Impacts of direction of arrival of direct sound and late reverberation on listening difficulty ratings of speech signals

Tatsuhiko Tanaka* and Makoto Otani†

Graduate School of Engineering, Kyoto University,
Kyoto daigaku-Katsura, Nishikyo-ku, Kyoto, 615–8540 Japan

(Received 21 July 2020, Accepted for publication 1 October 2020)

Keywords: Direction of arrival, Late reverberation, Listening difficulty

1. Introduction

Clarity of speech in an architectural space is important for people to communicate via acoustical information without misunderstanding.

Conventional evaluation of clarity of speech has been conducted using a monaural speech signal. However, such monaural speech signal does not include spatial characteristics observed in architectural spaces, such as direction of arrival of direct and reflected sounds. It has been reported that there is a certain spatial distribution in direction of arrival (DoA) of late reverberation in real sound fields [1], whereas sound fields in a room are often assumed to be perfectly diffuse when discussing late reverberation.

Morimoto *et al.* [2] proposed listening difficulty ratings as a subjective measure for evaluating the clarity of speech signals in reverberant spaces. Regarding the spatial characteristics of sound field, Takaoka *et al.* [3] reported that listening difficulty is not affected by interaural cross-correlations, thereby indicating that it is possible to evaluate the clarity of speech signals by using listening difficulty ratings for monaural speech signals. On the other hand, Pollack and Pickett [4] have reported that speech intelligibility in the presence of voice babbles of other speakers is improved by using stereophonic reproduction. Furthermore, Hawley *et al.* [5] have reported that speech intelligibility varies with changes in the direction of arrival of noise and speech. The latter two studies suggest that spatial characteristics of the sound field may affect the clarity of speech.

This study performed psychoacoustical experiments to investigate effects of the spatial characteristics of sound field in architectural space on listening difficulty of speech signal. We conduct two psychoacoustical experiments: one is to examine the effect of variation in DoA of direct sound with constant DoA of late reverberation and another is to examine the effect of variation in DoA of late reverberation with constant DoA of direct sound and early reflections.

2. Sound field simulation

In this study, speech signals as auditory stimuli were binaural signals generated from monaural speech signals and binaural room impulse responses (BRIRs) and were presented to participants through a set of headphones. This section

introduces a procedure for generation of auditory stimuli used in the psychoacoustical experiment.

All the audio signal was processed by MATLAB (Mathworks) with 48-kHz sampling frequency and 16-bits quantization. BRIRs were computationally simulated. Figure 1 demonstrates a temporal structure of room impulse responses and resulting BRIRs, assumed in this study. Namely, arrival time of direct sound (DS) was set to 0 ms. Then, early reflections (ER) arrived between 0 and 50 ms. Late reverberation (LR) arrived after 50 ms. The BRIRs were generated from a dataset of head-related transfer functions (HRTFs) [6], as described later. The length of the BRIRs was 10 seconds.

2.1. Reverberation time

The reverberation time (RT) measured in Große Musikvereinssaal in Vienna (Unoccupied) [7] and Notre Dame Cathedral in Paris [8] were employed for generating LR part of BRIR signals as depicted in Table 1. Both data (called “Wien” and “Notre,” respectively, hereafter) were available only in octave bands from 125 Hz to 4,000 Hz of center frequency. In the simulation, we added values of RT in the octave bands for 62.5 Hz and 8,000 Hz of center frequency in order to avoid deterioration of sound quality. One for 62.5 Hz was set equal to that of 125 Hz. One for 8,000 Hz was set to be half of that of 4,000 Hz.

The signal of LR was modelled by a random signal, i.e. white Gaussian noise with an attenuation applied to each octave band [9]. The white Gaussian noise would be suitable for modelling LR part assuming a diffuse sound field because it is able to emulate reflections in the LR part whose arrival times are random and its frequency characteristics are flat. The signal was generated as follows. First, 10 seconds of white noise was generated. Then octave bandpass filters with center frequencies of 62.5 Hz to 8,000 Hz were applied to the white noise, yielding eight octave band noises. Subsequently, attenuation corresponding to the RT at each octave band was applied to the octave band noises, resulting in LR signals in each octave band. Finally, they were summed up to yield an LR part of impulse response including all the octave bands. Therefore, the LR signal $h_{\text{rev}}(t)$ is expressed as

$$h_{\text{rev}}(t) = \sum_{i=1}^8 w_i(t) \sqrt{10^{-6t/T_i}} \quad (1)$$

where $w_i(t)$ is the noise signal at i -th octave band; T_i is the RT at i -th octave band ($i = 1, \dots, 8$).

Although the values of reverberation time for the concert

*e-mail: tanaka.tatsuhiko.38u@st.kyoto-u.ac.jp

†e-mail: otani@archi.kyoto-u.ac.jp

[doi:10.1250/ast.42.58]

Fig. 1 Diagram of room impulse response.

Table 1 Reverberation time (RT) of each octave band [7,8].

Freq. [Hz]		125	250	500	1,000	2,000	4,000
RT	Wien	2.97	3.03	3.06	3.05	2.67	2.10
	[s]	Notre	8.57	8.20	7.42	6.14	4.67
							3.20

hall and cathedral were generally too long and not suitable for listening speech, we employed them because such environments having higher energy of late reverberation would highlight their influences on the perception of speech.

2.2. Spatial characteristics

This study examined two aspects of spatial characteristics of room acoustics: variations in DoA of DS (Experiment 1) and DoA of LR (Experiment 2). For generating BRIRs used in both experiments, we employed HRTFs for 72 positions of sound source at every 5 azimuthal angles at 0-degrees elevation (in the horizontal plane), 1 m away from the center of the head, as illustrated in Fig. 2.

The LR part of BRIRs was modeled as a summation of HRTFs from 72 source positions, and its spatial distribution was characterized by weighting factors applied to each sound source, or HRTFs. The HRTFs were calculated by applying the boundary element method to the computer model of the dummy head (Brüel & Kjær, Head and Torso Simulator Type 4128 C) [6].

Fig. 2 HRTF settings.

2.2.1. DoA of direct sound

Five sound fields were assumed for BRIR calculation with variations in DoA of DS. First, referring Zahorik's method [9], the amplitude of the LR signal $h_{\text{rev}}(t)$ was adjusted so that the rms (root mean square) value at 0–50 ms of the impulse (DS) was equal to the rms value at 0–50 ms of the LR. Then, the signals from 0 to 50 ms of $h_{\text{rev}}(t)$ were replaced by zeros. Note that no ER component was included in the BRIRs used in Experiment 1. A DS arrived from one of five directions, that is 0, 5, 15, 45, and 90 degrees of azimuthal angle in the horizontal plane where 0 degrees corresponds to the frontal source and a positive value of azimuthal angle indicates anticlockwise direction, while LR was modeled as a set of sounds arriving coherently from all the directions without amplitude weighting. The RT of “Wien” was used for these five sets of BRIRs.

In addition to these five sets of BRIRs, we calculated BRIRs for two sound fields as reference stimuli, that is, BRIRs in an anechoic field (only DS) and those having the RT of “Notre.” These seven sets of BRIRs were used in Experiment 1.

2.2.2. DoA of late reverberation

Five sound fields were assumed for BRIR calculation with variations in DoA of LR. We used a BRIR measured at a multipurpose hall (approx. 4,886 m³) for the DS and ER (0–50 ms) of BRIRs. In the BRIR measurement, the distance between the dummy head and the sound source was about 8 m, and the dummy head was directed toward the sound source. The RT of these five sets of BRIRs was “Wien.” The DoA distribution of LR was characterized by multiplying the amplitude of the LR from each sound source by the weighting factors, which were determined from the normal distribution shown in Fig. 3. Here, the “main axis” of the DoA distribution, which corresponds to zero in the abscissa in Fig. 3, was set to 0, 5, 15, 45, and 90 degrees of azimuthal angle. The amplitude of the LR part of BRIRs for 0-degrees DoA distribution was adjusted so that the rms value of 40–50 ms of itself is equal to that of the BRIRs measured at the multipurpose hall, following the procedure in [9]. This amplitude adjustment rate was also applied to the BRIRs for other azimuthal angles.

In addition, we added two sets of BRIRs as reference stimuli: a sound field of only DS and ER and a set of BRIRs to which the RT of “Notre” was applied. These seven sets of BRIRs were used in Experiment 2.

Fig. 3 DoA distribution of late reverberation (LR).

Table 2 The scale of listening difficulty (in Japanese).

1	Not difficult
2	Slightly difficult
3	Difficult
4	Very difficult

3. Listening difficulty

A rating scale for listening difficulty was determined with reference to Morimoto *et al.* [2] and used in the psycho-acoustical experiment described in the following section. Participants evaluated the listening difficulty of the speech in four categories from (1) Not difficult to (4) Very difficult, as depicted in Table 2.

4. Experiment

Clarity of the speech signals convolved with the BRIRs were subjectively evaluated in the experiments by using the listening difficulty ratings described in the previous section.

Participants of the experiments were 10 male and female ranging from 22 to 26 years of age. None of them reported any hearing loss. The experiments were conducted in a semi-anechoic room. The experiments consisted of two sessions, labeled as Experiment 1 and Experiment 2, each of which is to investigate the effect of DoA of DS and LR, respectively. Total time required for these sessions was about 20 minutes. Speech signals, described later, were convolved with the BRIRs to generate auditory stimuli. All the audio signals were processed by MATLAB operating on a PC (OS: Windows 10), converted to analog signals via an audio interface (Roland, DUO-CAPTURE EX), and then presented to the participants via a set of headphones (Sennheiser, HD 598). Sound pressure levels of the auditory stimuli presented to the participants were 66 ± 2.5 dB (DS only, time constant: Fast, A-weighted, peak value) in Experiment 1 and 73 ± 3 dB (DS + ER only, time constant: Fast, A-weighted, peak value) in Experiment 2.

The participants were asked to listen to a three-word speech in a single trial, as illustrated in Fig. 4, then to dictate the speech they listened, and to rate the listening difficulty. Three words were randomly selected from words of the most familiar group in “Familiarity-controlled word lists 2003” (FW03) [10]. There were 28 trials in each session. Each of participants listened to seven sets of BRIRs four times in a random order, which was different among the participants. In addition, the combinations of three words were different among the participants and their trials. Seven practice trials including seven sets of BRIRs in a random order were conducted at the beginning of each session to familiarize the participants with the procedure of the experiments.

5. Results

One of the ten participants ambiguously corrected some responses, in other words, had multiple responses to one response field, so the participant’s responses were excluded from the results and discussions described hereafter. Statistical software JMP Pro 14 was used for data analysis. The significance level was set at 5%.

Fig. 4 Diagram of the auditory stimuli used in the experiments. **Fig. 5** Listening difficulty score of Experiment 1.

The evaluation of listening difficulty was carried out using integers from 1 to 4 in the experiments. The integers were converted to an interval scale by applying the method of successive categories [11].

5.1. Experiment 1

In Experiment 1, listening difficulty was evaluated for auditory stimuli with different DoA of DS. Figure 5 illustrates the results of Experiment 1. The error bar, which is not presented for the reference stimuli, indicates the standard deviation of the mean score for each participant. DS00 represents the sound field in which the direct sound arrives from the front, and DS90 represents the sound field in which the direct sound arrives from 90-degrees azimuthal angle, i.e. left. A repeated-measures analysis of variance (ANOVA) showed that the effect of sound field was not significant ($p = 0.3038$), indicating that the DoA of direct sound has no significant effect on listening difficulty.

5.2. Experiment 2

In Experiment 2, listening difficulty was evaluated for sound fields with different DoA distribution of the LR. Figure 6 depicts the results of Experiment 2. A repeated-measures ANOVA revealed that the effect of sound field was significant. Then Tukey’s HSD test was used to perform multiple comparisons of the population means of the listening difficulty scores among the sound fields. There was a significant difference in the DoA distribution of LR between the group of 0 (LR00), 5 (LR05), and 15 (LR15) and the group of 45 (LR45) and 90 (LR90), which thereby suggests that the DoA distribution of LR has a significant effect on listening difficulty. The lower ratings of listening difficulty for LR45 and LR90 may be explainable by spatial release from masking [12]: directional separation between a target speech and a masker improves word identification, while the masker corresponds to the late reverberation in the current study.

Fig. 6 Listening difficulty score of Experiment 2, * indicates a significant difference.

6. Conclusion

In this study, psychoacoustical experiments were conducted to investigate the influence of spatial characteristics, i.e. the DoA of direct sound and late reverberation, on speech intelligibility. The experimental results revealed that the DoA of direct sounds did not yield a significant difference in listening difficulty; however, the DoA of late reverberation led to a significant difference in listening difficulty, suggesting that DoA of late reverberation has an impact on listening difficulty.

We artificially created a situation where the reverberation time is very long and the DoA of late reverberation is strongly biased to a certain direction. Such a situation may occur between an indoor space with a long reverberation and an outdoor space, but in such a place, the speech intelligibility may not be a matter of concern. However, as shown in the reference [1], it is reported that certain biases occur in the DoA of late reverberation in room acoustics. For example, it often happens that the energy of the late reverberation arriving from the side walls, i.e. two directions, is dominant. In such a situation, it is possible that the similar perceptual phenomenon as in the current study. It will be future work to investigate whether the DoA of the late reverberation affects the speech intelligibility in environments with more realistic DoA distribution of late reverberation and/or shorter reverberation time.

Acknowledgement

A part of this research was supported by Grant-in-Aid for Scientific Research (Nos. 19H04153, 19H04145, and 17KT0137).

References

- [1] Y. Izumi and M. Otani, "Direction-of-arrival distribution analysis of reflected sounds using spherical microphone array," *Proc. Int. Symp. Room Acoust. 2019*, Amsterdam (2019).
- [2] M. Morimoto, H. Sato and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," *J. Acoust. Soc. Am.*, **116**, 1607–1613 (2004).
- [3] T. Takaoka, M. Morimoto, H. Sato and Y. Semba, "Effects of interaural cross-correlation of speech and background noise on listening difficulty," *J. Acoust. Soc. Jpn. (J)*, **63**, 520–528 (2007) (in Japanese).
- [4] I. Pollack and J. M. Pickett, "Stereophonic listening and speech intelligibility against voice babble," *J. Acoust. Soc. Am.*, **30**, 131–133 (1958).
- [5] M. L. Hawley, R. Y. Litovsky and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.*, **115**, 833–843 (2004).
- [6] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, **119**, 2589–2598 (2006).
- [7] L. Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, New York, 1996), p. 600.
- [8] B. Postma and B. Katz, "Acoustics of Notre-Dame cathedral de Paris," *Proc. 22nd Int. Congr. Acoust.*, Buenos Aires (2016).
- [9] P. Zahorik, "Perceptually relevant parameters for virtual listening simulation of small room acoustics," *J. Acoust. Soc. Am.*, **126**, 776–791 (2009).
- [10] S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo and T. Sone, "New lists for word intelligibility test based on word familiarity and phonetic balance," *J. Acoust. Soc. Jpn. (J)*, **54**, 842–849 (1998) (in Japanese).
- [11] Y. Tanaka, *Psychometric Methods*, 2nd ed. (University of Tokyo Press, Tokyo, 1977), pp. 156–158 (in Japanese).
- [12] R. Y. Litovsky, "Spatial release from masking," *Acoust. Today*, **8**(2), pp. 18–25 (2012).