# A Mix-Down Assistant Interface with Reuse of Examples

Haruhiro Katayose, Akio Yatsui
Kwansei Gakuin University
School of Science and Technology
Gakuen Sanda, 565-1337, Japan
{katayose, yatsui}@ksc.kwansei.ac.jp

Masataka Goto
National Institute of
Advanced Industrial Science and Technology (AIST)
Ibaraki, 305-8561, Japan
m.goto@aist.go.jp

## Abstract

*Mix-down (track down) occurs at the final stage of commercial music production. Selecting effectors for each sound track and setting the parameters of each effecter balances the sound from each track in stereo (mix-down design). The mix-down process greatly influences the final sound quality. Recently, professional mixing tasks have been done on digital audio workstations, i.e., software on PCs. In this sense, amateur musicians have entered the realm of professional production. However, it is difficult for mix-down beginners to obtain the design they want. One rational way of assisting mix-down is by using examples. In this paper we propose a mix-down assistance interface that copies an existing mix-down design to the given music, and describe its functions, evaluations and possibilities.*

## 1. Introduction

Mix-down (Trackdown) is the process of giving effects to each recorded music track and consolidating them to a stereo mix by adjusting volumes and spatialization. The mix-down process greatly influences the final sound quality. It is not an exaggeration to say that the musical quality is determined by the efficacy of mix-down design. That is, if the design in mixdown differs, the impression of the music differs a lot.

The mixdown design of a music piece is not unique. There are several ways to complete mixdown designs. Professional mixdown engineers elaborate the mixdown design, as they imagine the completed work. The number of avaialble effectors often exceeds 200. The mixdown task requires expertise regarding these effectors. The engineers have to select effectors and their parameters. Sometimes the order of using the effectors are crucial in accomplishing the mixdown design (See Figure 1).

Recent professional mixing tasks have been done on a digital audio workstation or hard-disk recording system [1] including software on personal computers. In this sense, amateur musicians have entered the realm of professional production. However, it is difficult for mix-down beginners to obtain the design they want. One rational way of assisting mix-down is by using examples that are designed by experienced engineers. We have been proposing and embodying interfaces to copy mix-down designs. In Section 2, we address the fundamental merit of design assistance by referring to examples. In Section 3, we describe a technical solution to realize design copying in the mix-down process; analysis of the musical structure, timbre and a style of rendition for each track. In Section 4, we introduce a preliminary analysis of the interfaces. In Section 5, we discuss applications of the proposed system.

## 2. An interface for mix-down design assistance by referring to examples

### 2.1. Efficiency of using examples in conveying a design concept

This subsection summarizes requirements for assisting the mix-down process for amateurs and proposes interfaces for each requirement.

Let us suppose that an amateur band (client) asks a professional mixing engineer to complete a piece that the band has recorded in a multi-track system. The client has to convey her/his design image to the engineer, except in the case where the client leaves everything in charge of the engineer. It is not easy to convey a design concept for non-verbal media using language. Using many adjectives might be a solution to that. It is, however, difficult to judge whether or not the design image has been conveyed adequately by that method. It is much easier to explain the features of the design by referring to concrete examples[4]. In fact, people

[1] A professional digital audio production system *Pro Tools* (Digidesign, Inc.) is a de facto standard.

**Figure 1. Tasks of mix-down design.**

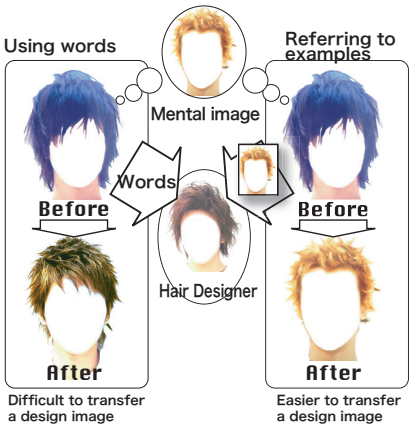


**Figure 2. Efficiency of using examples in conveying a design concept**

concerned with design tasks, especially in content production, communicate design concepts by referring to existing examples when elaborating their designs. It may be easier to realize the efficiency of referring to examples, if we imagine the conversation between a hair designer and a client (Fig.2).

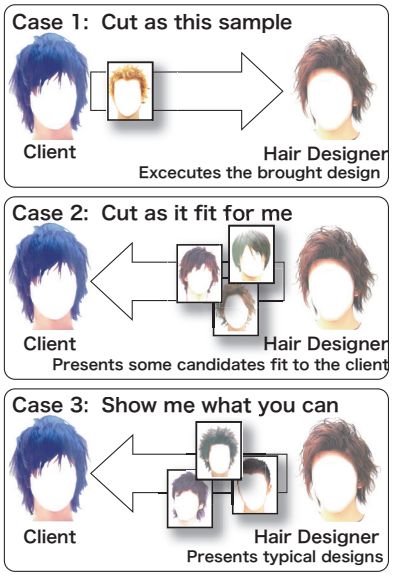


**Figure 3. Three types of design interfaces using examples**

## 2.2. Design interfaces referring to examples

The design interfaces that refer to examples are classified into the three types. Figure 3 shows three types in the case of hair design.

**Type 1) Copy this design** This is a type where a client provides or selects an example. A mix-down engineer copies the mix-down design of the example to the target music piece.

**Type 2) Make the best use of the source** This is a type where an engineer executes the design, trying to make the best use of the constituents of the target music. In case of hair design, face shape, hair type, and hair quantity would be constituents. The designer would select some examples of hair designs suitable for the client's constituents from her/his repository, and show the examples to the client. In case of music, the designer would find and show some similar completed examples considering music genre and constituents of the target music. After agreeing on a design concept, the mix-down engineers copy the design to the target.

**Type 3) Show me what you can** There is a case where a client wants to see design examples beforehand. For example, short lair, Mohicans, and Afro are typical designs in hair design. Although the engineer may not have to consider the best use of constituents of the target music, the engineer has to have prepared typical examples in advance.

For every case where we copy an existing mix-down design to a target, we have to provide functions for the analysis of music structure, the analysis of timbre and style of rendition for each track, and the utilization of annotation. These technique is necessary, especially for the realization of Type 2.

### 2.3. Use of annotation

The annotations to be used in mix-down design task can be classified into groups of incidental information and information related to the sound. The incidental information could be, "This music was used for a horror film" and "This piece was mix-downed by the engineer with an established reputation in dance music", for instance. Information related to the sound includes the name of the musical instrument, the way of rendition of each track and musical structures. In addition, some quantitative features, such as partial model and envelope (Attack Decay Sustain Release) model can be used to refer to sound. There are plenty mix-down designs so far. Yet the mix-down designers have not made annotations in the design tasks. It is meaningful to automatically annotate information related to the sound by analyzing music.

In the task of copying mix-down design, we have to match music tracks between the copy source and the copy target as shown in Figure 4. Constitution, number or order of musical instruments used in a musical piece differ, if the musical pieces differ. We should match each track based on the analysis of sounds of music tracks. To get heuristics that can be used for this analysis, we examined the mixdown settings for 100 pieces of the popular-music database "RWC Music Database: Popular Music" (RWC-MDB-P-2001) [3], which is an original database available to researchers around the world. The findings are summarized as follows;

- The constituents of music or the number of musical instruments tends to change at the beginning of music sections such as verse A, verse B, and refrain (chorus).

- A part played by the same musical instrument is separated into several tracks according to each music section.

Based on these findings, the timbre and rendition are analyzed for each track. Then, these features are used for the analysis of music sections. The features are also utilized for annotation. At the same time, these are used as the measure of similarity for the realization of the Type 2 interface.

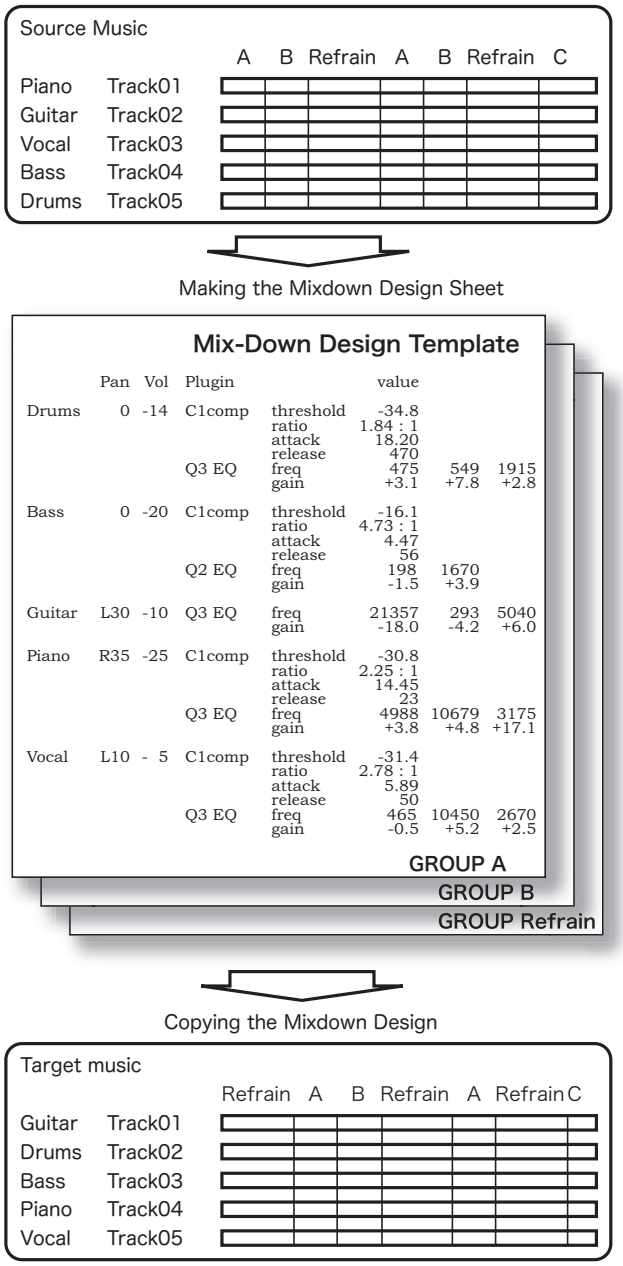


**Figure 4. Making the mixdown design sheet and copying to the target**

## 3. Analysis of timbre and rendition of each track and music sections

As mentioned above, music sound analysis is one of the crucial processing for the interfaces with reuse of examples. In the proposed system, input musical pieces are analyzed according to the following steps.

**step1** Input all of the musical materials.

**step2** Classification of each track into vocal (melody) part, percussion, and the other categories.

**step3** Beat detection in the tracks that are classified as percussion parts.

**step4** Music section analysis for the tracks which are classified as vocal (melody) parts.

**step5** Analysis of timber and rendition for each track.

### 3.1. Outline of studies of music section analysis

Analysis of repetitive patterns of music acoustics is one of the primal research areas in the research field of music information retrieval. The analysis of music sections in our system is based on Goto's RefraiD (Refrain Detecting Method)[2]. RefraiD extracts information of music sections by analyzing repetitive patterns based on autocorrelation of the reduced time-frequency map derived from music CDs.

The interface proposed here is for popular music. Most popular music includes a melody line and percussion part. The basic idea is to obtain music sections such as verse A, verse B, or refrain from the melody part. This task is regarded as a more simple one compared with RefraiD. The system first identifies the percussion track and detects beat. Next, the system analyses the pitch of the melody part. Then the system calculates autocorrelation of the pitch transition for each beat and finally extracts the music section information.

#### 3.1.1 Percussion Identification and beat detection

The common features of the timbre of the percussion instruments (snare drum, bass drum, tam tam, hi-hat, cymbal, ride cymbal, crush cymbal) is that the sound has larger amount of power over 1300 Hz at the attack time (within 10 ms), compared with other musical instruments. Therefore we decided to use an index the power ratio over 1300 Hz multiplying its frequency to determine whether or not the sound is that of a percussion instrument.

Then, the system identifies the percussive sound track and gives its name considering spectral envelope, tone duration, and beat frequencies. Then, the system obtains the beat unit, tracking IOI (Inter Onset Intervals) for the tracks that are identified as percussive.

#### 3.1.2 Estimation of vocal part

Most of main vocal parts are recorded, assigning an exclusive single track for the vocalist. Then, we employ content of non-chord note ratio as an index of identification of vocals.

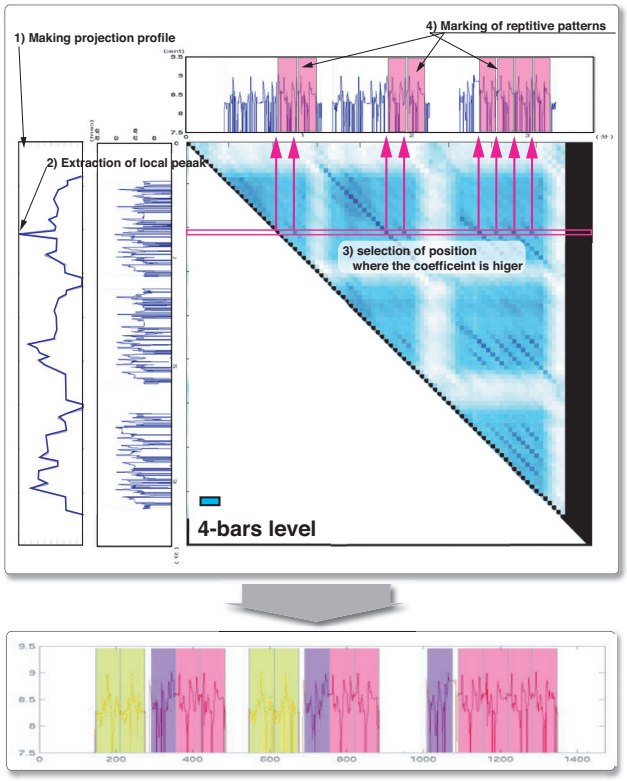


**Figure 5. Analysis of music sections**

F0 extraction using auto-correlation gives the correct result for the input of a monophonic melody, but it gives unstable values when chords are processed. On the contrary, harmonic summation operation for the time-frequency map obtained by STFT will be able to emphasize F0 even for chords.

We adopted vocal melody measure index by the combination of F0 stability calculated by auto-correlation and correspondence ratio in both pitch estimation methods. Furthermore, the system identifies the track into vocals, base, and other melodies, considering registers and vibratos.

#### 3.1.3 Analysis of musical sections using vocal melodies

The procedure described in the above section can identify the vocal melody and pitch has already been estimated. The repetitive section structure is analyzed by calculating autocorrelation of pitch transition of the melody, shifting by beat unit. This processing is shown in Figure 5. This is a correlation map of 4-bar level of the analysis for RWC-MDB-P-2001 No.13 in RWC Music Database. The high correlation part appears deeply colored in the correlation map. Next, we are going to show steps to obtain information of music section structure from the correlation map.

1. Make a projection profile for each column of the correlation map.
2. Search for local peaks in the projection profile in order to find the starting point of the repetitive pattern.
3. Search for the starting points of multiple repetitive patterns in the row.
4. Mark the extracted sections.
5. Repeat the processing 3 to 4 for every local peak of the projection profile.

The above processing is repeated to 2-bar, 4-bar level of correlation map. In the obtained repetitive patterns, the most frequent sections are labeled as refrain[2].

## 3.2. Analysis of timbre and style of rendition

### 3.2.1 Segmentation

Multiple features that we picked up for the timbre and style of rendition are calculated using the information of attack points (onset times). Therefore we have to extract the attack point, first of all. Most of popular music composes of keyboards and guitars. The system extracts attack point based on the following procedures.

1. Choose attack candidates, where the differential of the averaged power envelop is over the threshold.
2. Add the attack candidate, for the harmonic envelope from the time-frequency map to which harmonic summation is executed. The threshold in this step is set higher compared with the step 1.
3. Exclude the attack candidates just before the release of the notes. (Heuristics)

### 3.2.2 Timber analysis

Among a few researches that aim to classify of music timber, Kitahara et al. illustrated a new method to classify the given timber[5]. The analysis that we adopt is basically based on Kitahara's method. However, Kitahara's system is not designed to deal with buzzy sound such as a distorted guitar sound. We added some features to deal with buzzy sound, and eventually prepared the following 79 features.

**(1) Features of spectrum**

- 1: gravity of total frequency components (average frequency weighed its power)
- 2: F0 power / total power
- 3 to 30: Power sum from F0 to Fn / total power (n=2,3,...,29)
- 31: Power sum of odd order harmonics / Power sum of even order harmonics
- 32 to 40: The time ratio (p%) of the harmonics whose power is over a certain threshold, to the note is on. (p=10,20,...,90)

**(2) Features of envelope**

- 41: Inclination of a line a fitted to the note attack.
- 42 to 58: Decay index. Differential of the power envelope at $t_s$ after the attack point. (t=0.15,0.20,...,0.95)
- 59 to 75: Power $t_s$ after the attack point / Max power (t=0.15,0.20,...,0.95)

**(3) Features on noise**

- 76 to 79: Non-harmonic power / Total power at the attack point (average, a standard deviation, minimum, maximum)

### 3.2.3 Analysis of rendition

We adopted the following 17 features as the indices on styles of rendition.

- 1: Number of attack points
  Total extracted number of attacks
- 2 to 5: Interval of attacks
  An average, a standard deviation, minimum, maximum of attack intervals.
- 6 to 9: Dynamics
  Average, a standard deviation, minimum, maximum of the attack power distribution.
- 10 to 13: sharpness
  Average, a standard deviation, minimum, maximum of Onset-Offset Interval / Inter-Onset Interval
- 14 to 17: register
  Average, a standard deviation, minimum, maximum of the lowest F0 of the track.

# 4. Experiments

## 4.1. Experiments on music analysis

### 4.1.1 Estimation of vocal parts

We investigated the recognition accuracy of the vocal parts for three popular music pieces in RWC-MDB-P-2001. The

recognition accuracy was 94% and the error of 6% was for when the system judged the part as the chorus part. We may say the recognition accuracy for the main vocal part was 100%.

The exclusion ratio of non-vocal parts was 92%. Most of the errors were that the system judged solo guitar played in the vocal register at prelude or interlude, as vocal melodies. These errors are not those results in the additional errors in the section analysis processing.

#### 4.1.2 Analysis of musical sections

We investigated the ability of music analysis for No.13, No.18, No.45, No.64, No.82, No.88, No.98 of the RWC MDB-P-2001. The recognition rate has been100% so far. We would like to test the method with other genres for the next step of the evaluation.

### 4.2. User evaluation of the proposed interface

In order to evaluate the effectiveness of the proposed interface, we conducted user tests using music pieces to which several mix-down designs were copied.

The degree of difficulty depends much on the difference between the source music and the target music. If the target music is similar to the source music, according to the type of musical instruments included or the number of the recorded tracks, it might be an easier task. To the contrary, when the constituents of the recorded tracks differ, we have to face to the problem of track assignment. Considering this status, we made experiments for the case where the constituents of the source and target music are similar, and for the case where they are quite different.

For the experiments, we asked two professional engineers to execute three different mixdown patterns to a popular music and gathered six typical source music. As a target music, we selected one piece No.95 from the RWC Music Database, and prepared three music pieces which were originally composed and played by an amateur band [2]. We prepared 5 target music pieces in total. Among them, one piece by the amateur band is the case where the constituents are almost the same.

30 subjects, including four members of the band, participated in the experiment. The questionnaire items consisted of the following;

1. Source identification: Whether or not the subject can identify the source music from which the mix-down design is copied to the target music.

2. Design copy degree of similarity for the subjects (valid for the subjects who made correct answer for the item 1).

3. Design quality

4. Preference for the music

5. Free description

Items 2,3,4 were answered using five-grade scoring.

[2] It is not professional, but is getting a guarantee at a stage

**Table 1. Results of evaluating the design copy when the constituents of the source and target music are similar**

| | Design Copy Degree of Similarity | Design Quality | Preference to Music |
|---|---|---|---|
| Designer A, Type 1 | 4.2 | 4.2 | 4.2 |
| Designer A, Type 2 | 4.5 | 3.3 | 3.0 |
| Designer A, Type 3 | 4.5 | 3.0 | 3.8 |
| Designer B, Type 1 | 4.3 | 3.8 | 3.7 |
| Designer B, Type 2 | 4.2 | 3.3 | 3.0 |
| Designer B, Type 3 | 4.3 | 2.5 | 2.3 |

#### 4.2.1 Test for the design copy where the constituents of the source and target music are similar

In this experiment, the subjects evaluated how well mix-down designs were copied to a music piece prepared by an amateur band, for the case where the constituents of the source and target music are similar. The experimental materials are available from the following url:
**http://ist.ksc.kwansei.ac.jp/˜katayose/MXD/**

The results of the subjects' impressions are shown in Table 1. All of the 30 subjects guessed the correct sources for all of the targets, and the evaluation of design copy degree of similarity is high. This result shows that the proposed interface works well when the number of the instruments used in the source and the target piece is similar.

Evaluation for the items of design quality and preference to the music is not so high and divergent, but these values are more than the average value of 3. Especially, most of the subjects except for the band member commented that "Two type 1s look like those of commercial mixes." Actually, these sources were designed by two designers, asked to complete it as a current, common, popular music piece.

The band members' evaluations for these items were lower than that of the other subjects. We think the reason is that the band members have a higher goal for the completed piece, and they have more strict thresholds of difference.

The results obtained by automatically copying the mix-down design seemed a little bit inferior to those by a professional mix-down engineer. Nevertheless, it was much better than what an amateur could achieve.

#### 4.2.2 Test for the design copying of the constituents of the source and target music are quite different

When the number of the tracks of the source and the target music differs, we have to select the tracks to which the design is copied. The average number of target tracks was 8, and the average number of source tracks was 50. In this experiment, we considered the functions of the instruments when assigning tracks, and prepared several combinations of track assignments. We prepared 25 pieces to each one of which, one of the 6 mix-down design were copied.

The ratio of guessing correct source depends much on the subject. The subjects who were experienced with pop music got higher marks than the other subjects. The highest mark was 83% and the lowest was 38%. The chance level of this experiment is 17%.

The average value of evaluation for the items of design quality and preference to the music were 2.4 and 2.3 respectively. These marks are lower by more than 1 degree, compared with the test for the design copying where the constituents of the source and target music are similar.

Some free comments for this experiment were the following: "I feel uneasy about some noise", "The positions of each musical instrument are not obscure."

## 5. Discussion

### 5.1. User evaluation

All of the evaluations for the items of design quality and preference to the music were higher in the experiment of copying mix-down designs where the constituents of the tracks were similar than in the case where the constituents were different. This was an anticipated result. Gathering heuristics regarding track assignment is one of the crucial subjects as a future work.

In the experiments, some subjects pointed out, "The effect of the guitar is bad", and the others pointed out "Very interesting." We gathered such an opinion on setting parameters of effectors. There are some effectors, the parameter of which is regarded better to be adjusted to the beat length of the music piece. Some band members commented that a result without parameter adjustment is interesting. It is difficult to control and guess everyone's preference to designs. When we cannot avoid the assignment of tracks, it seems better to prepare several candidates using different track assignments, and then let the user select the favorite one. Instead, providing functions for evaluating similarity of sound track may be a plausible user interface for practical use.

### 5.2. Applications of the system

The main feature of proposed system is that the system tries to capture the intuitive cognitive structure of human beings. The system provides functions for music section analysis, timber analysis and rendition analysis. Many people evaluated that design copy ability is high enough when the constituents of the source and the target are similar.

The proposed system may be used at the first stage of mixdown design done by professional engineers. The system is expected to reduce the time required for mixdown design. Also it is expected to give inspiration, especially to amateur musicians. Another application target that we are planning is using the system as a special graphic equalizer for listeners in the general public. The user of the system will be able to enjoy music, using mixdown examples, to suit their taste.

There are two big issues to be solved, in order to put the proposed system to practical use. One is the track assignment and the other is gathering examples of mixdown designs. If the music contents are circulated in a form of separated multi-tracks, these problems will be solved at a stretch. It is anticipated that the music data of each separate track will be delivered simultaneously using broadband internet. At least, it is technically possible to deliver music data with separate tracks. It is likely that standardization about track usage will be started to discuss, if business and legal circumstances are conditioned in near future. We suppose our trial will be possible from the business point of view.

Currently the primal form of music circulated in the world is stereo. From a technical point of view, source separation from stereo is a promising and challenging theme in gathering a bigger archive of music designs. Differences of recording environments influence much on the quality of sound. It is crucial to absorb the differences caused by recording environments, in the task of mixdown design copy. Source separation studies are also expected as a drastic solution to this problem.

As mentioned above, the proposed system seems to have some potential. At the same time, we have to consider subsidiary problems when using examples. When we succeed in gathering a big archive of mixdown designs, effective data search scheme will be also desired, in our approach that utilizes examples to communicate design ideas. Annotation data will be effectively used for a data search task. Sound analysis techniques described in section 3 are applicable for this goal. Another subsidary problem is a scheme for dealing with copyrights. Although we have no concrete scheme at present, we believe that the business model to circulate mixdown designs will work functionally, considering the emergence of mp3 players in which plenty of commercial music is pre-installed and the diffusion of commercial

melodies for ring tones.

## 6 Conclusion

In this paper we proposed a mix-down assistance interface that copies an existing mix-down design to the given music, and described its functions, evaluations and potential. First, we introduced the merit of referring to examples to convey a design concept for non-verbal media using language. Then this paper described the analysis of music groups, timber and rendition. Next, this paper showed the experiment on music analysis and the user evaluation of the proposed interface. Recently there is a rapid progress of automatic clustering method such as ICA, SOM[1][6]. We would like to apply these methods for our music analysis function to improve efficiency as future work. In addition we are going to continue user evaluation for the goal of business use of the proposed system.

## References


[1] R. Basili, A. Serafuni, and A. Stellato. Classification of musical genre: A machine learning approach. *Proc. International Conference on Music Information Retrieval*, pages 505–508, 2004.

[2] M. Goto. A chorus-section detecting method for musical audio signals. *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, pages 437–440, 2003.

[3] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. Rwc music database: Popular, classical, and jazz music databases. *Proc. International Conference on Music Information Retrieval*, pages 287–288, 2002.

[4] H. Katayose, K. Hirata, K. Noike, T. Harada, A. Kasao, and R. Hiraga. Toward computer-supported design for non-verbal media. *Transactions of the Japanese Society for Artificial Intelligence*, 20(2):129–138, February 2005.

[5] T. Kitahara, M. Goto, and H. Okuno. Category-level identification of non-registered musical instrument sounds. *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, pages 253–256, 2004.

[6] C. McKay and I. Fujinaga. Automatic genre classification using large high-level musical feature sets. *Proc. International Conference on Music Information Retrieval*, pages 525–530, 2004.