Forum Acusticum 2011 - Aalborg - Denmark

Reverberation Perception with One and Two Ears

Ville Sivonen[1], Mikko Alanko, Hannes Gamper, Henri Kari, Olli Rummukainen and Ville Pulkki

Department of Signal Processing and Acoustics, Aalto University School of Science and Technology, Espoo, Finland.

[1] Now at Cochlear Nordic AB, Vantaa, Finland.

Summary

Binaural hearing is beneficial in suppressing the audibility of echoes in reverberant environments. Echo suppression has typically been investigated utilising a single reflection, and often at high direct-to-reverberant energy ratios (DRRs) where the reflection is barely audible. In natural environments, however, reverberation comprises a large number of reflections and DRR may be relatively low. The aim of this investigation was to quantify binaural echo suppression in the presence of reverberation by assessing its salience in monaural and binaural listening. A short speech segment and an impulsive finger snap were used as direct sound, and a simple reverberation model with frequency-dependent T_{60} between 0.5 and 2.0 s was used to create stimuli with different DRRs. The stimuli were spatialised using non-individual head-related transfer functions and played back over headphones. In an adaptive matching paradigm, a diotic or dichotic reference sound was fixed at 0 dB DRR, and the DRR of a monaural or diotic comparison sound was varied via a two-alternative forced-choice procedure. Twelve listeners participated, and on each trial, their task was to judge which of the two sounds, the reference or the comparison played back in random order, was more reverberant. The results show that between 2-6 dB higher DRR is needed for a monaural sound to be matched in perceived reverberation with the binaural sound, the effect being largest for the impulsive sound and when matched against the dichotic reference. This is in agreement with earlier studies on echo suppression of a single reflection.

PACS no. 43.66.Pn, 43.66.Rq

1. Introduction

The binaural auditory system offers many advantages compared to monaural listening. One of the most important advantages is better localization of sound due to interaural time and level differences (ITD and ILD, respectively), which are absent in monaural listening [1]. Binaural hearing is also more effective over monaural hearing in the transduction of acoustic signals to auditory percepts, due to binaural summation of loudness [2]. Binaural cues further allow the auditory system to enhance the perception of a sound signal in the presence of noise and competing sound sources [3], and to bolster the unreflected signal components in a reverberant environment [1].

In most natural enviroments, the sounds that we hear are accompanied by delayed reflections from many different directions. However, these reflections often go unnoticed [4]. This suppression of reverberation has been suggested to be largely due to binaural processing, although monaural mechanisms for echo suppression have also been proposed recently [5].

One of the earliest reports on binaural suppression of reverberation was published by Koenig in 1950 [6]. He listened to the signals of an acoustic reception system consisting of two microphones, placed at the ears of a manikin that provided signals to the observer's earphones. Switching between one-channel and two-channel (binaural) listening led Koenig to the conclusion that in binaural listening, reverberation and background noises were much less detectable. This informal observation was corroborated by Danilenko (as summarised in [1]), demonstrating that the perception of modulation of a signal in a reverberant room is more accurate binaurally than monaurally, while there is no difference in performance in an anechoic environment.

In order to quantify binaural suppression of reflected sound, diotic and dichotic echo thresholds were later measured by Zurek [7] with different echo delays. His hypothesis was that if the binaural system suppresses the effects of echoes, one might expect that the detection of delayed sound will be poorer in the dichotic case, in which there are ITDs and ILDs associated with the echo, than in the diotic case, in which the stimulus is identical at the two ears. The results showed that with echo delays of less than 5-10 ms, the thresholds for the diotic echo were about 10 dB lower than for the dichotic signal, in agreement with Zurek's hypothesis.

As opposed to a single reflection at a level where an echo is barely audible, natural reverberation comprises sound that has reflected a number of times from obstacles, and the acoustic energy of reverberation might be comparable to that of the direct sound. Thus, in this paper, the salience of reverberation is investigated at around 0 dB direct-to-reverberant energy ratio (DRR) by matching monaural and binaural listening in perceived reverberation via an adaptive paradigm.

2. Method

2.1. Participants

Twelve listeners (one female, 11 male) participated in the experiment. All listeners were recruited from an acoustics course held at Aalto University, and none of the listeners reported any known hearing problems. In addition, five listeners naïve to the task were recruited to an informal pilot study to test the functionality of the paradigm in matching different conditions for perceived reverberation. The listeners of the pilot study did not participate in the actual test.

2.2. Stimuli

Both a continuous and an impulsive sound sample were used as stimuli in the experiment. The samples were generated from anechoic recordings of continuous speech and a finger snap. The length of the speech sample (English phrase: *the number eight-one-nine*) and the finger snap was 1.5 and 0.03 s without reverberation, respectively, and A-weighting was applied to align them in level.

A simple reverberation model [8] was used to add reverberant energy to the stimuli. Twelve virtual loudspeakers were spaced 30° apart in the horizontal plane and direct sound was always reproduced with the loudspeaker at 0° azimuth. Reverberant sound was reproduced with all loudspeakers and it was generated from white noise, by using an independent noise sample for each loudspeaker. The noise samples were passed through an octave filter bank and different reverberation times (T_{60}) were assigned to each frequency band, decreasing towards high frequencies. For the speech sample, T_{60} was from 2 s at low frequencies to 0.5 s at high frequencies and for the finger snap it was from 1.5 s to 0.5 s, respectively. Reverberation for both stimuli was obtained by convolving them with the noise samples and applying an temporal envelope with a linear attack and an exponential decay. Finally, the direct and the reverberant sounds were spatialised by convolving each sound with head-related impulse responses (HRIRs) measured for an artificial head (Cortex Manikin Mk2) corresponding to the directions of the virtual loudspeakers. The HRIRs were prepared such that they were perfectly symmetrical across the two ears. The resulting sounds from the computations were then mixed to obtain various DRRs.

In order to prevent listeners to base their judgment merely on the length of the decay in reverberation, the reverberation tail was removed from the speech sample. The tail was removed by applying a raised-cosine ramp of 2 ms in length to the sample, resulting in a segment of reverberant speech but without an audible decay after the speech segment. For the impulsive finger snap, the reverberation tail was held intact. Finally, the stimuli were presented to the listeners at a comfortable listening level over headphones.

2.3. Apparatus

The participants were seated inside an anechoic chamber for the listening test. The stimuli were played back from a laptop (MacBook Pro) via a high-quality audio interface (MOTU Traveler Mk3) and open-back headphones (Sennheiser HD-595). The participants' responses were collected using a graphical user interface with two buttons. All signal processing for preparing the stimuli, programming the experiment and analysing the results was done using MatLab.

2.4. Procedure

An adaptive, two-interval, two-alternative, 1-up/1-down forced-choice procedure [9] was used in obtaining matches for perceived reverberation. On each trial, the participants were presented with a pair of sounds, the reference and the comparison in random order, and their task was to judge which of the two sounds was more reverberant. The DRR of the reference was kept constant at 0 dB DRR and the DRR of the comparison sound was altered via the adaptive procedure. If the comparison sound was perceived as more reverberant, its DRR was increased to make it less reverberant in the next trial. Vice versa, if the reference sound was perceived to be more reverberant in a trial, the DRR of the comparison sound was decreased. Two starting levels, +10 and -10 dB DRR, were used. The initial step size for the DRR was 4 dB, and after two reversals in the track it was decreased to 2 dB. Eight reversals were obtained for each adaptive track and the point of subjective equality (PSE) was computed as a median of the last four reversals of the track.

Matches for perceived reverberation were obtained for both sound samples between four different playback modes: In the *dichotic* mode, direct and reberberant sound was delivered to the listeners as such from the reverberation model, while in the *diotic*

Table I. Different test conditions listed by comparison vs. reference playback mode, sound sample type and the direct-to-reverberant ratio starting level.

Comparison vs. reference	Sample	Level
Monaural reverb vs. Dichotic	Speech	+10 dB
		−10 dB
	Finger snap	+10 dB
		−10 dB
Monaural vs. Dichotic	Speech	+10 dB
		−10 dB
	Finger snap	+10 dB
		−10 dB
Monaural vs. Diotic	Speech	+10 dB
		−10 dB
	Finger snap	+10 dB
		−10 dB
Diotic vs. Dichotic	Speech	+10 dB
		−10 dB
	Finger snap	+10 dB
		−10 dB

mode, the signal at one ear from the model was delivered to both ears. In the *monaural reverb* mode, direct sound was delivered to both ears, while all reverberant energy was summed and delivered to one ear only. Finally, in the *monaural* mode, the other earphone was muted, and the level of the monaural mode was increased by 6 dB to make it approximately equal in loudness with the samples delivered to both ears [10]. For half of the monaural cases the left-ear, and for the other half, the right-ear signal was used. The test conditions of the experiment are summarised in Table I with the comparison and reference playback mode to be matched via the adaptive procedure, as well with the sound sample and DRR starting level for each adaptive track.

The adaptive tracks were further organized into four blocks: one block included tracks with the monaural mode as a comparison sound and others included only diotic or dichotic sounds. This was adopted based on the feedback from pilot testing, where the listeners reported being confused with sound coming randomly to one or two ears. Placing the monaural tracks in a separate block aimed to reducing the confusing effect of different playback modes. Thus, each block consisted of one playback mode, with a track each for both sound samples (speech and finger snap) and with +10 dB and -10 dB DRR starting levels, totalling to four tracks per block. Within one block, the adaptive tracks were interleaved randomly such that the probability of choosing each track was based on the number of remaining reversals in the track.

Each block was repeated once resulting in a total of eight blocks and the presentation order of the blocks was governed by a 4x4 Latin square [11] resulting in a fully randomized testing procedure. Prior to the experiment, the listeners had one block of training (5–7 mins) to get acquainted with the test procedure. In the experiment proper, the listeners had a chance to have a break after each block and a mandatory longer break was held after four blocks. The total duration of the experiment was approximately 70 mins.

3. Results and Discussion

The hypothesis of the experiment was that reverberation is more salient in monaural than in binaural listening, and a higher DRR for monaural presentation would thus be required for a match in perceived reverberation with binaural presentation. For each of the conditions, the PSE was calculated as the median of the last four reversals obtained from each listener in each adaptive track. Given the noisiness of the data and the presence of outliers, the median was chosen as a robust measure of central tendency.

To test the data against the hypotheses, a Wilcoxon signed rank test was performed for each test condition. Data for all conditions was tested against the null hypothesis that it comes from a distribution with zero median, i.e., the median PSE of the reference and the comparison condition is zero. To increase the power of the statistical tests, data were collapsed across listeners and repetitions. Since no substantial differences between the two starting levels, i.e. +10 dB and −10 dB DRR were found in pilot testing, data from different starting levels were considered a repetition and thus collapsed as well. Bonferroni correction was applied to all p-values to account for multiple testing.

Figure 1 shows box plots of the data obtained from each test condition. A Wilcoxon signed rank test shows that the null hypothesis, i.e., that the data of each condition stem from a distribution with zero median, can be rejected at the 5% significance level for all tested conditions, except for the speech playback in the monaural reverb and in the diotic vs. dichotic condition. The analysis of the monaural playback condition suggests that reverberation is more salient if listened to with one ear only and compared to dichotic playback. For both speech and finger snap sounds, the interquartile range of the PSE data lies above 0 dB DRR. When comparing monaural and diotic playback to one another, the statistical analysis and the distribution of the data indicate that reverberation is also more salient in monaural than in diotic playback, although the effect size is smaller than in the case of the dichotic reference. This finding is in agreement with the observations by Koenig [6] on differences in perceived reverberation in monaural and binaural listening.

However, diotic vs. dichotic playback resulted in a median PSE fairly close to 0 dB DRR and the effect size is considerably smaller than what was demonstrated by Zurek [7] on differences in echo thresholds between diotic and dichotic presentation (up to 10 dB). Furthermore, although the null hypothesis may

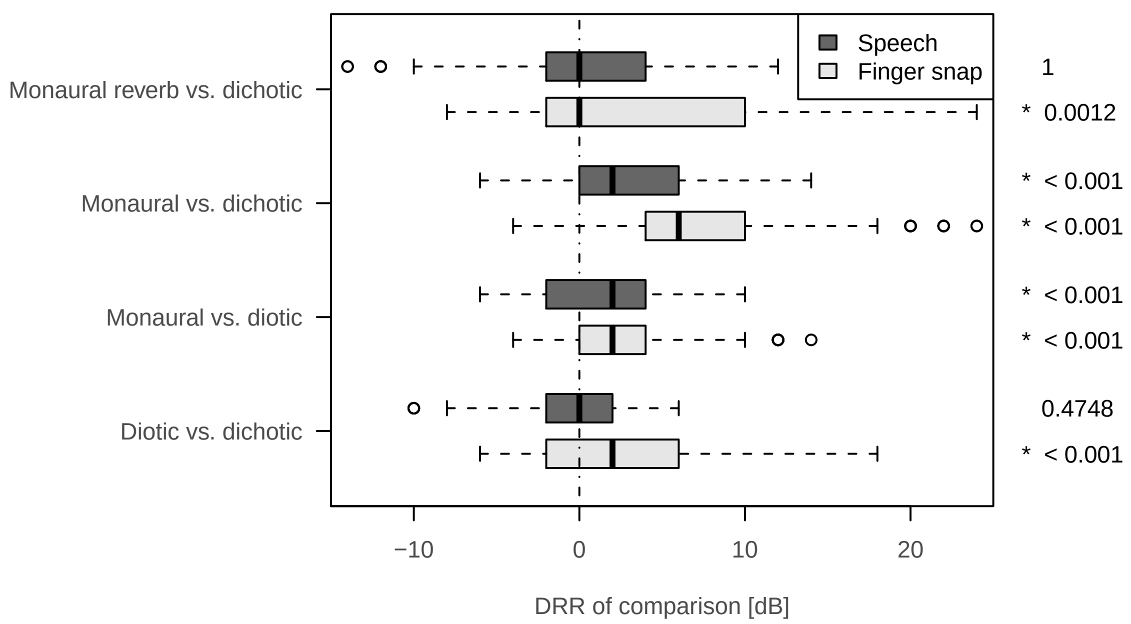

Figure 1. Box plot of matches (point of subjective equality; PSE) for perceived reverberation in various conditions. The conditions (comparison vs. reference) and their p-values of a Wilcoxon signed rank test for zero median PSE are shown on the left and right sides of each box plot, respectively. The direct-to-reverberant ratio (DRR) on the ordinate denotes the DRR required for the comparison to be matched in perveiced reverberation with the reference at 0 dB DRR.

Table II. Mean and median PSEs of perceived reverberation in various conditions. A Wilcoxon signed-rank test indicates significant differences of the median PSE between speech and finger snap sounds in all conditions except for monaural reverb.

Condition	**Monaural reverb**		**Monaural**		**Monaural vs. diotic**		**Diotic vs. dichotic**	
Sound type	Speech	Finger snap	Speech	Finger snap	Speech	Finger snap	Speech	Finger snap
Mean PSE (dB)	-2.0	3.6	1.7	7.5	1.3	2.7	-1.0	2.6
Median PSE (dB)	0	0	2	6	2	2	0	2
p-value	1		* < 0.001		* 0.033		* < 0.001	

be rejected for the finger snap sound in the monaural reverb and in the diotic vs. dichotic conditions, the distribution of the data indicates that the PSE is not substantially different from 0 dB DRR. Thus, the data do not support the hypothesis that reverberation is more salient if the reverb is summed and presented to one ear only (monaural reverb), or if the same signal is played back to both ears (diotic), when compared to dichotic playback. Informal interviews with the participants of the present study revealed that the monaural reverb condition was considered unnatural and difficult to compare with dichotic playback. Therefore, some listeners may have completed the monaural reverb test runs partly by issuing random answers, which could explain the large variance in the data and the presence of outliers.

To compare the median PSEs of the two different sound types, a test against the null hypothesis that the data for the speech sound come from a distribution with a median equal to the data for the finger snap sound was performed. The mean and median PSE values for each condition are summarised in Table II. A Wilcoxon signed rank test indicates significant differences of the median PSE between speech and finger snap sounds for all conditions except for monaural reverb. Although the difference is significant for the diotic vs. monaural and the diotic vs. dichotic conditions, it is not found to be substantial. In the monaural condition, however, speech yielded a PSE of 2 dB DRR, compared to 6 dB DRR with the finger snap sound (see Table II). This indicates that reverberation is more salient in monaural playback with a short impulsive sound than with a continuous complex sound. The reason for this might lie in the presence of a reverberation tail in the finger snap samples, whilst the reverberation tail was removed from the speech samples. However, a more thorough analysis and more experimental data would be necessary to investigate this further.

In this study, the suppression reverberation, as opposed to a single echo, was investigated. Therefore, a direct comparison to e.g., by Zurek [7], may not be warranted. In [7], the largests effects between diotic and dichotic presentation were seen for echo delays of less than 10 ms. Furthermore, monaural echo suppression has been shown to have the largest effect for delays on the order of 1-2 ms. It is thus unlikely that the present experimental paradigm is able to tap into these suppression mechanisms due to more complex, decaying reverberation present in all conditions. The findings of this study on quantifying the difference in DRR in various playback modes probably bear more resemblence to the observations by Koenig [6] on the salience of reverberation in monaural and binaural listening.

Understanding the perceptual effects of reverberation is important in many acoustical applications. In telecommunications, sound may be delivered to one or two ears and if the sound to be delivered is contaminated by large amounts of reverberant energy, the playback mode (monaural vs. binaural) may impose different requirements for algorithms to suppress reverberation. In binaural modeling of the loudness percept [2, 10], it is also important to know whether direct and reflected sounds need to be weighted differently and how they may be perceptually parsed by the binaural auditory system [12]. Finally, knowledge of reverberation perception is essential in the development of auditory models based on physiological processes [13], as well as in an attempt to improve the performance of auditory models in more complex, realistic acoustic environments. Future work is needed to uncover and quantify binaural suppression of reverberation with a larger range of stimuli and experimental conditions.

4. Conclusion

The salience of reverberation with various playback conditions and two sound types was investigated in this paper. Reverberation was generated with a simple computational model and matches in perceived reverberation between monaural, diotic and dichotic playback were obtained from listeners by varying the DRR in a listening experiment utilising an adaptive procedure.

The results show that reverberation is more salient in monaural (and partly in diotic) listening when compared to dichotic presentation, the effect being quantifiable and largest for an impulsive sound. At maximum, 6 dB higher DRR was required for monaural presentation to be equal in perceived reverberation with dichotic playback. The results are in agreement with earlier studies of binaural suppression of single echoes.

References

[1] J. Blauert: Spatial Hearing – The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA, 2001.

[2] V. Sivonen, W. Ellermeier: Binaural loudness - In: Springer Handbook of Auditory Research – Loudness, M. Florentine, A. N. Popper, R. R. Fay (eds.). Springer, New York, 2011.

[3] A. W. Bronkhorst: The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acta Acust. United Ac. **86** (2000) 117–128.

[4] W.G. Gardner: Reverberation algorithms - In: Applications of Digital Signal Processing to Audio and Acoustics, M. Kahrs, K. Brandenburg (eds.). Springer, New York, 1998.

[5] M. Bürck, J. L. van Hemmen: Modeling the cochlear nucleus: A site for monaural echo suppression?. J. Acoust. Soc. Am. **122** (2007) 2226–2235.

[6] W. Koenig: Subjective effects in binaural hearing. J. Acoust. Soc. Am. **22** (1950) 61–62.

[7] P. M. Zurek: Measurements of binaural echo suppression. J. Acoust. Soc. Am. **66** (1979) 1750–1757.

[8] M. Takanen, A. Jylhä, T. Pihlajamäki, J. Holm, I. Huhtakallio, V. Pulkki: Localization of consecutive sound events in reverberant environment. Proc. AES 126^{th} Int. Conv., Munich, Germany, May 7–10, 2009.

[9] H. Levitt: Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. **49** (1971) 467–477.

[10] B.C.J. Moore, B.R. Glasberg: Modeling binaural loudness. J. Acoust. Soc. Am **121** (2007) 1604–1612.

[11] N. Rapanos: Latin squares and their partial transversals - In: Harvard College Mathematics Review, S.D. Kominers (ed.) Harvard University, 2008, Vol. 2, 4–12.

[12] G.C. Stecker, E.R. Hafter: An effect of temporal asymmetry on loudness. J. Acoust. Soc. Am. **107** (2000) 3358–3368.

[13] V. Pulkki, T. Hirvonen: Functional count-comparison model for binaural decoding. Acta Acust. United Ac. **95** (2009) 883–900.