Auditory Stream Segregation in Auditory Scene Analysis with a Multi-Agent System

Tomohiro Nakatani, Hiroshi G. Okuno, and Takeshi Kawabata
NTT Basic Research Laboratories
3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-01 Japan
{nakatani, okuno, kawabata}@nuesun.ntt.jp

Abstract

We propose a novel approach to auditory stream segregation which extracts individual sounds (*auditory stream*) from a mixture of sounds in auditory scene analysis. The HBSS (*Harmonic-Based Stream Segregation*) system is designed and developed by employing a multi-agent system. HBSS uses only harmonics as a clue to segregation and extracts auditory streams incrementally. When the tracer-generator agent detects a new sound, it spawns a tracer agent, which extracts an auditory stream by tracing its harmonic structure. The tracer sends a feedforward signal so that the generator and other tracers should not work on the same stream that is being traced. The quality of segregation may be poor due to redundant and ghost tracers. HBSS copes with this problem by introducing monitor agents, which detect and eliminate redundant and ghost tracers. HBSS can segregate two streams from a mixture of man's and woman's speech. It is easy to resynthesize speech or sounds from the corresponding streams. Additionally, HBSS can be easily extended by adding agents of a new capability. HBSS can be considered as the first step to computational auditory scene analysis.

Introduction

Over the past years a considerable number of studies have been made on human auditory mechanisms. Although we have many techniques for processing particular sounds such as speech, music, instruments, and the sounds made by specific devices, we don't have enough mechanisms for processing and understanding sounds in real acoustic environments. Research into the latter is being made in the field of *Auditory Scene Analysis* (Bregman 1990), which is to speech recognition is what scene analysis is to character recognition. Auditory scene analysis is a difficult challenging area, partly because acoustic theory is not still rather inadequate (e.g., there is no good acoustic design methodology for concert halls), and partly because most research in acoustics has been focused exclusively on speech and music, ignoring many other sounds. Additionally, the *reductionist* approach to auditory scene analysis, which tries to sum up various techniques for handling individual sounds, is not promising.

Looking and listening are more active than seeing and hearing (Handel 1989). The essentials of our approach to auditory scene analysis are twofold:

- Active perception of observer — looking and listening rather than seeing and hearing, and
- Multi-sensor perception — may use multi-modal information perceived by means of sensor organs

The multi-agent system was recently proposed as a new modeling technology in artificial intelligence (Brooks 1986) (Maes 1991) (Minsky 1986) (Okuno 1993). We assume like Minsky that an agent has a limited capability, although in Distributed Artificial Intelligence, an agent is supposed to be much more powerful like a human being than ours. Each agent has its own goal and competes and/or cooperates with other agents. Through interactions among agents, intelligent behavior emerges (Okuno & Okada 1992).

Consider the approach that the multi-agent paradigm is applied to model auditory scene analysis. We expect that it will enhance the following functionalities: (1) *Goal-Orientation* — Each agent may have its own goal. (2) *Adaptability* — According to the current situation, the behavior of the system varies between reactive and deliberate. (3) *Robustness* — The system should respond sensibly even if the input contains errors, or is ambiguous and incomplete. (4) *Openness* — The system can be extended by adding agents of new capabilities. It can also be integrated into other systems as a building block.

In this paper, auditory stream segregation, the first stage of auditory scene analysis, is modeled and implemented by a multi-agent system. The rest of this paper is organized as follows: Section 2 investigates issues in auditory stream segregation. In Section 3, the basic system of auditory stream segregation with a multi-agent system is explained and evaluated to identify its problems. Section 4 presents and evaluates the HBSS (Harmonic-Based Stream Segregation) that copes with the problems. Related work and the conclusions are given in Section 5 and 6, respectively.

Auditory stream for auditory scene analysis

Auditory stream

Auditory scene analysis understands *acoustic events* or *sources* that produce sounds (Bregman 1990). An acoustic event consists of *auditory streams* (or simply *stream*, hereafter), each of which is a group of acoustic components that have consistent characteristics. The process that segregates auditory streams from a mixture of sounds is called *auditory stream segregation*.

Many techniques have been proposed so far. For example, Brown uses auditory maps in auditory stream segregation (Brown 1992) (Brown & Cooke 1992). These are off-line algorithms in the sense that any part of the input is available to the algorithm at any time. However, off-line algorithms are not well suited for many applications. Additionally, it is not easy to incorporate schema-based segregation and grouping of streams into such a system, since it does not support a mechanism of extending capabilities.

To design a more flexible and expandable system, we adopted a multi-agent system to model auditory stream segregation, and used a simple characteristic of the sounds, that is, the harmonic structure.

Definitions — Harmonic representation

We use only the *harmonic structure* or *harmonicity* of sounds as a clue to segregation. Other characteristics, including periodicity, onset, offset, intensity, frequency transition, spectral shape, interaural time difference and interaural intensity difference, may be used for further processing.

A harmonic sound is characterized by a fundamental frequency and its overtones. The frequency of an overtone is equal to an integer multiple of the fundamental frequency. In this paper, *harmonic stream* refers to an auditory stream corresponding to a harmonic sound, *harmonic component* refers to a single overtone in the harmonic stream, and *agent's stream* refers to the stream an agent traces. We also define the *harmonic intensity* $E(\omega)$ of the sound wave $x(t)$ as

$$E(\omega) = \sum_{k=1}^{n} \| H_k(\omega) \|^2, \quad (1)$$

where

$$H_k(\omega) = \sum_t x(t) \cdot \exp(-jk\omega t), \quad (2)$$

t is time, k is the index of the harmonic components, and ω is the fundamental frequency. We call the absolute value of H_k the *intensity* of the harmonic component, and call the phase of H_k the *phase* of the harmonic component. In this paper, the term *common fundamental frequency* is extended to include the case where the fundamental frequency of one sound coincides with overtone of another sound.

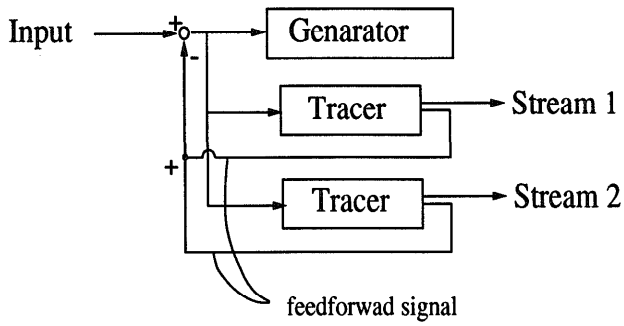


Figure 1: Structure of basic system

Issues in segregation

To extract an auditory stream from a mixture of sounds, it is necessary to find out the harmonic structure, its fundamental frequency and the power of each overtone. The system should segregate auditory streams incrementally, since it will be used as a building block for real-time applications. The important issues to cope with these requirements are summarized below:

1. How to find the beginning of a new harmonic structure,
2. How to trace a harmonic structure,
3. How to reduce the interference between different tracings, and
4. How to find the end of a harmonic structure,

Basic stream segregation

Agents for Basic system

The basic system (Nakatani et al. 1993) consists of two types of agents, the *stream-tracer generator* (hereafter, *the generator*) and *stream tracers* (hereafter, *tracers*). The generator detects a new stream and generates a tracer. The tracers trace the input sound to extract auditory streams. Figure 1 shows the structure of these agents. The input signal consists of the mixed audio waveform.

System parameters The basic system uses three parameters to control the sensitivity of segregation:

1. Power threshold array θ_1 to check for overtones,
2. Power threshold θ_2 to check for fundamental frequencies,
3. Duration T_1 to check for the continuity of sounds.

These three parameters are global and shared among all the agents. The parameter θ_1 is a array of thresholds for frequency regions and plays the most important role in controlling the sensitivity.

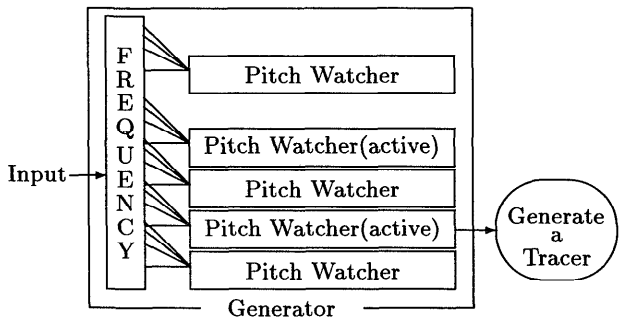


Figure 2: Structure of Generator
Active pitch watch detects a sound.

Generator

The generator detects the beginning of harmonic structures included in the input sounds and generates a new tracer agent. It consists of agents called *pitch watchers* (Figure 2), which monitors the harmonic structure at each frequency ω by evaluating the harmonic intensity defined by Equation 1. Each pitch watcher treats ω as a candidate fundamental frequency, and is activated if the following conditions are satisfied:

1. There is at least one overtone of ω whose power is larger than θ_1,
2. the power of the fundamental frequency, ω, is larger than θ_2, and
3. there is a peak near ω in the acoustic spectrum.

The active pitch watcher with the largest harmonic intensity generates a new tracer, which traces the new stream whose fundamental frequency is ω in Equation 2.

Tracer

Each tracer searches for the fundamental frequency ω_n within the neighborhood of the frequency ω_{n-1} of the previous input frame by maximizing the harmonic intensity (Equation 1). In evaluating Equation 1, overtones whose power is less than θ_1 are discarded. Then, the tracer calculates the intensity and the phase of each harmonic component by using Equation 2.

The tracer terminates automatically if one of the following conditions is satisfied for a period of T_1:

- there is no overtone whose power is larger than θ_1, or
- the power of ω is less than θ_1.

Reducing interference between tracers

A stream should be extracted *exclusively* by one tracer. For this purpose, two tasks are performed by each agent.

Table 1: Benchmark mixtures of two sounds

No	$sound_1$	$sound_2$
1	man's speech	synthesized sound (Fundamental Frequency is 200 Hz)
2	man's speech	synthesized sound (F.F. is 150 Hz)
3	man's speech	woman's speech

Male and female speech utter "aiueo" independently.

Subtract signal As shown in Figure 1, a tracer guesses the input of the next frame and makes a feedforward signal (called *a subtract signal*), which is subtracted from the input mixture of sounds. The waveform to be subtracted is synthesized by adjusting the phase of its harmonic components to the phase of the next input frame. The remaining input (called *the residual input*) goes to all the tracers and to the generator. Each tracer restores the sound data, $x(t)$, by adding the residual signal to its own subtract signal. By this mechanism, the generator does not generate a new tracer for existing streams and one tracer cannot trace another tracer's stream.

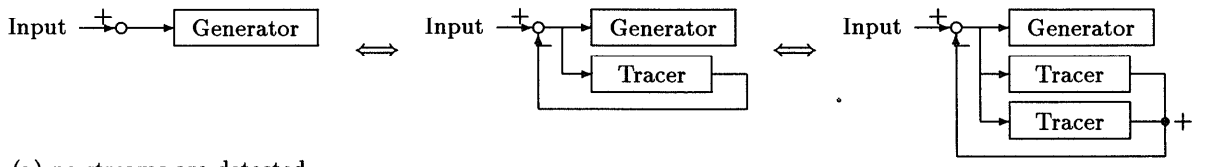
Updating the global parameters θ_1 and θ_2 Each tracer increases the array elements of θ_1 for the regions in the vicinity of the frequency it is tracing. The increase is in proportion to the estimated trace error of each harmonic component, and results in lower sensitivity around the neighboring frequency regions. When terminating, each tracer decreases the array elements of θ_1 in its frequency regions, thereby raising the sensitivity.

Let A be the intensity of a traced harmonic component, ω be the frequency of the harmonic component, and ω' be the representative frequency for each frequency region. We estimate the trace error for the harmonic component at frequency ω' as

$$T(\omega') = c \cdot \| \sum_t A \sin(\omega t) \exp(-j\omega' t) \|,$$

where c is a constant. Since the frequency of higher-order harmonic components is more sensitive to the fundamental frequency than that of lower-order components, the threshold for a higher-order component should be increased over a wider region. Consequently, we use $T(\omega + (\omega_0/\omega) \cdot (\omega' - \omega))$ to increase the local threshold for the harmonic component at frequency ω'.

Each tracer also updates the global parameter θ_2 for every input frame. This is increased by the amount in proportion to the square root of the harmonic intensity. In most regions in vicinity of harmonic components, this value is set much lower than θ_2.



(a) no streams are detected.

(b) one stream is being traced.

(c) two streams are being traced.

Figure 3: Dynamic generation and termination of Tracer agents

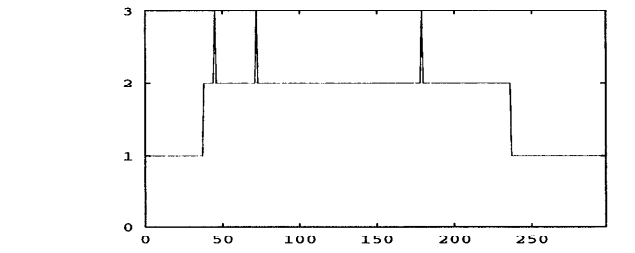


Figure 4: Dynamics of tracer agents (Exp. 1)
(Total number of generated agents = 5)

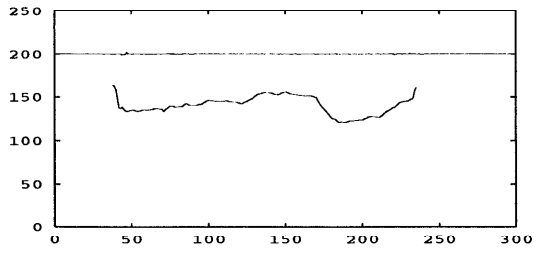


Figure 5: Segregated streams (Exp. 1)

System Behavior

Figure 3(a) shows the initial state of the system. No sound is input to the generator and no tracer is generated. When a new sound enters the system, the generator is activated and a new tracer is generated (Figure 3(b)). Since a tracer is not complete, some errors may be fed into the generator. However, the tracer increases the threshold values adaptively according to the harmonic components, so this mechanism inhibits the generation of inappropriate tracers due to trace errors. The next stream is detected in almost the same way as the first stream (Figure 3(c)). When two or more tracers are generated, each tracer ignores competing components that are excessively influenced by the other streams. As a result, each stream is expected to be well segregated.

Evaluation

We evaluate this system by using three sets of sound mixtures as shown in Table 1. The input signals consisted of combinations of a male speaker and a female speaker uttering Japanese vowels "aiueo", and a stationary synthesized sound with an exponentially attenuating sequence of harmonic components up to 6 kHz.

The input sounds were sampled at 12 kHz, 16-bit quantized, and analyzed with a 30-ms Hamming window. The frame period was 7.5 ms.

Experiment 1 Figure 4 depicts the dynamic generation and termination of tracers in response to the first set of input sounds (Table 1). It shows that three inappropriate tracers (called *redundant tracers*) follow a stream assigned to another tracer, but terminate immediately. The segregated streams are depicted in Figure 5. Both of the sounds resynthesized[1] from the segregated streams are very similar to the original sounds. Additionally, the formants of the original voice are preserved in the resynthesized voice.

Experiment 2 Figure 6 depicts the dynamic generation and termination of tracers in response to the second set of input sounds. The first redundant tracer terminates immediately, while three inappropriate tracers continue to operate for as long as the second sound lasts. One of the three tracers is a redundant tracer. The rest are two *ghost tracers* that traces non-existing streams. The segregated streams are depicted in Figure 7. The sound resynthesized from the segregated stream corresponding to the 150-Hz synthesized sound was very similar to the original sound, but the man's speech was not so good, sounding more like "aiueo-h". Most formants of the original voice are preserved in the resynthesized voice.

[1]At the presentation, the original and resynthesized sounds will be demonstrated.

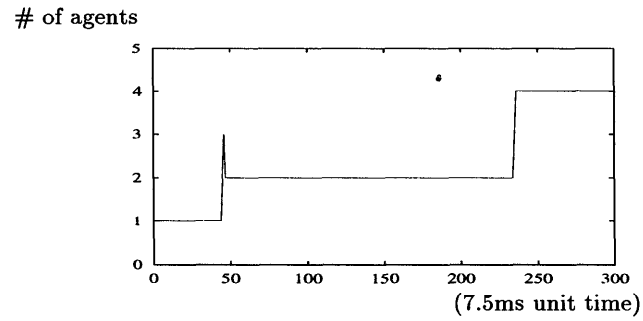


Figure 6: Dynamics of Tracer agents (Exp. 2)
(Total number of generated agents = 5)

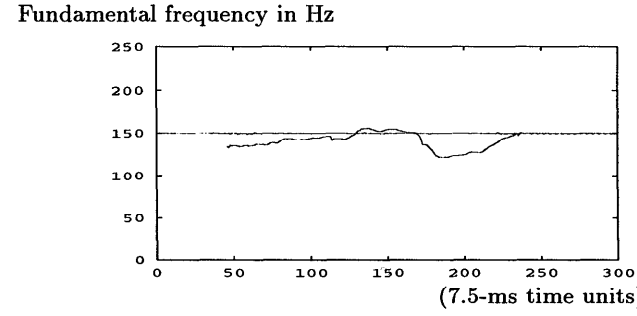


Figure 7: Segregated streams (Exp. 2)

Experiment 3 The third input signal results in the generation of 70 tracer agents, many of which were short-lived as shown in Figure 8. There are many redundant and ghost tracers. However, none of these agents traced both the man's and woman's speech at the same time, as shown in Figure 9. Each of the sounds resynthesized from the corresponding segregated stream was quite poor compared with the original. Additionally, it is not easy to resynthesize a sound by grouping segregated streams. Some formants of man's and woman's original voice are destroyed in each resynthesized voice, respectively.

Summary The basic system occasionally generates redundant and ghost tracers. A redundant tracer is caused by imperfect subtract signals and poor termination detection. A ghost tracer, on the other hand, is caused by self-oscillation, because the phase of the subtract signals is not considered. A pair of ghost tracers usually trace two streams with opposite phases.

Since each tracer extracts a stream according to the current internal status of the tracer and the current residual signal, it is difficult to determine which tracer is inappropriate. In the next section, we extend the basic system to cope with this problem.

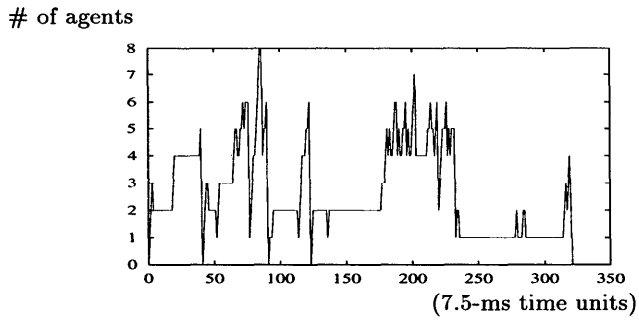


Figure 8: Dynamics of tracer agents (Exp. 3)
(Total number of generated agents = 70)

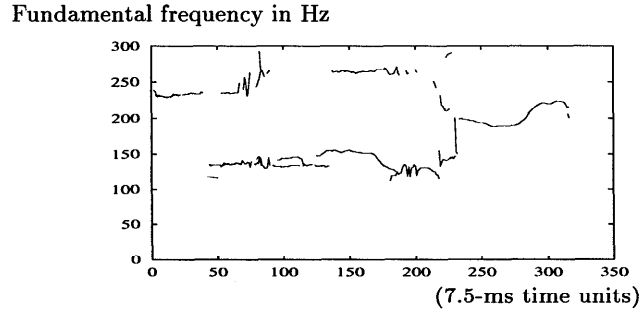


Figure 9: Segregated streams (Exp. 3)

Advanced stream segregation

An advanced stream segregation is proposed to cope with the problems encountered by the basic system (Nakatani et al. 1993). The advanced system is also called the HBSS (*Harmonic-Based Stream Segregation)* system.

Monitors

We introduce agents called *monitors* to detect and kill redundant and ghost tracers. A monitor is generated simultaneously with a tracer, which it supervises (Figure 10). The monitor starts a log for its tracer, and uses it to do the following.

1. Eliminate a redundant tracer, and
2. Adjust the intensity of harmonic components according to the input sound.

Eliminating redundant tracers Redundant tracers should be killed for stable segregation. When the following conditions are met, the monitor judges that its tracer is tracing the same stream as some other tracer.

1. The tracer shares a common fundamental frequency with others for a constant period of time, and

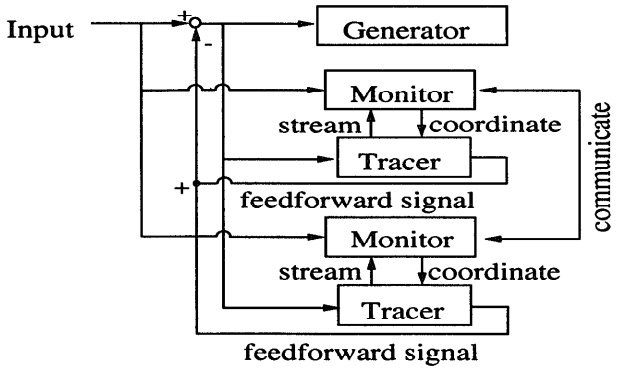


Figure 10: Structure of advanced system (HBSS)

2. The tracers have a common harmonic balance.

The *harmonic balance* is defined as the vector B,

$$B = \frac{1}{\sum_i h_i}(h_1, \ldots, h_m), \tag{3}$$

where $(h_1, \ldots, h_m)$ is a sequence of components each of which is nearly equal to some overtone of the other stream. Two streams have a *common harmonic balance* if the following condition is met:

$$\sum_{i=1}^{m} \frac{\alpha_i}{m} < 1.0 + \epsilon, \tag{4}$$

$$\begin{cases} \alpha_i = r_{1,i}/r_{2,i} & \text{if } r_{1,i} > r_{2,i}, \\ \alpha_i = r_{2,i}/r_{1,i} & \text{otherwise.} \end{cases}$$

Here $(r_{1,1} \ldots r_{1,m})$ and $(r_{2,1} \ldots r_{2,m})$ are the harmonic balances of two streams and ϵ is a constant.

The first condition is easily detected by comparing the trace logs. When a number of monitors detect that their tracers are monitoring the same stream, they all kill each other, along with their tracers, except for the monitor that was generated earliest to trace their common fundamental frequency.

Adjusting the stream Since the feedforward signals of the tracers are subtracted from the waveform, sound components not included in the original sound may be created through tracer interactions. The monitors continuously reduce such sound components in the following way.

Let E_{comp} be the intensity of a harmonic component whose frequency is ω, E_{in} be the intensity of the actual input sound at the corresponding frequency ω in the acoustical spectrum, and r be E_{comp}/E_{in}. If r is greater than the constant c for the past τ frames, the monitor substitutes the value of E_{comp} with the value given by

$$E_{comp} = \alpha(\log(r/c)/\tau + 1.0) \cdot E_{in},$$

where α is a constant. If $\tau = 1$, the change in E_{comp} is small. As τ becomes larger, E_{comp} approaches αE_{in}.

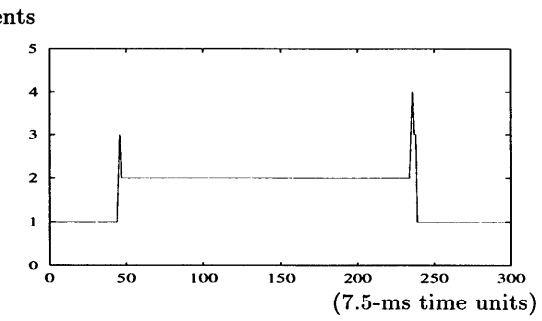


Figure 11: Dynamics of tracer agents (Exp. 4)
(Total number of generated agents = 7)

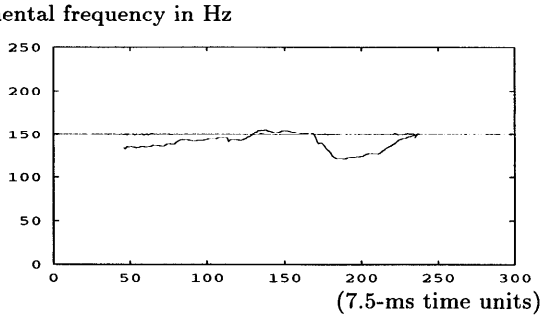


Figure 12: Segregated streams (Exp. 4)

System Behavior

We will briefly explain why inappropriate tracers cannot survive for long. A ghost tracer will be killed as follows:

- A tracer which does not trace an existing stream will be attenuated by the adjustment of its monitor.
- A tracer which traces an existing stream of another tracer will be terminated.

On the other hand, a tracer that is tracing an actual stream is influenced little by the adjustment of its monitor. A redundant tracers will be killed, leaving the oldest tracer to trace the stream stably.

Evaluation

The performance of the advanced system (HBSS) was evaluated using the same set of benchmark signals. Since the first mixture was well segregated even by the basic system, we skipped the result of this experiment with the advanced system.

Experiment 4 Figure 11 shows the dynamic generation and the termination of tracers in response to the second set of input sounds, and segregated streams are depicted in Figure 12. These figures show that

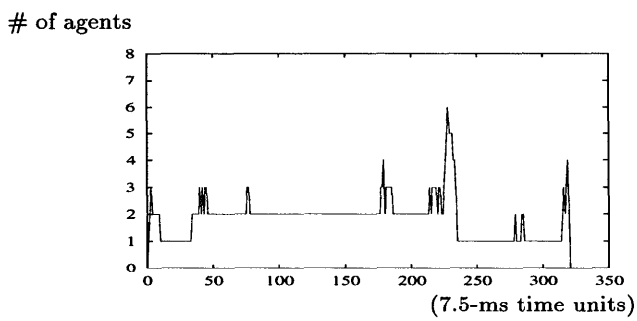


Figure 13: Dynamics of tracer agents (Exp. 5)
(Total number of generated agents = 37)

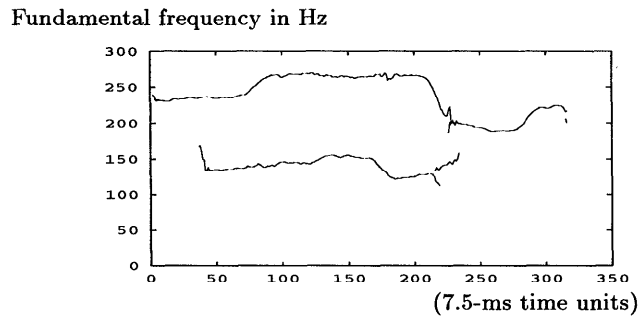


Figure 14: Segregated streams (Exp. 5)

redundant and ghost tracers are killed well. Both sounds resynthesized from the corresponding segregated streams are very similar to the original. Additionally, the formants of the original voice are well preserved in the resynthesized voice.

Experiment 5 The third input signal results in total of 37 generated agents, and Figure 13 shows that redundant and ghost tracers are killed soon. The segregated streams are depicted in Figure 14. Both sounds resynthesized from the segregated streams are not too bad. Additionally, it is easy to resynthesize sounds, because the women's speech was resynthesized from only one stream and the man's speech from two streams. The formants of the man's and woman's original voice are preserved in each resynthesized voice, respectively.

Related Work

Auditory Scene Analysis Bregman classifies the mechanisms of auditory scene analysis into two categories: *simultaneous (spectrum)* and *sequential* grouping (Bregman 1990). The former extracts auditory streams from a mixture of sounds, while the latter groups together auditory streams that belong to the same acoustic source. The Experiment 3 with the third mixture of two sounds in Table 1 shows that it is very difficult to segregate man's and woman's speech by simultaneous grouping followed by sequential grouping. The proposed system integrates both grouping processes and proves to be effective.

Brown and Cooke proposed computational auditory scene analysis (Brown 1992) (Brown & Cooke 1992), which builds auditory map to segregate speech from the other sound such as siren and telephone rings. This system extracts various acoustic characteristics on batch basis, but the extension or interface to other systems is not considered.

Integrated Architecture IPUS (*Integrated Processing and Understanding Signals*) (Lesser & Nawab 1993) integrates signal processing and signal interpretation into blackboard system. IPUS has a small set of front-end signal processing algorithms (SPAs) and choose correct parameters setting for SPA and correct interpretations by dynamic SPA reconfiguration. In other words, IPUS views the reconfiguration as a diagnosis for discrepancy between top-down search for SPA and bottom-up search for interpretation. IPUS has various interpretation knowledge sources which understand actual sounds such as hair driers, footsteps, telephone rings, fire alarms, and waterfalls (Nawab 1992). Since IPUS is a generic architecture, it is possible to implement any capability, but IPUS is fully-fledged. The initial perception can be much simplified. Additionally, a primitive SPA (or agent, in our terminology) that segregates a stream incrementally is not considered so far.

Okuno (Okuno 1993) proposed to use subsumption architecture (Brooks 1986) to integrate bottom-up and top-down processing to realize cognition capabilities. The term "subsumption architecture" is often confused with "behavior-based control" (Brooks 1986), but they are different. The former indicates that interaction between agents is specified by inhibitors and suppressors, or activation propagation (Maes 1991), while the latter indicates that it is behavior that is subsumed. We will use subsumption architecture rather than blackboard architecture, because the former allows agents to interact directly with the environment and can make it easier to extend the capabilities of system.

Wada (Wada & Matsuyama 1993) employed a multi-agent system in deciding regions of image. An agent is placed to a candidate region and then communicates with adjacent agent to determine the boundary of two regions. The interaction of agents is similar to that of HBSS. This and our result proves that a multi-agent system is promising in pattern recognition.

Conclusions

We have presented basic and advanced methods for auditory steam segregation with a multi-agent system, which use only the harmonic structure of input sounds. The advanced system, HBSS, is able to segregate man's

and woman's speech. This result suggests a clue to understanding the *cocktail party problem.* We are about to investigate this problem by designing a new agent that extracts only human voice including consonants by using the information HBSS extracted. This new agent will be added to HBSS with subsumption architecture so that its output subsumes (overwrites) human voice stream segregated by HBSS.

HBSS is currently being evaluated with a wide range of sound mixtures such as a mixture of speech and, white noise of a sound of breaking glass. The performance of segregating human voice from white noise is shown to become worse as the power of white noise increases. However, it is known that constant white noise can be reduced by the spectral subtraction (Boll 1979). We will develop a new agent that reduces white noise by employing the spectral subtraction and use it as a front-end of HBSS.

One might argue that HBSS would not treat transient or bell sounds. This is somewhat true, but is not fatal, because the current HBSS holds and uses just a previous state to segregate auditory streams. We are working to design a new agent that holds longer previous states to restore missing phonemes caused by loud noise. This process is similar to *phonemic restoration* in auditory perception.

In case a mixture of sounds comprises only harmonic sounds and any pair of sounds have not any common fundamental frequency, HBSS would be able to segregate all sounds. This situation is an extension of the first benchmark mixture. Of course, as the number of pairs of sounds that have common fundamental frequency increases, it becomes more difficult to segregate such sounds. This is also the case for human perception. Therefore, we think that, active hearing, or listening, is essential. The typical example of listening is a cocktail party problem.

HBSS uses only harmonics in segregation. This is because we don't either have enough acoustic characteristics to represent a sound or know their hierarchy. In vision, there are a set of visual characteristics and Marr (Marr 1982) proposed their hierarchy, that is, primary and $2\frac{1}{2}$ sketch. It is urgent and important in the research of auditory scene analysis to develop a methodology to represent general acoustics, not restricted to speech or music.

Acknowledgments.

We would like to thank M. Kashino of NTT, H. Kawahara and M. Tsuzaki of ATR for discussions on auditory perception. We would like to thank S.H. Nawab of Boston University and other participants of Abstract Perception Workshop held at Japanese Advanced Institute of Science and Technology for comments on an earlier draft. We would also like to thank I. Takeuchi and R. Nakatsu of NTT for their continuous encouragement of our inter-group research.

References

Boll, S.F. 1979 A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech, In Proceedings of International Conference on Acoustics, Speech, and Signal Processing, IEEE, 200-203.

Bregman, A.S. 1990. *Auditory Scene Analysis – the perceptual organization of sound*, The MIT Press.

Brooks, R.A. 1986. A Robust Layered Control System for a Mobile Robot, *IEEE Journal of Robotics and Automation* RA-2(1): 14–23.

Brown, G. 1992. Computational auditory scene analysis: A representational approach, *PhD thesis*, Dept. of Computer Science, University of Sheffield.

Brown, G.J.; and Cooke, M.P. 1992. A computational model of auditory scene analysis, In Proceedings of International Conference on Spoken Language Processing, 523-526, IEEE.

Handel, S. 1989. *Listening.* The MIT Press.

Lesser, V.; Nawab, S.H.; Gallastegi, I.; and Klassner, F. 1993. IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation in Complex Environments. In Proceedings of the Eleventh National Conference on Artificial Intelligence, 249–255.

Maes, P. ed. 1991. *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, special issue of *Robot and Autonomous Systems*, The MIT Press/Elsevier.

Marr, D. 1982. *Vision.* Freeman.

Minsky, M. 1986. *Society of Minds.* Simon & Schuster, Inc.

Nakatani, T.; Kawabata, T.; and Okuno, H.G. 1993. Speech Stream Segregation by Multi-Agent System. In Proceedings of International Workshop on Speech Processing (IWSP-93), 131–136, The Institute of Electronics, Information and Communication Engineers. Also numbered Technical Report, SP93–97.

Nawab, S.H.; and Lesser, V. 1992. Integrated Processing and Understanding of Signals, 251–285. in Oppenheim, A.V.; and Nawab, S.H. eds. 1992. *Symbolic and Knowledge-Based Signal Processing*, Prentice-Hall.

Okuno, H.G.; and Okada, M. 1992. Emergent Computation Model for Spoken Language Understanding (*in Japanese*). Technical Report SIG-AI 82-3, 21–30, Information Processing Society of Japan.

Okuno, H.G. 1993. Cognition Model with Multi-Agent System (*in Japanese*), 213–225. In Ishida, T. ed. 1993. *Multi-Agent and Cooperative Computation II (Selected Papers from MACC '92)*, Tokyo, Japan: Kindai-Kagaku-sha.

Wada, T.; and Matsuyama, T. 1993. Region-Decomposition of Images by Distributed and Cooperative Processing. Proceedings of the Workshop on Multi-Agent and Cooperative Computation (MACC '93). Japanese Society for Software Science and Technology.