Spatial Resolution of Early Reflection for Speech and White Noise

Xiaoli Zhong¹, Hao Song², Xuejie Liu³

¹School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China
²School of Management, Guangdong University of Technology, Guangzhou, China
³School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou, China

xuejie.liu@m.scnu.edu.cn; songhao@mail2.gdut.edu.cn

Abstract

In virtual auditory display, the accurate simulation of early reflection is helpful to guarantee audio fidelity and enhance immersion. However, the early reflection may not be easily distinguished from the direct sound due to the masking effect. This work investigated the spatial resolution of early reflection for speech and white noise under different conditions, in which three-down-one-up adaptive strategy with three-interval-three-alternative forced-choice (3I-3AFC) was employed. Results show that, for both speech and white noise, the spatial resolution of early reflection decreases with the increasing deviation of reflection orientation relative to the direct sound, and has no relationship with the time delay; Moreover, the spatial resolution of early reflection for speech is always lower than that for white noise under the same condition.

Index Terms: spatial resolution, early reflection, virtual auditory display, room acoustics

1. Introduction

In our daily life, sound fields are complicated with various spatial and temporal information of the direct sound, early reflection and late reverberation [1–2]. The early reflection is defined as the sound observed within 50ms–80ms time delay after the arrival of the direct sound [3–4], and it can influence distance localization, perceptual source width, speech intelligibility, as well as cause fluctuations in loudness, timbre and spatiality [5]. However, the early reflection is often masked by the direct sound and reverberation, making it hard to be detected by the listeners. Therefore, the spatial resolution of early reflection is introduced as the audible threshold that the reflection is just able to be perceived.

Some literatures have investigated the spatial resolution of early reflection [6–9]. Olive and Toole studied the relationships between the absolute spatial resolutions of early reflection and multiple experimental parameters, and found that the spatial resolution of early reflection was relevant to sound signal types [6]. Begault further adopted virtual auditory display technique to explore the spatial resolution of early reflection, and found that spatial resolution of early reflection had relationships with signal types, incidence angles, and the room masking effect [7]. In later work, Begault et al. used a room simulation to examine the effect of different time delays and incidence angles on the spatial resolution of reflections [8]. Moreover, the study of Grantham et al. reported the spectral information in the horizontal and vertical plane contributed to the spatial resolution of reflections independently [9]. In summary, the spatial resolution of early reflection varies with different conditions in a complicated way. Considering speech is a commonly-used stimulus in virtual auditory display, while white noise is often used in scientific research as a standard stimulus, this work aims to comprehensively measure the spatial resolution of early reflection for speech and white noise. Three-down-one-up adaptive strategy with three-interval-three-alternative forced-choice (3I-3AFC) was adopted as the experiment paradigm, and results were analyzed by the ANOVA method to evaluate the statistical significance.

2. Methods

In an enclosure, there are various temporal and spatial distribution patterns of the early reflection. Considering the spatial resolution of early reflection will decrease if multiple early reflections exist due to mutual masking [10], this work used a simplified sound field consisting of a single direct sound and a single early reflection. It represents the worst-case in which humans are most sensitive to the early reflection. This model has also been used in the research of the precedence effect [11–12] and simplification of the binaural response impulse response [13].

The transformed up-down adaptive method is recognized as a robust and effective way to evaluate resolution in psychoacoustic experiment design [14], in which the spatial resolution on each trial is determined by the preceding stimuli and response. In a specific experimental design, four factors, including choice of up-down strategy, choice of response paradigm, choice of initial value and step size of stimulus resolution, and choice of termination condition, need to be carefully set.

The up-down strategy determines the convergence point on a psychometric function. We used three-down-one-up adaptive strategy, which produces a resolution targeting 79.4% correct responses. In this strategy, the adaptive rule prescribes that three consecutive positive responses lead to an increase in the resolution of early reflection, whereas a negative response leads to a decrease.

On each trial, we used a random sequential presentation according to three-interval three-alternative forced-choice (3I-3AFC) paradigm for efficiency and robustness [15]. In 3I-3AFC, a stimulus presentation consisted of three segments, and each segment was chosen from either reference A or comparison B. Thus, there were totally three kinds of stimulus presentations: A-A-B, A-B-A, and B-A-A. In this case, the reference A contained a direct sound and an early reflection; the comparison B was the same to the reference A except the...
reflection orientation varied according to the step size of the three-down-one-up adaptive strategy, see fig. 1 for details. In
the experiment, the subjects were asked to judge which segment was different from the other two segments according
to whatever differences perceived, and then gave a response.

The orientation of the early reflection in the first trial is
termed as the initial value. To facilitate the subjects to make
positive responses, the initial reflection orientation in
comparison B was chosen to be 45° deviating from the
reflection orientation in reference A. On the other hand, the
step size is often changed from a high to a low value after a
certain number of trials, implying gradual convergence. In the
current work, the initial step size of the reflection orientation
in comparison B was set to be 10°, and reduced by half at each
reversal until the 2.5° step size was reached.

In the up-down adaptive method, a run refers to a
sequence of trials where the changes in reflection orientation
are all in one direction (“increasing” or “decreasing”); while a
reversal is the point where the direction of reflection orientation adjustment changes. As recommended in Ref. [16],
a test could be terminated after obtaining six to eight reversals.
In this work, termination reached after a total of eight
reversals (No. 4–8) was the spatial resolution obtained from this experimental block. Note that, if the
measurement is conducted along the horizontal plane, then the
above-mentioned angle adjustment refers to the azimuth adjustment; if the measurement is conducted along the median
plane, then the above-mentioned angle adjustment refers to the
elevation adjustment.

3. Apparatus and Procedure

This work used the head-centered coordinate system, in which the sound orientation was specified by azimuth θ from 0° to
360° and elevation from –90° to 90°. Here, (θ = 0°, φ = 0°) refers to the directly front and right of the
subject in the horizontal plane, respectively; (θ = 0°, φ = 90°) refers to the top of the subject in the median plane.

This experiment was implemented via headphone-
rendered virtual auditory technique [17–19], in which the
sound transmission from sound source to ears through direct
or reflective path were synthesized by filtering the mono stimulus with head-related transfer function (HRTF) at
intended orientation. Here, an HRTF dataset of KEMAR was
used because KEMAR is regarded as a representative manikin
and has been widely used in the research of binaural hearing
[20]. Moreover, speech (a segment of Chinese sentence “Mei
Tan Bu Mei”) and white noise were adopted as the mono
stimuli. In the experiment, the direct sound was always fixed
directly in front of the subject, while the early reflection
distributed at different spatial orientations (see Table 1). For
the white noise, the spatial resolution of early reflection was
measured not only in the horizontal plane (i.e., θ = 0°, 30°, 60°,
φ = 0°), but also in the median plane (i.e., θ = 0°, φ = 0°,
30°, 60°), respectively. However, as to the speech, the
measurement in the median plane was cancelled because of
the difficulty in localization reported by the subjects.

A pair of circumaural headphone (Sennheiser HD 250 II)
was used to render the synthesized binaural signals. In order to
eliminate the adverse influence caused by non-ideal
headphone transfer functions on reproduction performance,
headphone equalization was implemented [21]. The three-
down-one-up adaptive procedure with a 3I-3AFC paradigm
described in Sec. 2 was implemented through a graphical user
interface (GUI) created in MATLAB. For each trial, the
subject gave his or her response via pushing corresponding
button in the GUI interface.

Twelve subjects aging between 21–25 years participated in
the experiment. To guarantee experimental stability, the
subjects were exposed to an extensive training program before
the formal experiment, including a procedural training and an
auditory training aiming to familiarize the subjects with
experimental signals and procedure. In total, each subject
conducted 45 different experimental conditions, including 9
reflection orientations (see Table 1) and 5 reflection time
delays (from 10ms to 50ms with an interval of 10ms). Note
that, in each block, the number of trials varied with subjects
due to subject differences in auditory discrimination ability.
and experimental stability. For example, in Fig. 1, 32 trails were carried out till eight reversals reached. Generally, 30–40 trails were needed before terminating the test. This means that each subject responded to 1350–1800 stimulus presentations.

Table 1: Reflection orientations in reference A with one for each test condition

<table>
<thead>
<tr>
<th>Signal Type</th>
<th>Reflection Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal plane</td>
<td>$(\theta=0^\circ, \phi=0^\circ)$, $(\theta=30^\circ, \phi=0^\circ)$, $(\theta=60^\circ, \phi=0^\circ)$.</td>
</tr>
<tr>
<td>White noise</td>
<td>$(\theta=0^\circ, \phi=0^\circ)$, $(\theta=0^\circ, \phi=30^\circ)$, $(\theta=0^\circ, \phi=60^\circ)$.</td>
</tr>
<tr>
<td>Median plane</td>
<td>$(\theta=0^\circ, \phi=0^\circ)$, $(\theta=0^\circ, \phi=30^\circ)$, $(\theta=0^\circ, \phi=60^\circ)$.</td>
</tr>
<tr>
<td>Speech</td>
<td>Horizontal plane $(\theta=0^\circ, \phi=0^\circ)$, $(\theta=30^\circ, \phi=0^\circ)$, $(\theta=60^\circ, \phi=0^\circ)$.</td>
</tr>
</tbody>
</table>

4. Results

4.1. Orientation and time delay

The spatial resolution of early reflection was obtained by averaging over twelve subjects, see Fig. 2. According to Fig. 2 (a), the spatial resolution of early reflection for speech decreases when the reflection orientation gradually deviates from the direct sound $(\theta = 0^\circ, \phi = 0^\circ)$. Similar tendency can be observed in Fig. 2 (b) and (c) for white noise, except at $(\theta = 30^\circ, \phi = 0^\circ)$ and $(\theta = 60^\circ, \phi = 0^\circ)$ the spatial resolutions of early reflection are close at different time delays.

The spatial resolutions in different conditions were submitted to a multi-factor ANOVA to examine the influence from reflection orientation, time delay and their interaction. Table 2 shows that the effect of reflection orientation is significant for speech in the horizontal plane ($P<0.001$), white noise in the horizontal plane ($P<0.001$) and white noise in the median plane ($P<0.001$). This suggests that the reflection orientation has a significant impact on the spatial resolution of early reflection. Moreover, Table 2 shows that the time delay and the interaction between reflection orientation and time delay have no significant influence on the spatial resolution of early reflection ($P>0.05$).

Table 2: Results of multi-factor ANOVA.

<table>
<thead>
<tr>
<th>Data</th>
<th>Source</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech in the horizontal plane</td>
<td>Reflection orientation</td>
<td>137.380</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Time delay</td>
<td>0.302</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>Orientation * time delay</td>
<td>0.319</td>
<td>0.958</td>
</tr>
<tr>
<td>White noise in the horizontal plane</td>
<td>Reflection orientation</td>
<td>24.873</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Time delay</td>
<td>0.717</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>Orientation * time delay</td>
<td>0.802</td>
<td>0.602</td>
</tr>
<tr>
<td>White noise in the median plane</td>
<td>Reflection orientation</td>
<td>31.785</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Time delay</td>
<td>0.847</td>
<td>0.497</td>
</tr>
<tr>
<td></td>
<td>Orientation * time delay</td>
<td>0.129</td>
<td>0.998</td>
</tr>
</tbody>
</table>

Figure 2: The spatial resolution of early reflection, (a) for speech in the horizontal plane, (b) for white noise in the horizontal plane, (c) for white noise in the median plane.
4.2. Signal type

Figure 3 shows the differences in the spatial resolution of early reflection between speech and white noise across 5 time delays and 3 reflection orientations, in which the positive value means that the spatial resolution of white noise is higher than that of speech, and visa versa. According to the figure, in most cases, the white noise has a higher spatial resolution of early reflection compared with speech, except at ($\theta = 0^\circ$, $\phi = 0^\circ$) and ($\theta = 30^\circ$, $\phi = 0^\circ$) with time delay of 40ms. Repeated measures ANOVA was performed on the spatial resolution of early reflection between speech and white noise for 5 time delays and 3 reflection orientations. Results show that speech has significant lower spatial resolution of early reflection than that of white noise ($F=94.352, P<0.001$).

4.3. Spatial distribution

As to the white noise, the spatial resolution of early reflection was measured along both horizontal and median plane with the same angle interval relative to the direct sound (i.e., $0^\circ$, $30^\circ$, $60^\circ$), see Table 1. The corresponding differences in the spatial resolution of early reflection between the horizontal and median plane are shown in Fig. 4, where the positive value means, as to the same reflection deviation relative to the direct sound, the spatial resolution of early reflection in the horizontal plane is higher than that in the median plane. According to the figure, for all the time delays and reflection orientations, the spatial resolution of early reflection in the horizontal plane is higher than that in the median plane. The differences of spatial resolution of early reflection between horizontal and median plane is significant higher than that in the median plane ($F=95.190, P<0.001$).

5. Conclusions

This work measured and compared the spatial resolution of early reflection for speech and white noise by using a simplified model (one direct sound plus one early reflection). Results indicate that: (1) the reflection orientation has a significant influence on the spatial resolution of early reflection, while the time delay of reflection relative to the direct sound has not; (2) compared with speech, the white noise has a higher spatial resolution of early reflection in most cases; (3) as to the white noise, the spatial resolution of early reflection in the horizontal plane is always higher than that in the median plane. The conclusions of this work can serve as instructions to optimize algorithms in virtual auditory display and related applications.

6. Acknowledgements

The authors thank Wenyong Guo, Zhuowei Lai, and Ningming Dai for their help and participation in the experiment. This work is supported by the National Nature Science Foundation of China (No.11474103), and Guangdong Modern Vision-audio Information Engineering Technology Research Center (Guangzhou University, 2017.12-2019.12).

7. References


