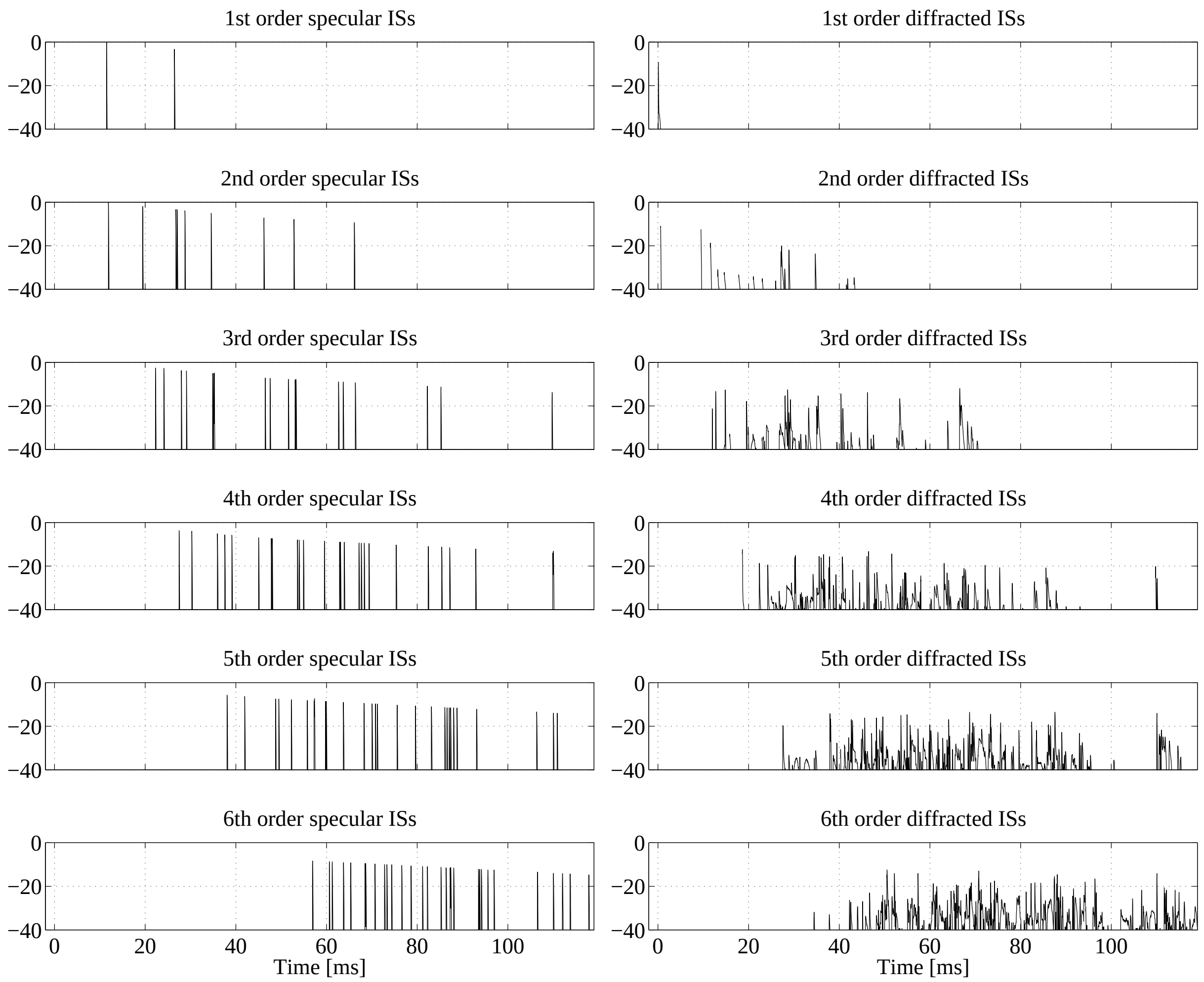

Figure 3: *Energy time curves in dB at receiver position R1. On the left column from top: 1st order to 6th order specular reflections. On the right column from top: 1st order to 6th order diffractions. The y-axis values are decibels*

[7] V. Pulkki, T. Lokki, and L. Savioja, "Implementation and visualization of edge diffraction with image-source method," in *the 112nd Audio Engineering Society (AES) Convention*, Munich, Germany, May 10-13 2002, preprint no. 5603.

[8] T. Lokki, U.P. Svensson, and L. Savioja, "An efficient auralization of edge diffraction," in *AES 21st Int. Conf. on Architectural Acoustics and Sound Reinforcement*, St. Petersburg, Russia, June 1-3 2002, Accepted for publication.

[9] T. Lokki and M. Karjalainen, "Analysis of room responses, motivated by auditory perception," *Journal of the New Music Research*, 2002, Accepted for publication.

[10] ISO Standard 226, *Acoustics – Normal equal-loudness level contours*, International Standards Organization, 1987.

[11] M. Slaney, "An efficient implementation of the Patterson – Holdsworth auditory filter bank," Tech. Rep. 35, Apple Computer, Inc., 1993, Available at: http://www.slaney.org/malcolm/apple/tr35/PattersonsEar.pdf.

[12] B.C.J. Moore, R.W. Peters, and B.R. Glasberg, "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 132–140, July 1990.

[13] B.C.J. Moore and B.R. Glasberg, "A revision of Zwicker's loudness model," *ACUSTICA united with acta acustica*, vol. 82, pp. 335–345, 1996.

[14] C.J. Plack and A.J. Oxenham, "Basilar-membrane nonlinearity and the growth of forward masking," *J. Acoust. Soc. Am.*, vol. 103, no. 3, pp. 1598–1608, Mar. 1998.

[15] American Standars Association, "USA standard acoustical terminology," 1960, S1.1-160, American Standars Association, New York.

[16] R. Pratt and P. Doak, "A subjective rating scale for timbre," *Journal of Sound and Vibration*, vol. 45, pp. 317–328, 1976.

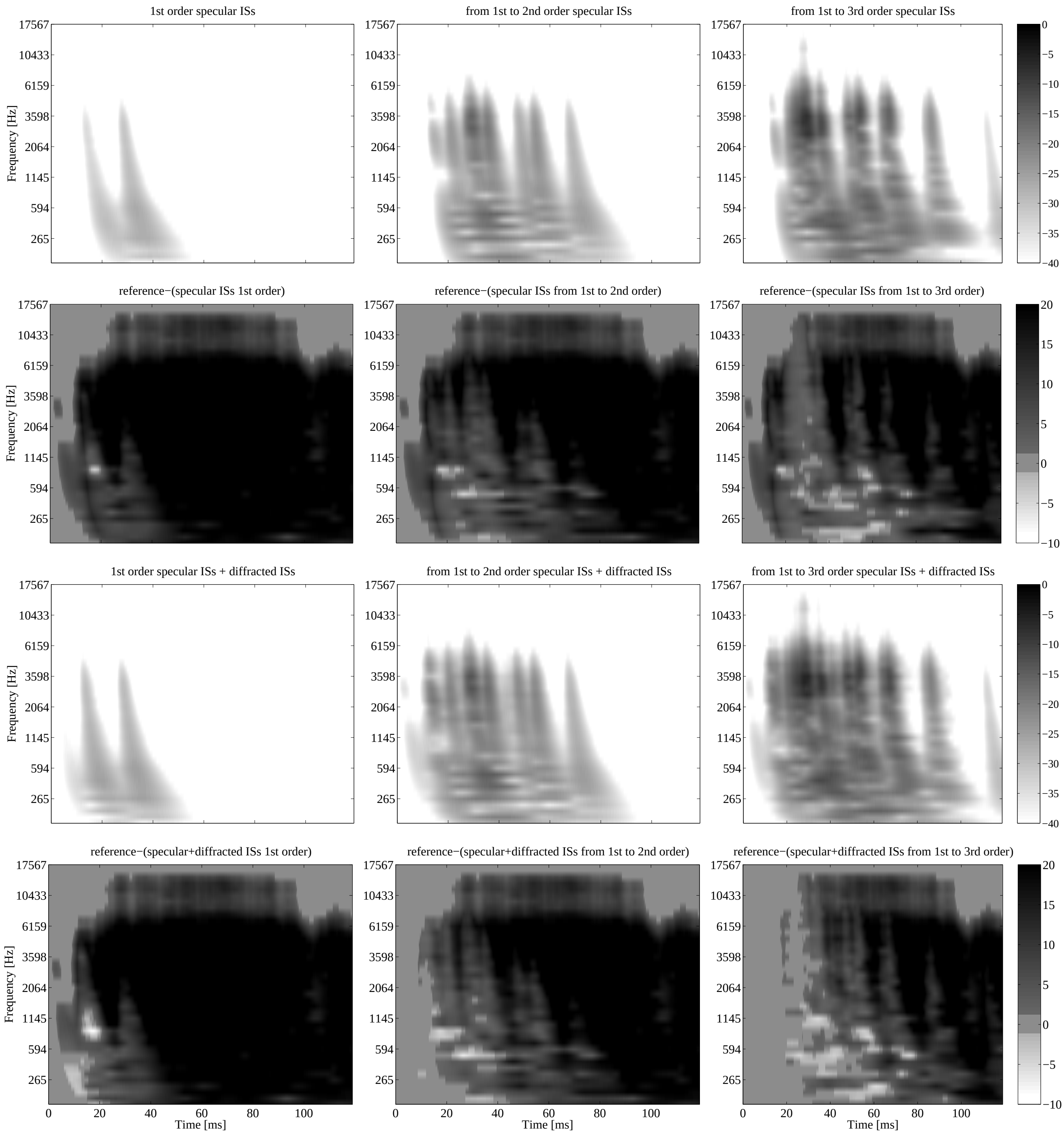

Figure 4: *Visualizations of predicted timbres with different orders of specular and diffracted image sources. Orders from one to three are visualized from left column to the right column correspondingly.*

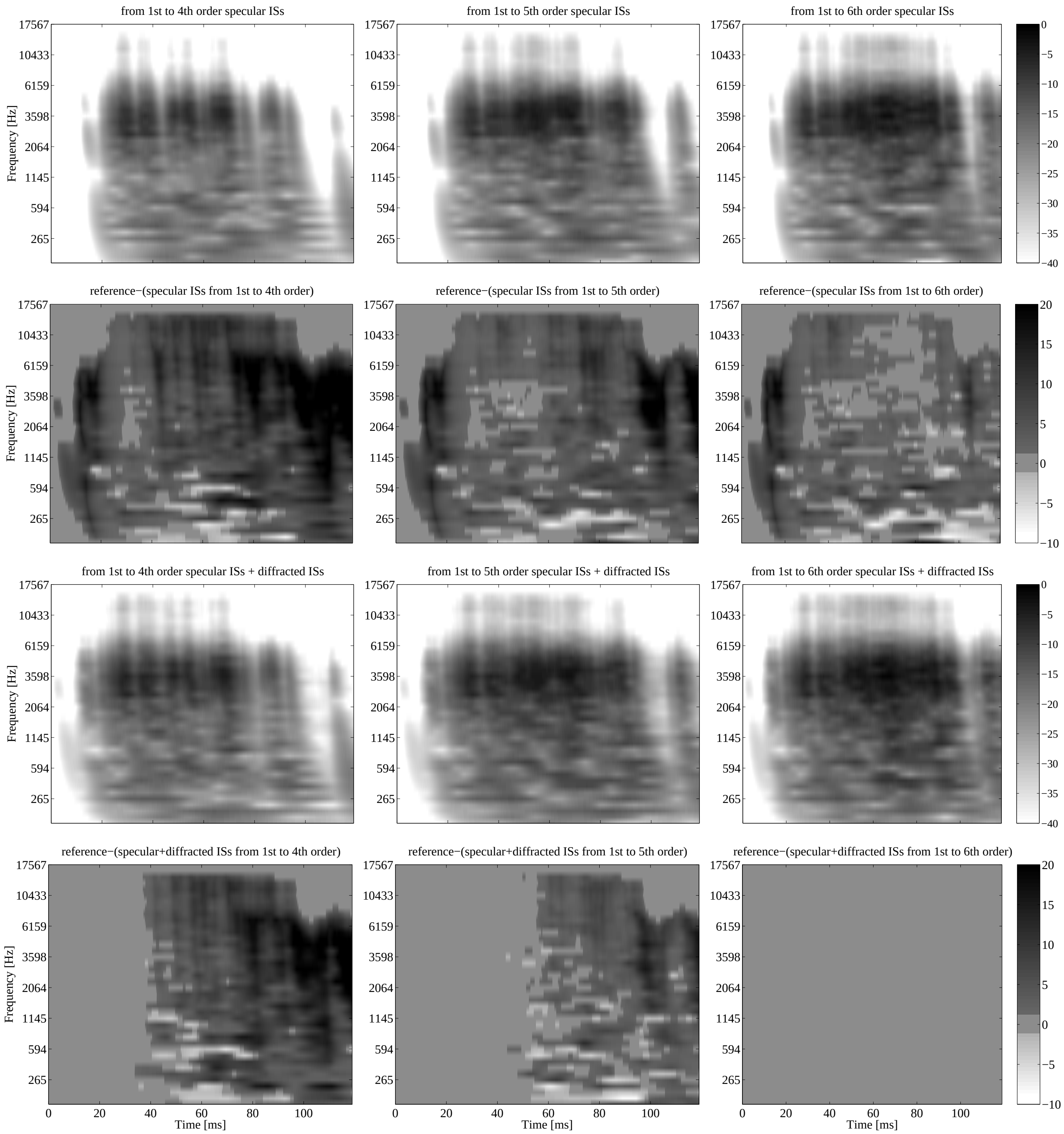

Figure 5: *Visualizations of predicted timbres with different orders of specular and diffracted image sources. Orders from four to six are visualized from left column to the right column correspondingly.*