Applied Acoustics 110 (2016) 119–127

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A perceptual space for underwater man-made sounds towards target classification

Lixue Yang *, Kean Chen

Department of Environmental Engineering, School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 23 March 2016
Accepted 23 March 2016
Available online 29 March 2016

Keywords:
Underwater man-made targets
Dissimilarity evaluation
Multidimensional scaling
Perceptual space
Auditory features
Target classification

ABSTRACT

This work is the second part in a series of studies about the auditory features for underwater target classification, focusing on man-made vehicle targets (i.e. submarines, patrol boats and large surface ships). A psychoacoustic method, which is suitable for a small number of samples, was used. An optimal model with three common dimensions, specificities and latent classes was selected on the basis of the dissimilarity ratings among representative sounds and with the use of an extended version of the multidimensional scaling algorithm CLASCAL. However, such a three-dimensional space could not absolutely separate targets, whereas the first dimension in the four-dimensional space discriminated the submarines, patrol boats and ships; thus, the four-dimensional space was superior in target classification. The stepwise regression method was used to establish the relationships between individual dimensions and typical auditory features. Results showed that the first dimension was represented by the linear combination of *zero-crossing rate* and *spectral variation*, whereas the second dimension was described by *attack slope*. The last two dimensions were not associated with any features, and they were proved to include meaningless data noises. Finally, through a contrastive analysis, the perceptual space obtained in this study was found to be a good 'local' representation of the space in the first part of the study series.

1. Introduction

In modern naval warfare, the performances of sonar operators in underwater target classification are still better than those of automatic systems. The advantages of human beings may lie in the excellent anti-interference and features of their auditory system [1,2]. This study focuses on the latter, i.e. obtaining the auditory features used by human beings. The acoustic characteristics of underwater targets are well known to be complex and diverse; thus, no unique feature vector can discriminate all targets. According to a general theory on sound-source recognition [3], when an unknown target is classified, it is attributed to various categories at different levels of abstraction by human beings step by step, i.e. from an abstract macro-category to a specific micro-category, and the cues used become increasingly fine. Therefore, this series of studies investigates the hierarchical categories used by human beings to classify underwater targets, explores the variation patterns in the auditory features used at different levels and finally lays a foundation for obtaining effective and robust features.

A psychoacoustic method can be adopted at each level. Firstly, the dissimilarity ratings among representative underwater sounds are obtained via a paired comparison experiment, assuming that the sounds perceived as being similar are from the same target category. Secondly, the multidimensional scaling (MDS) technology is employed to represent the perceptual relationships among the sounds in a Euclidean space. Finally, the acoustic features that the dimensions may represent are easy to find, often by correlating the ordering of sounds along each dimension with an acoustic feature. The features obtained can be directly used in automatic classification if the dimensions can arrange different target categories in order. Compared with traditional machine learning methods, this psychoacoustic method usually needs a small number of sounds, because the amount of paired comparison among these sounds is proportional to the square of this number. A large dataset will bring a heavy task for subjects, and then cause subjects' fatigue and inconsistent results.

This method has been widely used underlying musical timbre properties [4–7], and its generalisation capacity has further improved owing to the sophistication of MDS algorithms. The CLASCAL method is the most accurate MDS algorithm. It includes two other properties: the existence of additional dimensions that are specific to individual sounds (called 'specificities') and the

* Corresponding author at: No. 127, Youyi West Road, Xi'an, Shaanxi 710072, PR China.

E-mail address: yanglixue.2008@163.com (L. Yang).

http://dx.doi.org/10.1016/j.apacoust.2016.03.036
0003-682X/

differences in the perceptual importance of each dimension between subpopulations of listeners (called 'latent classes').

This study is the second part in a series of studies. In part I [8], a paired comparison experiment was conducted on a dataset composed of 20 underwater sounds from man-made (vehicles) targets (e.g. surface ships and submarines) and natural targets (e.g. cetaceans, fish, shrimps and water-based background targets) to derive the perceptual categories of underwater targets at the first level. The obtained perceptual space provided a coarse description of underwater target classification, in which man-made targets were separated from natural targets. However, the fine perceptual structures within the man-made targets, particularly between the large surface ships and submarines, were not clearly displayed in this space. Thus, a further study on man-made targets is necessary.

The perceptual structures among man-made underwater sounds have been investigated in several studies. Soloman [9] used the semantic differential method to investigate the timbre of 20 sounds from 4 categories (i.e. submarines, light craft, naval ships and cargo ships) and found 7 orthogonal dimensions using the factor analysis method. A subsequent study [10] identified the auditory cues used by subjects with the use of cluster analysis. Mackie et al. [11] used the same method as the one used in the current study. The sounds in the stimulus set came from 14 man-made targets, including 3 submarines, 3 naval ships, 3 cargo ships, 2 light craft and 3 unidentifiable targets from [12]. The obtained five dimensions were *beat clarity*, *beat tonality*, *squeaky beats*, *beat rate* and *dual beats*. However, these auditory cues cannot be used in automatic classification directly because they have not been quantified well. This problem is addressed in the current study.

In this study, we aim to obtain the perceptual categories of man-made targets and to identify the auditory features for classification. Firstly, a paired comparison experiment was conducted on a carefully selected stimulus set, and the dissimilarity data were analysed with the CLASCAL method to obtain the perceptual space. Then, the optimal dimensionality was selected considering both the goodness of data fit and the ability of the space to classify different targets. Some typical auditory features were also derived to provide the quantitative description of each individual dimension using the stepwise regression method. Finally, the obtained perceptual space was compared with the perceptual space in part I [8] to examine the variation patterns between the 'local' and the 'global' perceptual properties.

2. Review of part I

In this section, we provide a brief review of part I [8] to explain the background of the current study and facilitate the comparative analysis between these two studies.

2.1. Methods

2.1.1. Stimuli

To obtain primary descriptions of underwater target categories, a representative and comprehensive stimulus set was selected. The sounds were divided into two main categories, i.e. man-made targets and natural targets, and each category comprised 10 exemplars. The man-made targets included four submarines, four large surface ships and two patrol boats, and the natural targets consisted of one fish, one croaker, four cetaceans and four water-based background targets. Based on the auditory sensations and intuitive observations of the spectrograms (not presented here), both intra-class and inter-class variations in the acoustic features were high; this result was in accordance with real underwater environments. Obtaining features (used by human beings) that only underlie inherent inter-class variations was preferred to achieve robust automatic recognition.

2.1.2. Subjects

Twenty-four undergraduates and graduates participated in the experiment. Three had rich experiences in similar experiments, whereas the others had no prior experience.

2.1.3. Procedure

In the experiment, the subjects were asked to provide dissimilarity ratings between 210 sound pairs, which consisted of 190 pairs of different sounds (A–B) and 20 pairs of identical sounds (A–A). A seven-point scale was adopted, with 1 representing 'extremely similar' and 7 representing 'extremely dissimilar'.

2.2. Results

The CLASCAL method (Section 4.2) was used to analyse the dissimilarity data, and an optimal model with three common dimensions, specificities and three latent classes was chosen.

2.2.1. Common dimensions (D1–D3)

The coordinate distribution of the sounds in each individual dimension was examined, and the gammatone filterbank [13] was used to derive the auditory features they represented.

The common space had three dimensions, which were labelled as D1–D3. D1 discriminated natural targets from man-made targets, i.e. the former usually possessed occasional transients, whereas the sounds of the latter were steady and regular. The measure of *maximum subband envelope variance* (MSEV) was derived to characterise this feature. The sounds of the water-based background targets and patrol boats were isolated from the other sounds in D2 and D3; the former had high low-frequency energy, whereas the latter possessed strong tonalities. These two dimensions were separately quantified with the *spectral rolloff* (SRO) and *pitch duration* (PD).

2.2.2. Specificities and individual differences

The specificity values of most of the sounds, particularly the natural sounds, were considerably large, thus suggesting the large variations in their spectral–temporal features and source structures.

The different weight patterns in the three latent classes illustrated the individual differences among them. Three experienced subjects (Class 2) assigned the highest weight to the most important dimension (D1) and put little emphasis on the specificities. By contrast, Classes 1 and 3 assigned balanced weights to the different dimensions, with the former assigning weights lower than those assigned by the latter. The results indicated that the performance of the target classification benefited from the training of the subjects.

In summary, we obtained in part I [8] the primary categories of underwater targets and related auditory features. However, the perceptual differences among different man-made targets were masked in this perceptual space.

3. Methods

3.1. Stimuli

The first step in the current work was to select a representative stimulus set. This stimulus set should cover the major categories of interest, and the variations in each category, including the types, operating modes and acoustic characteristics, should be sufficiently large to be representative. Then, the acoustic features thus