| | |
|---|---|
| Title | 音の分離抽出における聴覚の計算理論に関する研究 |
| Author(s) | 鵜木, 祐史 |
| Citation | |
| Issue Date | 1999-03 |
| Type | Thesis or Dissertation |
| Text version | author |
| URL | http://hdl.handle.net/10119/877 |
| Rights | |
| Description | Supervisor:赤木 正人, 情報科学研究科, 博士 |

# A research of the computational theory of audition concerned with sound segregation

Masashi Unoki
School of Information Science
Japan Institute of Science and Technology

14 January 1999

## Abstract

The aim of this paper is to construct a computational theory of audition. This work is to explain the following questions: "what is a purpose of auditory processing?" and "why must auditory system compute it?", based on researches on psychology, physiology, and information science. This computational theory corresponds to the auditory edition of the computational theory of vision proposed by Marr. If the computational theory of audition can be constructed, it can not only clarify human auditory functions but also contribute to some applications such as a signal processing, robust speech recognition, and a modeling of psychoacoustical phenomena. However the computational theory of audition in analogy to Marr's theory has not been constructed completely because psychoacoustical and physiological knowledge of audition is not sufficient to construct it in analogy to Marr's theory.

This paper proposes a computational theory of audition concerned with sound segregation based on the following approaches in analogy to Marr's theory: constraints on sound waves and environment conditions are necessary in order to uniquely solve the problem (ill-posed inverse problem) of segregating the desired signal from mixed signals. This paper adopts the following idea as a construction method of the computational theory: psychoacoustical constraints that auditory system uses to solve the problem of auditory scene analysis, that is the four regularities proposed by Bregman, can be used to uniquely solve the signal segregation problem as mathematical constraints. This paper focuses on "segregation of two sounds" as a fundamental auditory function. Therefore, the problem of segregating sounds is set to "the problem of segregating two acoustic sources." It is supposed that the desired signal is "an AM-FM harmonic complex tone" such as vowel and instrumental sound. Moreover, a computational theory of audition is defined as a strategy of sound segregation, "how are the problem of segregating two acoustic sources solved uniquely using the constraints ?"

In this paper, the problem of segregating two acoustic sources based on an amplitude and phase spectra was formulated. The four regularities proposed by Bregman were formulated as mathematical constraints: (i) common onset and offset for the component of the complex tone, (ii) continuity defined by the piecewise-polynomial approximation and the spline interpolation, (iii) harmonicity, and (iv) correlation between the amplitude envelopes. A method of segregating AM–FM harmonic complex tone from the mixed signal using the constraints was proposed. This method was examined whether it could be segregated the desired signal from the mixed signal or not. This examination showed sufficient constraints for the problem of segregating two acoustic sources. The strategy of sound segregation was derived by examining benefit of the sufficient constraints. The derived strategy is to uniquely solve the problem of segregating two acoustic sources by regarding it as the piecewise linear problem and by constraining the temporal fluctuations of the amplitude and the phase of the desired signal. Finally, the derived strategy was examined by applying it theory into two real segregation problems: (1) the problem of segregating the desired real speech (vowels) from noisy speech and (2) the problem of segregating pure tone from masked signal, that is co-modulation masking release (CMR). These examinations showed the derived strategy of sound segregation can be used to lead the solution of the problems.

This strategy can contribute to the applications such as a preprocessor of the robust speech recognition system and as a modeling of psychoacoustical phenomena. Moreover, it can also contribute to a new construction method of the computational theory of audition in analogy of Marr's computational theory.

**Key words: computational theory, auditory scene analysis, the problem of segregating two acoustic sources, Bregman's four regularities, mathematical constraint**