

## 2ch ステレオコンテンツからの分離音源を用いた リミックスとそのプリファレンス評価

小池 遥<sup>†1</sup> 石井佑哉<sup>†1</sup> 西口正之<sup>†1</sup> 渡邊貫治<sup>†1</sup> 安倍幸治<sup>†1</sup> 高根昭一<sup>†1</sup>

**概要:** 2ch ステレオコンテンツの分離音源を用いて、リミックスを簡便に行うためのアプリケーションを開発した。このアプリケーションでは、2ch ステレオ信号を構成する各分離音源のボリュームとパニングをユーザーが自由に設定出来るようにして居り、ユーザーは自分の好みのリミックスが出来る。更に本アプリケーションでは、ユーザーのリミックスに基づいてそれと類似の異なるリミックスを自動生成出来る様にした。元の2ch コンテンツ、ユーザーがリミックスしたもの、及び自動生成した類似のリミックスを混ぜて主観評価実験を行った。その結果、ユーザー自身のリミックスが統計的有意差を持って原音の2ch コンテンツよりも高い評価を得た。このことから、ユーザーは自身の好みのリミックスを類似の他のリミックスとは聞き分け、それを確かに好んでいるということが確認された。これより個人の嗜好に合わせたリミックスが手軽に行えるツールの提供には意味があると言えそうである。

**キーワード:** リミックス, 音源分離, アップミックス,

### Remixing of sound sources separated from 2 channel stereo content and its subjective evaluation

HARUKA KOIKE<sup>†1</sup> YUYA ISHII<sup>†1</sup> MASAYUKI NISHIGUCHI<sup>†1</sup>  
KANJI WATANABE<sup>†1</sup> KOJI ABE<sup>†1</sup> SHOUICHI TAKANE<sup>†1</sup>

**Abstract:** A GUI based application has been developed for the purpose of easy remixing of two channel CD content. The application allows volume and panning control of each individual sound source separated from 2 channel CD tracks. Test subjects made their own remixed version of the CD contents, and they were evaluated together with original tracks and other similar remixed versions. Subjective listening tests showed the subjects' own remixed version achieved the best score in preference, meaning that the subjects distinguish their own remixed version from others and prefer it. This implies it is meaningful to enable easy remixing of CD content by end users.

**Keywords:** remixing, sound source separated, upmixing

#### 1. はじめに

近年、映画やテレビ放送などの分野でサラウンド音声の映像コンテンツが増えており、5.1ch サラウンドスピーカーやバーチャル 3D ヘッドフォンなどのサラウンド再生システムが一般家庭にも普及しつつある。また、スマートフォンの普及に伴い、音楽関連のアプリケーションが増え、ユーザーがどこでも手軽に音楽を加工し楽しめる時代になった。しかし、このような多ch再生システムが普及しているにも関わらず、音楽配信サービスやCDで提供されている商用音楽コンテンツをはじめ多くのコンテンツは依然2chステレオで提供されているものが主流となっている。そのため、従来の2ch音楽コンテンツを高い臨場感で受聴したり、自分の好みのリミックスで受聴するには、音楽コンテンツを音源毎に分離・抽出し、サラウンド信号のチャンネル数に合わせてアップミックスを行うことや分離音をリミックスすることが望まれている。

一方、音源分離に関しては、左右チャンネルの音声スペク

トルの振幅の比から分離音源を推定する手法 [1][2]、非負値行列因子分解による手法[3]、深層学習を利用した手法[4]など多くの研究が行われているが[5]、分離後の音源の活用に関する研究は少ない。そこで本研究では、音源分離技術の応用として分離音のリミックスに着目し、ユーザー好みのリミックスを行えるツールの開発を行った。そして、そのユーザー好みの音というものが雰囲気では好んでいるのではなく、再現性を持って統計的に有意差をもって好まれているものなのかを主観評価実験により確認した。

ユーザー好みのリミックスをベースに、将来的には自分好みのアップミックスを行うことや、サラウンド再生環境を活かしたコンテンツの生成に発展させていくことができるものと考えられる。

#### 2. アプリケーションの開発

##### 2.1 リミックスアプリケーションのねらい

あるひとつの楽曲を構成音ごとに分離し、各分離音の左

<sup>†1</sup> 秋田県立大学  
Akita Prefectural University

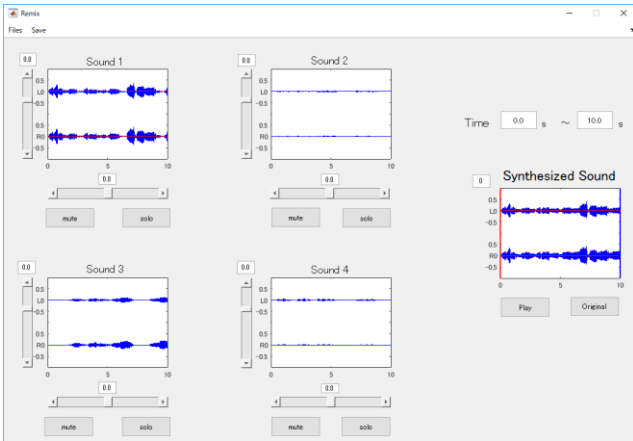


図 1 アプリケーションの概要図

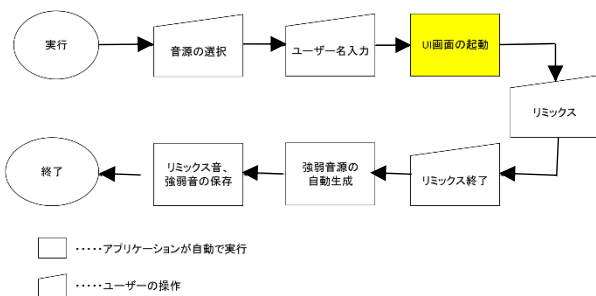


図 2 アプリケーションの実行フロー

右バランスや強度を変更することで仮想的にその定位を変更し、手軽にリミックスを行えるツールとして MATLAB 環境下で使用可能なアプリケーションの開発を行った。ただし、本研究での再生環境はヘッドフォンとして設計しているため、クロストークに関しては考慮していない。

## 2.2 アプリケーションの概要

本研究で開発したアプリケーションの UI は、図 1 のようになっており、図 2 のようなフローで実行される。

### 2.2.1 音量、音量バランスの実行について

図 1 のスライダーにより、各分離音の音量および音量バランスの調整ができる。各分離音の Lch, Rch 信号の調整はそれぞれ以下の式で実行する。ただし、 $i$  は分離音の番号、 $Vol_i[\text{dB}]$  ( $-60 \leq Vol_i \leq 24$ ) は分離音  $i$  の音量調整スライダーの調整値、 $Pan_i$  ( $-1 \leq Pan_i \leq 1$ ) は分離音  $i$  の音量バランススライダーの調整値、 $x'_{iL}(n)$ ,  $x'_{iR}(n)$  は Lch, Rch の調整前の信号、 $y_{iL}(n)$ ,  $y_{iR}(n)$  は Lch, Rch の調整後の信号である。

$$y_{iL}(n) = x'_{iL}(n) \times \sqrt{(1 - Pan_i)} \times 10^{\frac{Vol_i}{20}}$$

$$y_{iR}(n) = x'_{iR}(n) \times \sqrt{(1 + Pan_i)} \times 10^{\frac{Vol_i}{20}}$$

各分離音  $i$  についての上記の処理結果を合成すると、

$$y_L(n) = \sum_{i=1}^4 y_{iL}(n)$$

$$y_R(n) = \sum_{i=1}^4 y_{iR}(n)$$

となり、原音中の各分離音の音量及び音量バランスを調整した信号が得られる。尚、本アプリケーションで扱う分離音の数は、最大 4 とした。

### 2.2.2 リミックス後のエネルギー補正

リミックスを行うことでエネルギーが変わると、聴いたときの印象が大きく変わってしまう。そこで、リミックス前後で再生音のエネルギーが変わらないように補正を行う。原音の左右信号をそれぞれ  $x_L(n)$ ,  $x_R(n)$ 、リミックス後 (音量及び音量バランス調整後) の左右信号をそれぞれ  $y_L(n)$ ,  $y_R(n)$  とし、左右のクロストークはないものとする。

原音のエネルギーを

$$\sum_n \{x_L(n)^2 + x_R(n)^2\}$$

リミックス後のエネルギーを

$$\sum_n \{y_L(n)^2 + y_R(n)^2\}$$

として考える。

リミックス前後のエネルギーを等しくするように、リミックス後の各 ch に定数  $k$  をかけ、以下のようにおく。

$$\begin{aligned} \sum_n \{x_L(n)^2 + x_R(n)^2\} &= \sum_n \{[ky_L(n)]^2 + [ky_R(n)]^2\} \\ &= k^2 \sum_n \{y_L(n)^2 + y_R(n)^2\} \end{aligned}$$

$$k = \sqrt{\frac{\sum_n \{x_L(n)^2 + x_R(n)^2\}}{\sum_n \{y_L(n)^2 + y_R(n)^2\}}}$$

この定数  $k$  をリミックス後の出力データ  $y_L(n)$ ,  $y_R(n)$  にかけて補正を行う。

## 3. リミックス音の主観評価実験

### 3.1 主観評価実験のためのリミックス音の加工処理

本研究での主観評価実験を行うにあたり評価音を 4 音用意する。4 音は、原音、被験者がリミックスを行った音 (以下、リミックス音)、被験者のリミックスを強めた音 (以下、強リミックス音)、被験者のリミックスを弱めた音 (以下、弱リミックス音) である。リミックス音の他に強弱リミックス音を用意する目的としては、原音、リミックス音の 2 音の比較では聞き分けが容易であると考えられるためである。この 4 音の関係のイメージは、図 3 のようになる。図 3 の横軸はパンニング、縦軸は音量の調整量である。

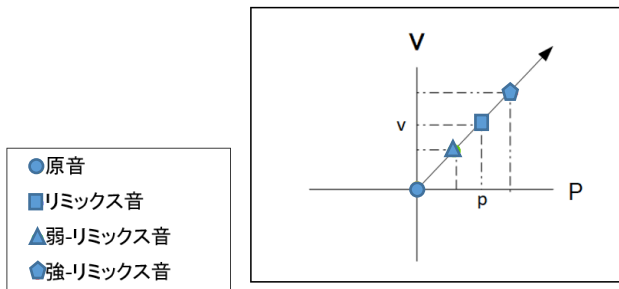


図 3 評価音 4 音の関係イメージ

### 3.2 強弱リミックス音の生成 (音量の調整)

強弱リミックス音の音量調整値は、ユーザーの音量調整値をスケールリングすることで求めた。各分離音に乗ずるゲイン  $Vol_i$ [dB] に定数をかけることで行う。本研究では、強リミックスの定数を 1.5, 弱リミックスの定数を 0.5 とした。

従って 2.2.1 節の、 $10^{\frac{Vol_i}{20}}$  の代わりに、 $10^{\frac{1.5 \times Vol_i}{20}}$  または  $10^{\frac{0.5 \times Vol_i}{20}}$  をそれぞれ各分離音にかけることになる。

### 3.3 強弱リミックス音の生成 (音量バランスの調整)

強弱リミックス音の音量バランスの調整値は、ユーザーのパンニングの調整値に対し、同じ調整量でもスライダの感度を仮想的に変更することで強弱をつけた。スライダの感度を強くする、すなわち同じ移動量でも本来のスライダーよりも値が大きく動くようにすれば強リミックス音が生成できる。また逆に、スライダーの感度を弱くする、すなわち同じ移動量でも、本来のスライダーよりも値が小さく動くようにすれば弱リミックス音が生成できる。このようなイメージで、強弱リミックス音を自動生成した。

図 4 にスライダーの感度を示す。横軸がユーザーの設定したスライダー位置  $Pan_i$ 、縦軸に感度変更後の仮想的なスライダーの位置を  $Pan'_i$  として示す。このグラフに基づいて強弱リミックスを定義する。パンニングのスライダーの

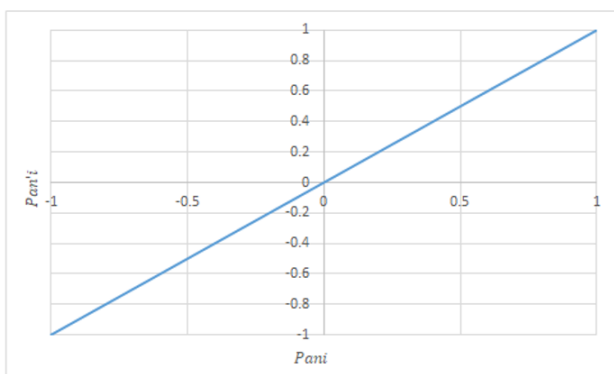


図 4  $Pan_i$  及び  $Pan'_i$  によるスライダーの感度の設定

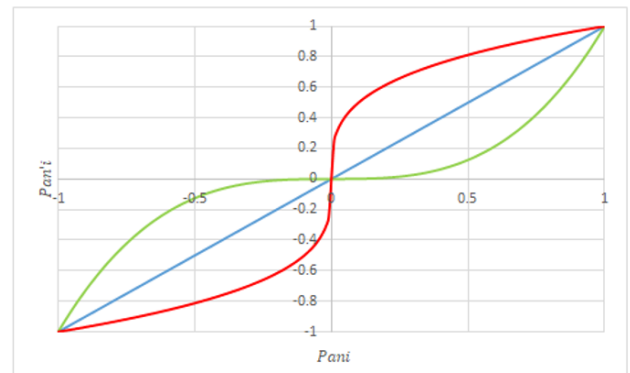


図 5 強弱パンニングのための  $Pan_i$  から  $Pan'_i$  へのマッピング

範囲である  $-1 \leq Pan_i \leq 1$  の範囲内でグラフに膨らみをつけることで、感度の変更を可能にするよう設計していく。 $-1 \leq Pan_i \leq 0$  の範囲では、 $Pan'_i = -(-Pan_i)^n$ ,  $0 \leq Pan_i \leq 1$  の範囲では、 $Pan'_i = Pan_i^n$ , とすると  $Pan_i$  と  $Pan'_i$  の関係は図 5 のようになる。ただし、図 5 は  $n=0.3, n=3.0$ , および  $n=1.0$  の場合である。 $n$  を適切に設定してスライダーの感度を変更できるため、ユーザーが行ったリミックスのパンニングの値に対し強弱をつけることができる。

### 3.4 実験の流れ

本実験は、健康成人である大学生 16 人を被験者に、開発したリミックスアプリケーションを使用し自分好みの音をリミックスしてもらい、その後本当に自分好みの音がリミックスできているかを確認するための主観評価実験を行う 2 段階の流れで行った。実験に使用したリミックスの音源は 10 秒のクラシック、ロックの 2 種類の音源を使用した。サンプリング周波数は 48kHz, ビット数は 16bit の固定小数型である。リミック操作は音を聴きながら何度でも行え、制限時間は設けなかった。

評価音の 2 音源に対し、それぞれリミックスの条件として、アプリケーション起動時の音量及び音量バランスの初期値が原音のままのもの(以下、初期値 0), およびアプリケーション起動時の音量及び音量バランスの初期値がランダム(以下、初期値ランダム)の異なる 2 パターンで実験を行った。これは初期値に依存してリミックスが偏る可能性があるためである。したがって、各被験者には計 4 パターンの実験を行ってもらった。実験の流れを図 6 に示す。

### 3.5 主観評価の方法

前節で述べたようにリミックス作業が終了した直後、自分好みの音が本当に作られているかを確認する評価実験を行った。評価方法には MUSHRA 法[6]を使用し、名称を伏せた 4 音(原音, リミックス音, 強リミックス音, 弱リミックス音)を一度に提示し比較評価してもらった。評価点は 0

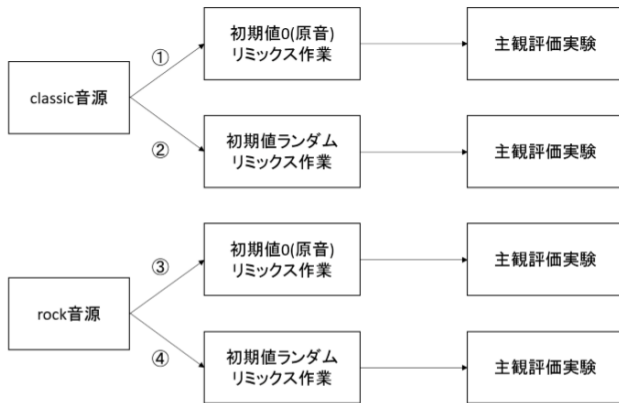


図6 実験の流れ

から 100 までの連続値であり、この値は、平均オピニオン  
評点 (Mean Opinion Score: MOS) で用いる「非常に悪い」  
から「非常に良い」までの 5 段階評価を拡張したものであ  
る。被験者はすべての音源を自由に切り替えて何回でも聴  
くことができる。自分の 1 番好みの音に 100 点を与えても  
らい、他の 3 音源についても直感的に採点してもらった。

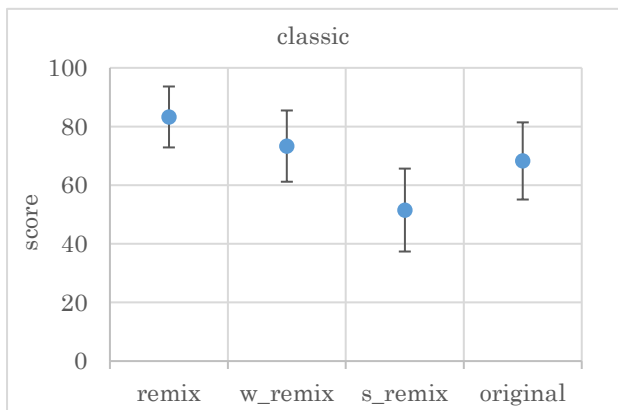


図7 クラシックの初期値0における主観評価実験結  
果の平均値と 95%信頼区間

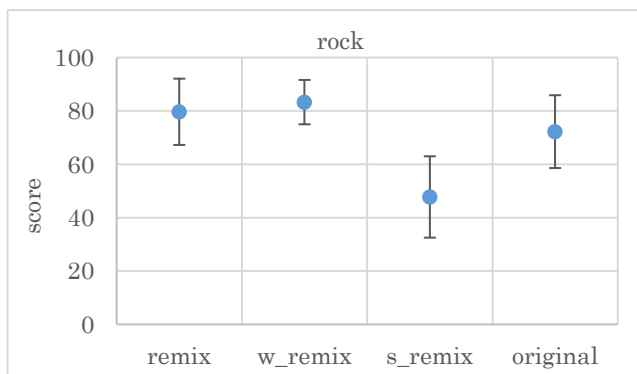


図8 ロックの初期値0における主観評価実験結果の  
平均値と 95%信頼区間

## 4. 実験結果

### 4.1 評価実験結果

図 7,8,9,10 はそれぞれクラシックのリミックス条件が初期  
値 0, ロックのリミックス条件が初期値 0, クラシックのリ  
ミックス条件が初期値ランダム, ロックのリミックス条件  
が初期値ランダムの主観評価実験結果の平均値と 95%信  
頼区間をプロットしたものである。remix がリミックス音,  
w\_remix が弱リミックス音, s\_remix が強リミックス音,  
original が原音である。

### 4.2 評価実験の考察

#### 4.2.1 異なるリミックス条件, 異なる音源の合算

リミックス条件(リミックスの初期値)に関わらず、共通  
の得点傾向がみられるか検討するために、各音源毎に異な  
るリミックス条件のデータを合算した。ただし、エラーバ  
ーは 95%信頼区間であり、データ数は楽曲毎における被験  
者の合算になるため 32 である。その結果が以下の図 11,12  
のようになり、図 7,9 あるいは図 8,10 に示す結果よりも

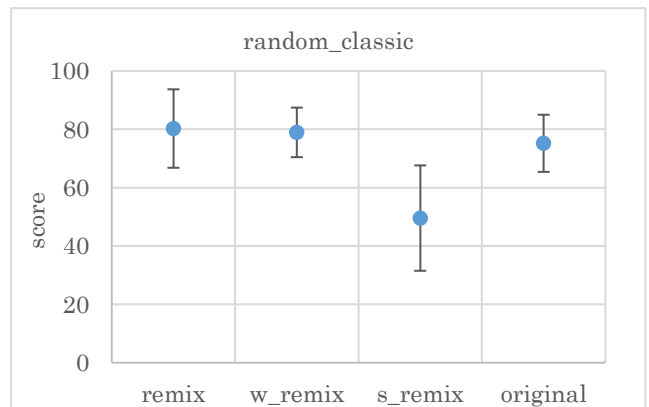


図9 クラシックの初期値ランダムにおける主観評価  
実験結果の平均値と 95%信頼区間

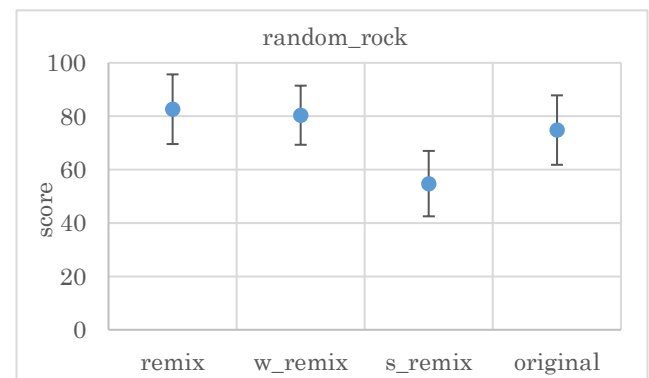


図10 ロックの初期値ランダムにおける主観評価  
実験結果の平均値と 95%信頼区間

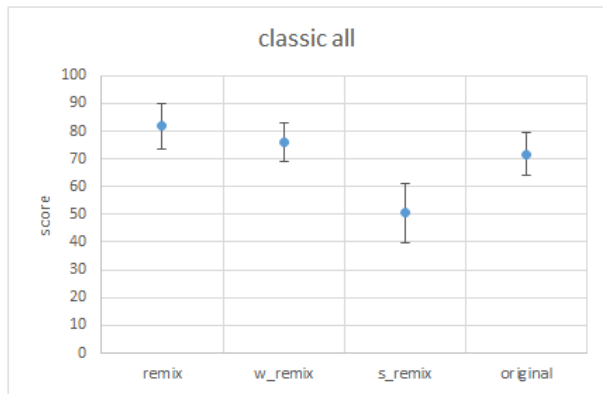


図 11 クラシックの異なるリミックス条件の合算結果の平均値と 95%信頼区間

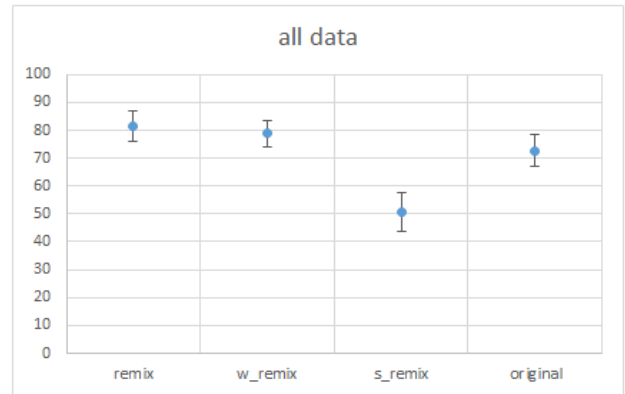


図 13 異なる楽曲, 異なるリミックス条件の合算結果の平均値と 95%信頼区間

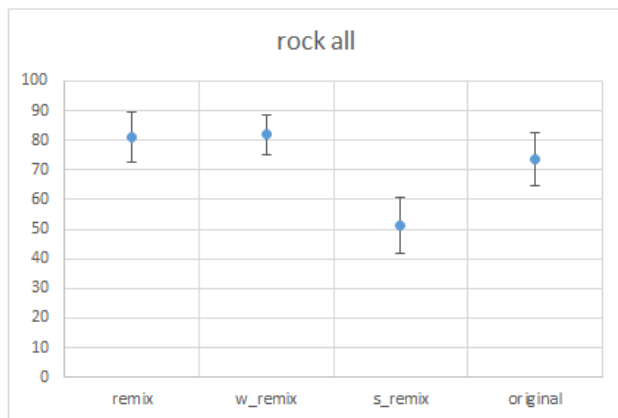


図 12 ロックの異なるリミックス条件の合算結果の平均値と 95%信頼区間

95%信頼区間が縮まっていることがわかる。つまり、リミックス条件に関わらず共通の傾向がみられるといえる。

更に楽曲, リミックス条件に関わらず, 共通の得点傾向がみられるか検討するため図 11,12 に示したプロットを算出するのに用いたデータを全て合算した。その結果を図 13 に示す。ただし, エラーバーは 95%信頼区間であり, データ数は楽曲毎における被験者数の合算のため 64 となる。この結果より, 更にエラーバーが縮まっていることがわかり, 異なる音源あるいは異なるリミックス条件であっても同一の傾向があることが確認された。

#### 4.2.2 有意差検定

本実験で得られたリミックス音と原音の得点の平均値と 95%信頼区間の結果について, 両側 t 検定を行った[7]。その結果を表 1 に示す。これにより, 全条件のデータの合算である all においてリミックス音と原音の得点差に有意差ありの判定となった。

表 1 各条件でのリミックス音と原音の得点間の t 検定の結果(両側検定)

サンプル	データ数	t値	有意水準5%のt値 (t分布表より)	有意差
classic	16	1.907	2.042	なし
random_classic		0.859		
rock		0.649		
random_rock		0.904		
classic all	32	1.827	1.999	なし
rock all		1.265		
all	64	2.182	1.979	あり

## 5. まとめ

本研究では, 2ch 音楽信号の音源分離技術を用いユーザー好みのリミックスが手軽に行えるツールとしてアプリケーションの開発を行い, そのリミックス音が本当に自分の好みの音になっているかを評価すること, また, エンドユーザーがリミックスを行う意義の確認を研究目的とした。

主観評価実験の結果より, 自分がリミックスした音が原音よりも平均点が高いことが分かった。リミックス条件に関わらず共通の得点傾向がみられるか検討するために, 各音源毎に異なるリミックス条件のデータをまとめて集計した。その結果, 異なるリミックス条件のデータの合算前の結果よりも評価実験結果の平均値の 95%信頼区間が縮まっていることが分かり, 得点の傾向が同一であることが確認できた。更に, 楽曲, リミックス条件に関わらず共通の得点傾向があるか検討するために, 全ての楽曲, 全てのリミックス条件でのデータをまとめて集計した。その結果, まとめて集計したものはどの条件よりも 95%信頼区間が

縮まっていることがわかり、音源が異なっても、あるいはリミックス条件が異なっても同一の得点傾向があることが確認できた。この結果を両側  $t$  検定で有意差検定を行ったところ、全条件を合算したものについてはリミックス音と原音の評価結果に有意差ありの判定となった。以上の結果より、ユーザー好みのリミックスが存在すること及びユーザーが自分好みのリミックスを聞き分けていることが確認できた。

今までは、ユーザー好みのリミックスというものが、面白いアプリケーションとして語られることはあったが、そのような個人個人の好みというものが本当に意味のあるものなのかが確認されることはなかった。しかし本研究の結果から、そのような個人の好みは確かに存在して、人はそれを聞き分けていることが統計処理によって確認されたといえる。以上、ユーザー好みのリミックスが行えるアプリケーションの開発を行い、ユーザーが自分の好みを聞き分けているかを確認し、エンドユーザーがリミックスを行う意義を確認することができた。

## 参考文献

- [1] Masayuki Nishiguchi, Ayumu Morikawa, Yuya Ishii, Kanji Watanabe, Koji Abe, and Shouichi Takane, “Sound Source Separation and Synthesis for Audio Enhancement based on Spectral Amplitudes of Stereo Signals”, *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 8, No. 6, pp. 1362-1371, November (2017)
- [2] 石井佑哉, 西口正之, 渡邊貫治, 高根昭一, 安倍幸治, “高臨場感再生のための 2ch ステレオ信号からの音源分離手法の検討,” *日本音響学会講演論文集*, 711-712 (2017)
- [3] C. Fevotte, N. Bertinand J.-L. Durrieu, Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis, *Neural Computation*, vol. 21, no. 3, pp. 793-830, 2009.
- [4] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, P. Smaragdis, Deep learning for monaural speech separation, *ICASSP*, 2014.
- [5] 牧野昭二 他, 電子情報通信学会『知識の森』 (<http://www.ieice-hbkb.org/>) 2 群-6 編-2 章 音源分離
- [6] International Telecommunication Union, “Method for the subjective assessment of intermediate quality level of audio systems”, *Recommendation ITU-R BS.1534-3*, 2015.
- [7] 田中 敏夫・山際勇一郎, *ユーザーのための教育・心理統計と実験計画法*, 教育出版株式会社 (1997)