Application of Bayesian Probability Network to Music Scene Analysis

Kunio Kashino*, **Kazuhiro Nakadai**, **Tomoyoshi Kinoshita** and **Hidehiko Tanaka**

H.Tanaka Lab. Bldg#13, Department of Electrical Engineering,
Faculty of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-Ku, Tokyo 113 Japan.
kashino@MTL.t.u-tokyo.ac.jp

Abstract

We propose a process model for hierarchical perceptual sound organization, which recognizes *perceptual sounds* included in incoming sound signals. We consider perceptual sound organization as a scene analysis problem in the auditory domain. Our current application is a *music scene analysis* system, which recognizes rhythm, chords, and source-separated musical notes included in incoming music signals.

Our process model consists of multiple processing modules and a probability network for information integration. The structure of our model is conceptually based on the blackboard architecture. However, employment of a Bayesian probability network has facilitated integration of multiple sources of information provided by autonomous modules without global control knowledge.

1 Introduction

We humans recognize or understand existence, localization and movements of external entities through five senses. We call this function "scene analysis". Scene analysis is viewed here as an information processing which produces valid symbolic representation of external entities or events based on sensory data and stored knowledge. We use the term visual scene analysis for the scene analysis through optical (or visual) information, and the term auditory scene analysis we use for the scene analysis through acoustic (or auditory) information. It is widely admitted that the research which addresses artificial realization of those functions, as well as physiological and psychological approaches, becomes notably important in the upcoming multimedia society.

From engineering point of view, however, current state of research on auditory scene analysis is still in its infancy, when compared with a wide spectrum of the work on visual scene analysis, though several pioneering works on recognition or understanding of non-speech acoustic signals can been found in the literature[Oppenheim and Nawab, 1992; Nakatani *et al.*, 1994; Ellis, 1994; Brown and Cooke, 1994]. Here we consider two directions of development: flexibility of processing and hierarchical structure of auditory scene.

First, we note that the flexibility of existing systems has been rather limited when compared with human auditory abilities. For example, automatic music transcription systems which can deal with given ensemble music played by multiple music instruments have not yet realized, although several studies have been conducted [Roads, 1985; Mont-Reynaud, 1985; Chafe *et al.*, 1985].

Regarding flexibility of auditory functions in humans, recent progress in physiological and psychological acoustics has offered significant information. Especially, the property of information integration in the human auditory system has been highlighted, as demonstrated in the "auditory restoration" phenomena [Handel, 1989]. To achieve flexibility, machine audition systems must have this property, since auditory scene analysis is an inverse problem in general formalization and cannot be properly solved without such information as memories of sound or models of the external world, as well as given sensory data.

Using the blackboard architecture, information integration for sound understanding has already been realized [Oppenheim and Nawab, 1992; Lesser *et al.*, 1993]. However, it is still necessary to consider a quantitative and theoretical background in information integration.

Second, we introduce the concept of *perceptual sounds* as hierarchical and symbolic representation of acoustic entities. The auditory stream[Bregman, 1990] has already been a familiar concept in auditory scene analysis: an auditory stream can be thought as a cluster of acoustic energy formed in our auditory processes. On the other hand, a perceptual sound is a symbol which corresponds to an acoustic (or auditory) entity. In addition, an essential property of perceptual sounds is its hierarchical structure, as discussed in the following sections. Thus auditory scene analysis will be also referred to more precisely as perceptual sound organization in this paper.

With these points as background, we provide a novel process model of hierarchical perceptual sound organization with a quantitative information integration mechanism. Our model is based on probability theory and characterized by its autonomous behavior and theoretically proved stability.

*Currently at NTT Basic Research Laboratories.

2 Problem Description

2.1 Perceptual Sound Organization

An essential problem of perceptual sound organization is a clustering of acoustic energy to create such clusters that humans hear as one sound entity. Here it is important to note that humans recognize various sounds in a hierarchical structure in order to properly grasp and understand the external world. That is, a perceptual sound is structured in both spatial and temporal hierarchy. For example, when one waits for a person to meet standing in a busy street, the waiting person sometimes hears a whole traffic noise as one entity, while sometimes hears a noise of one specific car as one entity. If he or she directs attention to the specific car's sound, an engine noise of the car and a frictional sound from the road surface and the tires of the car might be heard separately as two entities.

Figure 1 shows an example of snapshot of perceptual sounds for music. Note that there is not only spatial structure as shown in this figure but also temporal clusters of perceptual sounds, typically melodies or chord progression, though the temporal structure of perceptual sounds has not been depicted in Figure 1 for simplicity of the figure.

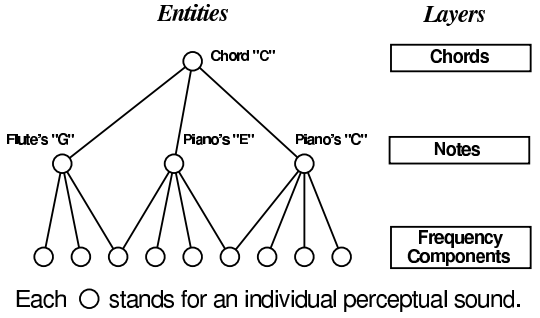

Figure 1: **An example of snapshot of perceptual sounds**

The problem of perceptual sound organization can be decomposed into the following sub problems:

1. Extraction of frequency components with an acoustic energy representation.
2. Clustering of frequency components into perceptual sounds.
3. Recognition of relations between the clustered perceptual sounds and building a hierarchical and symbolic representation of acoustic entities.

Note that we consider the problem as extraction of *symbolic* representation from flat energy data, while some approaches toward "auditory scene analysis" have treated their problem as (*e.g.* evaluated their systems in terms of) restoration of target sound *signals*[Nakatani *et al.*, 1994; Brown and Cooke, 1992]. In the computer vision field, the scene analysis problem has been considered as extration of symbolic representation from bitmap images and clearly distinguished from the image restoration problem which addresses recovery of target images from noise or intrusions.

2.2 Music Scene Analysis

Here we have chosen music as an example of applicable domain of perceptual sound organization. As summarized in Table 1, we use the term music scene analysis in the sense of perceptual sound organization in music. Specifically, music scene analysis refers to recognition of frequency components, notes, chords and rhythm of performed music.

In the following sections, we first introduce general configuration of the music scene analysis system. We then focus our discussion on hierarchical integration of multiple sources of information, which is an essential problem in perceptual sound organization. Then behavior of the system and results of the performance evaluation are provided, followed by discussions and conclusions.

3 System Description

Figure 2 illustrates our processing architecture OPTIMA (Organized Processing toward Intelligent Music Scene Analysis). Input of the model is assumed to be monaural music signals. The model creates hypotheses of frequency components, musical notes, chords, and rhythm. As a consequence of probability propagation of hypotheses, the optimal (here we use the term "optimal" in the sense of "maximum likelihood") set of hypotheses is obtained and outputted as a score-like display, MIDI (Musical Instrument Digital Interface) data, or resynthesized source-separated sound signals.

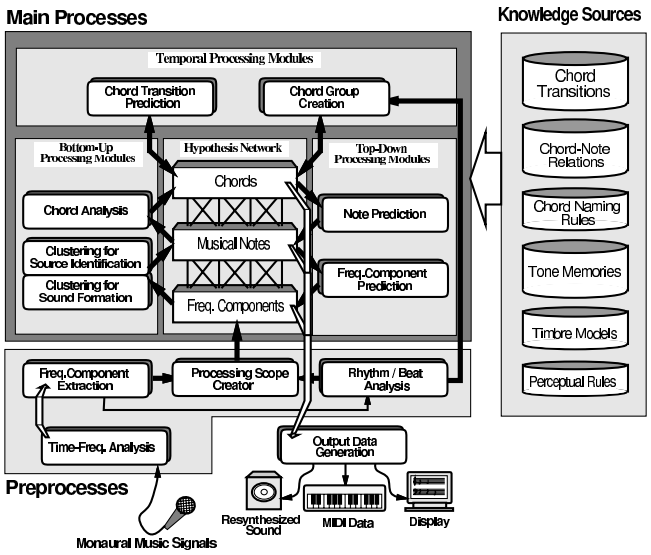

Figure 2: **OPTIMA processing architecture**

OPTIMA consists of three blocks: (A) preprocessing block, (B) main processing block, and (C) knowledge sources. In the preprocessing block, first the frequency analysis is performed and a sound spectrogram is obtained.

With this acoustic energy representation, frequency components are extracted. This process corresponds to

Table 1: **Summary of our terminology**

Words	:	Meanings
perceptual sound	:	A symbol which represents an arbitrary acoustic event in the external world.
perceptual sound separation	:	Extraction of perceptual sounds from incoming sound signals
perceptual sound organization (auditory scene analysis)	:	Construction of an internal model of external acoustical events in a spatial and temporal structure with separation and restoration of perceptual sounds
music scene analysis	:	auditory scene analysis for music sound signals

the first sub problem discussed in the previous section. Since it is difficult to achieve practical accuracy by a simple threshold method, we developed the pinching plane method in peak picking and tracking. As illustrated in Figure 3, This method uses two planes pinching a spectral peak in order to find temporal continuity of spectral peaks. The planes are the regression planes of the peak, calculated by a least-squares fitting.

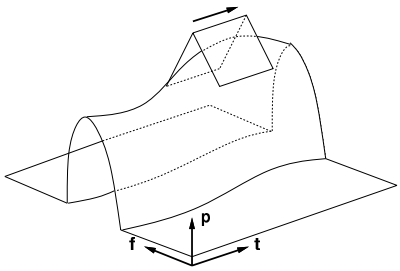

Figure 3: **Extraction of frequency components using pinching planes**

In the case of complicated spectrum patterns, it is difficult to recognize onset time and offset time solely by bottom-up information. Thus the system creates several terminal point candidates for each extracted component.

With Rosenthal's rhythm recognition method [Rosenthal, 1992] and Desain's quantization method [Desain and Honing, 1989], rhythm information is extracted for precise extraction of frequency components and recognition of onset/offset time. Based on the integration of beat probabilities and termination probabilities of terminal point candidates, the candidates were fixed their status: continuous or terminated, and consequently *processing scopes* are formed. Here a processing scope is a group of frequency components whose onset times are close: it is clarified by the experiments on human auditory characteristics that if onset asynchrony of two frequency components is greater than a certain threshold, the two components cannot form one *note* (the value of the threshold is typically 80ms, though the value differs by the frequencies or onset gradients of the components). The processing scope is utilized as a basic time clock for succeeding main processes of OPTIMA, as discussed later.

When each processing scope is created in the preprocessing block, it is passed to the main processing block, as shown in Figure 2. The main block has a hypothesis network with three layers corresponding to levels of abstraction: (1) frequency components, (2) musical notes and (3) chords. Each layer encodes multiple hypotheses. That is, OPTIMA holds an internal model of the external acoustic entities as a probability distribution in the hierarchical hypothesis space.

Multiple processing modules are arranged around the hypothesis network. The modules are categorized into three blocks: (a) bottom-up processing modules to transfer information from a lower level to a higher level, (b) top-down processing modules to transfer information from a higher level to a lower level, and (c) temporal processing modules to transfer information along the time axis. The processing modules consult knowledge sources if necessary. The following sections discuss the information integration at the hypothesis network and behavior of each processing module.

4 Information Integration by the Hypothesis Network

For information integration in the hypothesis network, we require a method to propagate impacts of new information through the network. We employ Pearl's Bayesian network method [Pearl, 1986], which can fuse and propagate new information represented by probabilities through the network using two separate links (λ-link and π-link) if the network is a singly connected (*e.g.* tree-structured) graph.

Figure 4 shows our application of the hypothesis network. As shown in the previous section, the network has three layers: (1) C(Component)-level, (2) N(Note)-level, and (3) S(Chord)-level. The link between the C-level node and the N-level node is the S(Single)-Link, which corresponds to one processing scope. The link between the S-level and the N-level becomes the M(Multiple)-Link, as a consequence of temporal integration: multiple notes along time axis may form a single chord. The S-level nodes are connected along time by the T(Temporal)-Link, which encodes chord progression.

To discuss information integration scheme, assume we wish to find the belief (BEL) induced on the Node A in Figure 4, for example. Letting D_A^- stand for the data contained in the tree rooted at A and D_A^+ for the data contained in the rest of the network, we have

$$BEL(A) = P(A|D_A^+, D_A^-) \tag{1}$$

where A is a probability vector: $A = (a_1, a_2, \cdots, a_M)$. Using Bayes' theorem and assuming independence of hy-

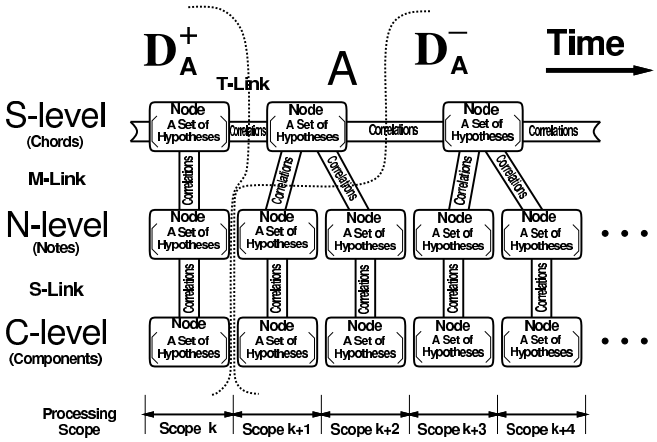

Figure 4: **Topology of the hypothesis network**

potheses

$$P(D_A^+, D_A^-|a_j) = P(D_A^+|a_j)\ P(D_A^-|a_j), \qquad (2)$$

we have

$$P(A|D_A^+, D_A^-) = \alpha P(D_A^-|A)\ P(A|D_A^+), \qquad (3)$$

where α is a normalization constant.

Substituting as $\lambda(A) = P(D_A^-|A)$ and $\pi(A) = P(A|D_A^+)$, Equation (3) can be written as

$$BEL(A) = \alpha\ \lambda(A)\ \pi(A). \qquad (4)$$

Given conditional probabilities P(Child|Parent) between any two adjacent nodes, $\lambda(A)$ can be derived from λ(Children of A) and $\pi(A)$ from π(Parent of A)[Pearl, 1986]. This derivation is considered as propagation of diagnostic (λ) or causal (π) support to A.

A minimum set of processing modules required in each node of the network is shown in Figure 5. B-Holder holds the belief (BEL) and passes new information as λ and π messages to the adjacent B-Holders. In our OPTIMA model, B-Holders are embedded in the hypothesis network and not explicitly drawn in Figure 2. H-Creator creates the hypotheses with initial probabilities. H-Correlator is for evaluating conditional probabilities $P(\text{Node}_1|\text{Node}_2)$, where Node_2 is a parent of Node_1, which are required in the information propagation process.

5 System Behavior

Based on the OPTIMA processing architecture, a music scene analysis system has been implemented. This section discusses configuration of knowledge sources and the behavior of processing modules in the main processing block in Figure 2.

5.1 Knowledge sources

Six types of knowledge sources are utilized in OPTIMA.

The **chord transition dictionary** holds statistical information of chord progression, under the N-gram assumption (typically we use N=3); that is, we currently assume that the length of Markov chain of chords is

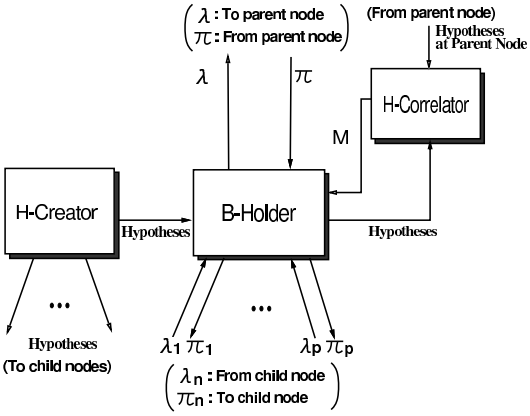

Figure 5: **Processing modules for each node of hypothesis network**

three, for simplicity. Since each S-level node has N-gram hypotheses, one can note that the independence condition stated by Equation (2) is satisfied even in S-level nodes. We have constructed this dictionary based on statistical analysis of 206 traditional songs (all western tonal music), which are popular in Japan and other countries.

In the **chord-note relation** database, probabilities of notes which can be played under a given chord are stored. This information is obtained by statistical analysis of the 2365 chords.

The **chord naming rules**, based on a music theory, are used to recognize chord when hypotheses of played notes are given.

The **tone memory** is a repository of frequency components data of a single note played by various musical instruments (Figure 6). Currently it maintains notes played by five instruments (clarinet, flute, piano, trumpet, and violin) at different expressions (forte, medium, piano), frequency range, and durations. We recorded those sound samples at a professional music studio.

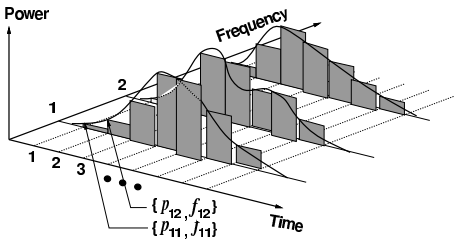

Figure 6: **Tone memory**

The **timbre models** are formed in the feature space of the timbre. We first selected 43 parameters for musical timbre, such as onset gradient of the frequency components and deviations of frequency modulations, and then reduced the number of parameters by the principal component analysis. With the proportion value of 95%, we have an eleven-dimensional feature space where at least timbres of above mentioned five instruments are completely separated with each other. Assuming that

one category of timbre has the normal distribution in the timbre space, we use $\bar{x}_j$, the averaged value of j-th parameter ($j = 1, 2, ..., m$), and a variance-covariance matrix V as timbre model parameters for a timbre category A. Using Mahalanobis' distance D_i^2, the probability P for the i-th note to belong to the category A can be calculated as

$$P = \frac{1}{(2\pi)^{m/2}\sqrt{|S|}} \exp\left\{-\frac{1}{2}D_i^2\right\} \quad (5)$$

where $S = V^{-1}$ and

$$D_i^2 = \sum_j \sum_k (x_{ij} - \bar{x}_j)\ S_{jk}\ (x_{ik} - \bar{x}_k). \quad (6)$$

Finally, the **perceptual rules** describe the human auditory characteristics of sound separation[Bregman, 1990]. Currently, the harmonicity rules and the onset timing rules are employed[Kashino and Tanaka, 1993] based on psychoacoustic experiments.

5.2 Bottom-up processing modules

There are two bottom-up processing modules in OPTIMA: **NHC** (Note Hypothesis Creator) and **CHC** (Chord Hypothesis Creator). **NHC** is a H-Creator for the note layer, and performs the clustering for sound formation and the clustering for source identification to create note hypotheses (Figure 7). It uses the perceptual rules for the clustering for sound formation, and the timbre models for discrimination analysis of timbres to identify the sound source of each note. **CHC** is a H-Creator for the chord layer, which creates chord hypotheses when note hypotheses are given. It refers to chord naming rules in the knowledge sources.

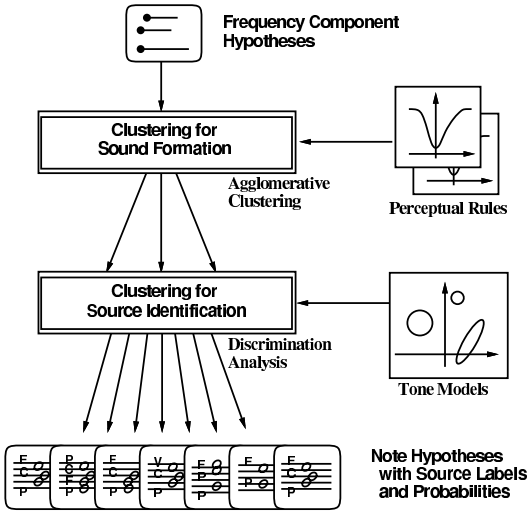

Figure 7: **Note hypothesis creator**

5.3 Top-down processing modules

FCP (Frequency Component Predictor) and **NP** (Note Predictor) are the top-down processing modules. **FCP** is a H-Correlator between the note layer and the frequency component layer, and evaluates conditional probabilities between hypotheses of the two layers, consulting tone memories (Figure 8). **NP** is a H-Correlator between the chord layer and the note layer, to provide a matrix of conditional probabilities between those two layers (Figure 9). **NP** uses the stored knowledge of chord-note relations.

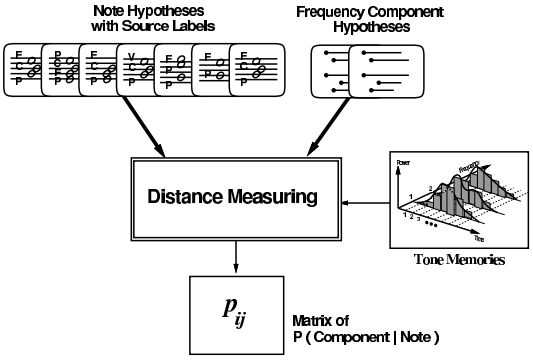

Figure 8: **Frequency component predictor**

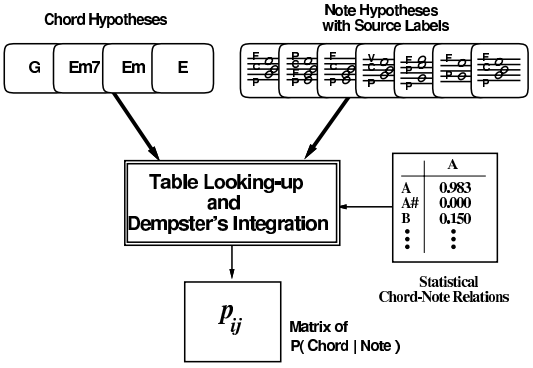

Figure 9: **Note predictor**

5.4 Temporal processing modules

There are also temporal processing modules: **CTP** (Chord Transition Predictor) and **CGC** (Chord Group Creator). **CTP** is a H-Correlator between the two adjacent chord layers, which estimates the transition probability of two N-grams (not the transition probability of two chords), using the chord transition knowledge source (Figure 10). **CGC** decides the M-Link between the chord layers and the note layers. In each processing scope, **CGC** receives chord hypotheses and note hypotheses. Based on rhythm information extracted in the preprocessing stage, it tries to find how many successive scopes correspond to one node in the chord layer, to create M-Link instances. Thus the M-Link structure is formed dynamically as the processing progresses.

6 Note Level Evaluation using Benchmark Test Signals

We have performed a series of evaluation tests on the system: frequency component level tests, note level tests,

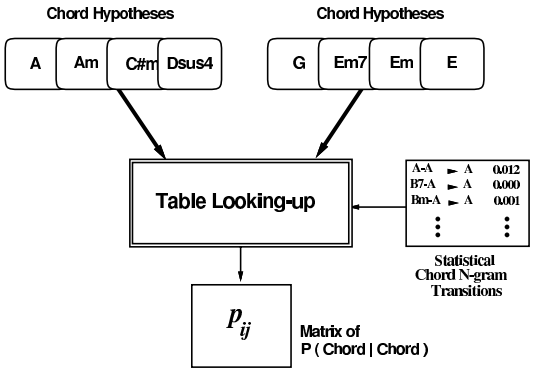

Figure 10: **Chord transition predictor**

chord level tests, and tests using sample song performances. In this section, however, we will concentrate on the note-level evaluation and present a part of the results of benchmark tests.

6.1 Test Patterns

In note-level benchmark tests, we used simultaneous note patterns such as shown in Figure 11. In those patterns, a given number (typically two or three) of simultaneous notes were performed repeatedly by a MIDI sampler using digitized acoustic signals (16bit, 44.1kHz) of natural musical instruments (clarinet, flute, piano, trumpet, and violin). Note patterns were randomly composed by a computer, with one of the following constraints:

Class 1 pattern: At least two of simultaneous notes are always in a octave relation in pitch. That is, in the case of a two simultaneous note pattern, the fundamental frequencies of simultaneous two notes are always in harmonic relations.

Class 2 pattern: At least two of simultaneous notes are always in a 0.5 octave (fifth) relation in pitch. That is, in the case of a two simultaneous note pattern, the second (fourth, sixth, ...) harmonic of one note and the third (sixth, ninth, ...) harmonic of another note always overlap.

Class 3 pattern: Note patterns which do not belong to class 1 nor class 2.

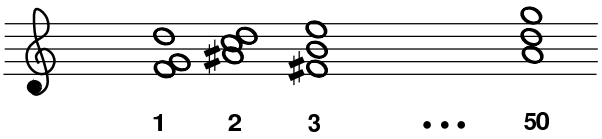

Figure 11: **An example of note patterns for note-level benchmark tests**

6.2 Parameters for Evaluation

The note recognition index R was defined as

$$R = 100 \cdot \left(\frac{right - wrong}{total} \cdot \frac{1}{2} + \frac{1}{2} \right) \quad [\%], \qquad (7)$$

where *right* is the number of correctly identified and correctly source-separated notes, *wrong* is the number of spuriously recognized (surplus) notes and incorrectly identified notes, and *total* is the number of notes in the input. Since it is sometimes difficult to distinguish surplus notes from incorrectly identified notes, both are included together in *wrong*. Scale factor 1/2 is for normalizing R: when the number of output notes is the same as the number of input notes, R becomes 0 [%] if all the notes are incorrectly identified and 100 [%] if all the notes are correctly identified by this normalization.

In addition, we also use the retrieval index α and the precision index β for the note-level evaluation:

$$\alpha = \frac{a}{n}, \qquad \beta = \frac{b}{n}, \qquad (8)$$

where n is the number of total notes in the input, a is the number of correctly identified and correctly source-separated notes in the output, b is the number of the other notes in the output (Figure 12). The R can be written using α and β as

$$R = \frac{1}{2}(\alpha - \beta) + \frac{1}{2}. \qquad (9)$$

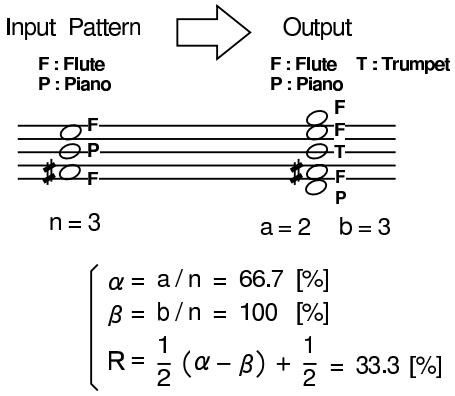

Figure 12: **Parameters for note-level evaluation**

6.3 Results

In our experiments, tests have been performed in three ways: perceptual sound organization (case 1) without any information integration, (case 2) with information integration at the N-level only, and (case 3) with all (*i.e.* N-level and S-level) information integration. In the first case, the best note hypothesis produced by the bottom-up processing (**NHC**) is just viewed as the answer on the system, while the other two cases the tone memory information given by **FCP** is integrated.

Experimental results for the N-level evaluation is displayed in Figure 13 and 14 (class 1), Figure 15 and 16 (class 2), and Figure 17 and 18 (class 3). In the $\alpha - \beta$ plot figures, results for the case 1 and case 2 are displayed.

These results clearly indicate that integration of tone memory information has significantly improved the note recognition accuracy of the system. Especially, in the

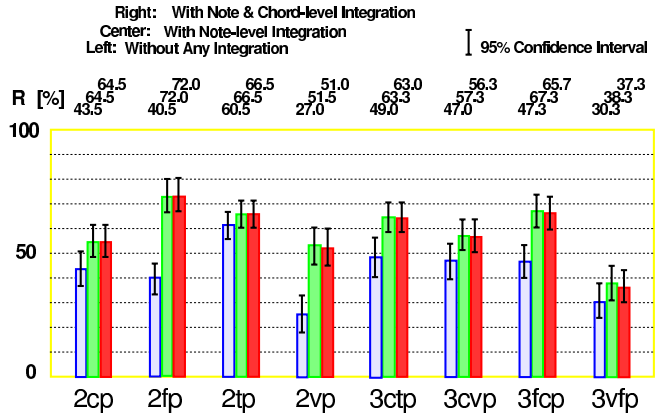

Figure 13: **Results of benchmark tests for note recognition (class 1)**

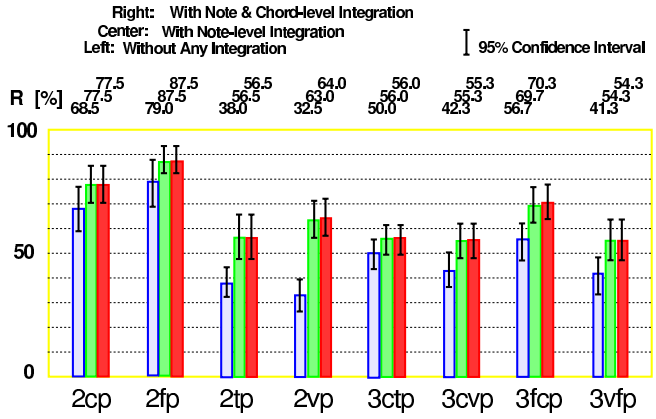

Figure 15: **Results of benchmark tests for note recognition (class 2)**

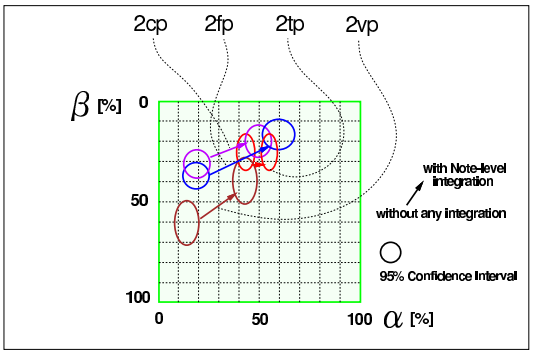

Figure 14: $\alpha - \beta$ **plot (class 1)**

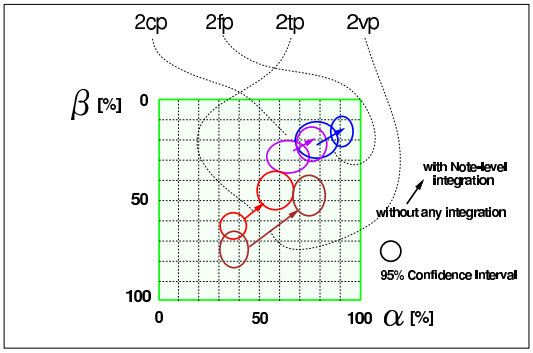

Figure 16: $\alpha - \beta$ **plot (class 2)**

class 1 patterns, where most frequency components are overlapped, the α values before information integration is quite low. However, **NHC** creates octave hypotheses when there are only low probability hypotheses. Thus the note-level information integration can effectively work and consequently the α values are improved.

It is natural that chord-level integration did not affect the results, since the note patterns used in these tests are randomly (under the class constraint) composed and have nothing to do with the stored chord transition knowledge.

7 Conclusion

We first discussed problem of perceptual sound organization in terms of construction of valid symbolic hierarchical representation from incoming acoustic energy. Since we believe that an essential technical issue on this problem is quantitative integration of multiple sources of information, we then discussed information integration scheme based on Bayesian probability network, which have been applied to a music scene analysis system.

The Bayesian probability network enables us stable information integration without any global control knowledge. The experimental results show that the integration of tone memory information significantly improves the recognition accuracy for perceptual sounds, in comparison with a conventional bottom-up based processing. Especially, in the situation that most frequency components in multiple notes are overlapping (*e.g.* class 1 and class 2), use of the tone memory information is found to be essential.

While we concentrated on the note-level evaluation in this article, we have found that the chord level information is also very effective: for the efficacy of chord transition information, see also our another paper[Kashino and Tanaka, 1995].

One of the limits of our method lies in the topological constraint of the probability network: probability propagation scheme described in this paper cannot be applied to a multiply-connected graph (e.g. mesh-structured graph). In our music scene analysis system, most common error at the note level is misidentification of instruments, and the second major error is the overtone pitch error. If we could introduce note-level transition information to the system, as well as the chord-level transition information, the recognition accuracy would be further improved; for example, such information that "it is not usual that one note of trumpet suddenly cuts into a flute and piano duet", or "this note is unusually high in pitch judging from this melody stream" would be use-

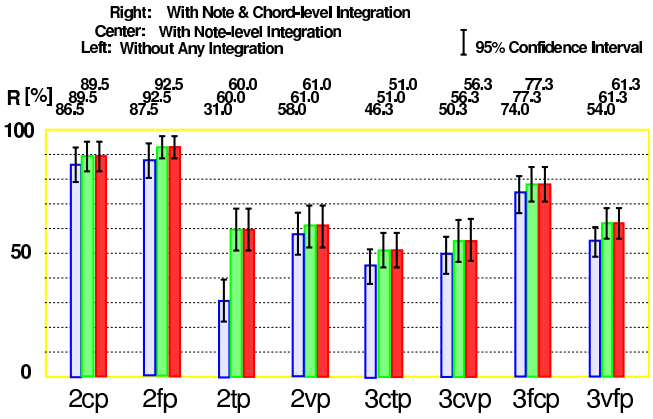

Figure 17: **Results of benchmark tests for note recognition (class 3)**

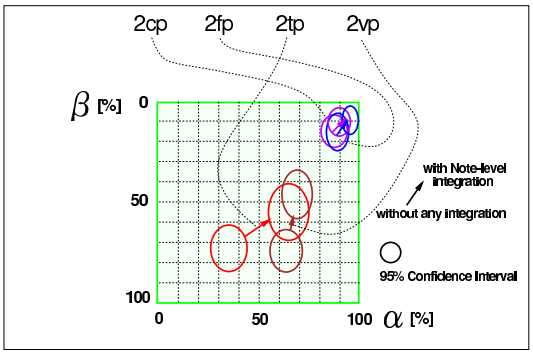

Figure 18: $\alpha - \beta$ **plot (class 3)**

ful. Thus we anticipate that developing an information integration method applicable to a multiply-connected graph will be the next step of our approach.

References

[Bregman, 1990] Bregman A. S. *Auditory Scene Analysis.* MIT Press, 1990.

[Brown and Cooke, 1992] Brown G. J. and Cooke M. A Computational Model of Auditory Scene Analysis. In *Proceedings of International Conference on Spoken Language Processing*, pages 523–526, 1992.

[Brown and Cooke, 1994] Brown G. J. and Cooke M. Perceptual Grouping of Musical Sounds: A Computational Model. *Journal of New Music Research*, 23(1):107–132, 1994.

[Chafe *et al.*, 1985] Chafe C., Kashima J., Mont-Reynaud B., and Smith J. Techniques for Note Identification in Polyphonic Music. In *Proceedings of the 1985 International Computer Music Conference*, pages 399–405, 1985.

[Cooke *et al.*, 1993] Cooke M. P., Brown G. J., Crawford M. D. and Green P. D. Computational auditory scene analysis: Listening to several things at once. Endeavour, 17(4):186–190, 1993.

[Desain and Honing, 1989] Desain P. and Honing H. Quantization of Musical Time: A Connectionist Approach. *Computer Music Journal*, 13(3):56–66, 1989.

[Ellis, 1994] Ellis D. P. W. : A Computer Implementation of Psychoacoustic Grouping Rules. In *Proceedings of 12th International Conference on Pattern Recognition*, 1994.

[Handel, 1989] Handel S. *Listening.* MIT Press, 1989.

[Kashino and Tanaka, 1993] Kashino K. and Tanaka H. A Sound Source Separation System with the Ability of Automatic Tone Modeling. In *Proceedings of the 1993 International Computer Music Conference*, pages 248–255, 1993.

[Kashino and Tanaka, 1995] Kashino K., Nakadai K., Kinoshita T. and Tanaka H. Organization of Hierarchical Perceptual Sounds : Music Scene Analysis with Autonomous Processing Modules and a Quantitative Information Integration Mechanism. In *Proceedings of the 1995 International Joint Conference on Artificial Intelligence*, 1995.

[Lesser *et al.*, 1993] Lesser V., Nawab S. H., Gallastegi I. and Klassner F. IPUS: An Architecture for Integrated Signal Processing and Signal Interpretation in Complex Environments. In *Proceedings of the 11th National Conference on Artificial Intelligence*, pages 249–255, 1993.

[Mellinger, 1991] Mellinger D. K. *Event Formation and Separation of Musical Sound.* Ph.D. Thesis, Department of Music, Stanford University, 1991.

[Mont-Reynaud, 1985] Mont-Reynaud B. Problem-Solving Strategies in a Music Transcription System. In *Proceedings of the 1985 International Joint Conference on Artificial Intelligence*, pages 916–918, 1985.

[Nakatani *et al.*, 1994] Nakatani T., Okuno H. G., and Kawabata T. Auditory Stream Segregation in Auditory Scene Analysis with a Multi-Agent System. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 100–107, 1994.

[Oppenheim and Nawab, 1992] Oppenheim A. V. and Nawab S. H. *(eds.). Symbolic and Knowledge-Based Signal Processing.* Prentice Hall, 1992.

[Pearl, 1986] Pearl J. Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, 29(3):241–288, 1986.

[Roads, 1985] Roads C. Research in Music and Artificial Intelligence. *ACM Computing Surveys*, 17(2):163–190, 1985.

[Rosenthal, 1992] Rosenthal D. *Machine Rhythm: Computer Emulation of Human Rhythm Perception.* PhD. Thesis, Department of Computer Science, Massatussetts Institute of Technology, 1992.