

令和 2 年 6 月 3 日現在

機関番号：12612

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K00229

研究課題名(和文) 時間周波数平面を自在に操るハイパーミキサーの研究と研究者用音源データベースの構築

研究課題名(英文) Hyper mixer manipulating time-frequency plane and construction of sound source database for researchers

研究代表者

高橋 弘太 (Takahashi, Kota)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：10188005

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究は、過去に行ってきた「効率的視聴法」の研究と、「時間周波数平面を用いるミキシング法」の研究の2つの研究成果をあわせて、人間にとってより聴きやすい音声の混合法を研究するものである。具体的には、音成分のゲインと位相を調整するだけでなく、時間軸方向や周波数軸方法に木目細かく自在に成分移動させることによって、より良いミキシングが可能であるかを調査し、その有効性を示すのが本研究の目的である。移動量を客観的に決める指標の提案を行い実験した結果、時間方向や周波数方向の成分移動が有効であることが確認された。あわせて、研究者用に公開している音源データベースを拡充して整備しより使いやすくすることができた。

研究成果の学術的意義や社会的意義

高齢化社会においては、音で伝える情報が、できるだけ多くの人に確実に届くように配慮する必要がある。一方、自動車の車内で大音量の音楽を楽しみながらカーナビの音声を聞きたいというように、同時に複数の音を人間に提示するとき、いかに正確に情報を伝えるかという問題もあり、これは年齢によらずに多くの国民が問題意識を持っていることである。本研究は時間周波数平面上での成分移動によって音を混合することで、より正確に音の情報を伝えることができることを示したものである。また、この研究のために収録された音源データベースは研究者が自由に利用できるようにインターネット上で公開されており、研究者に利用してもらっている。

研究成果の概要(英文)：This research combines the results of two researches, "Efficient playback system" and "Mixing method using time-frequency plane", which have been conducted in the past. The aim of this research is finding an appropriate method of mixing speech that is easier for humans to hear. Specifically, not only adjusting the gain and phase of the sound component, but also moving the component finely and freely in the time axis direction and frequency axis method, we investigated whether better mixing is possible. As a result of experimenting by proposing an index that objectively determines the amount of movement, it was confirmed that component movement in the time direction and frequency direction is effective. At the same time, we were able to expand and maintain the sound source database open to researchers for easier use.

研究分野：信号処理

キーワード：時間周波数平面 音声データベース

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

テレビ放送の多チャンネル化と、ネットによる番組配信により、視聴覚コンテンツ量は爆発的に増大している。過去に行われたある調査機関の調査結果によれば、現代人がレコーダに録り溜めている番組の平均数は37.3番組であり、1年で4割の伸びがあるという。このような背景のもと、究極の高速再生を目指してはじまった「効率的視聴法」の研究が本研究の第一の源流である。

「効率的視聴法」の研究では、局所的な再生速度と再生加速度を制約条件とし所要再生時間を目的関数とする最適化問題を解く方法を提案し研究を進めてきた。

一方、本研究の第二の源流として、ミキシングの研究があげられる。この研究の根幹は、ソース信号を時間周波数平面に展開し、平面同士の非線形操作としてミキシングを行う枠組みである。具体的には、平面上の各所で任意のゲインと位相調整を行う仕組みとしてミキシングをとらえ、あたかも信号同志を互いに編み込むような木目細かい混合を実現し、ミキシングエンジニアが経験と勘を駆使して初めて成し得ていたミキシングを自動的に実現する

また、本研究を進めるにあたっては音源のデータベースが必要である。本研究の開始前より本研究の研究グループでは音源のデータベースを構築し、研究者がだれでも自由に利用できるように公開してきていた。本研究を進めるにあたってこのデータベースを利用することはもちろん、本研究や類似の研究が円滑に行えるようにデータベースを整備することが重要であるとの状況にあった。以上が研究開始当時の背景である。

2. 研究の目的

以上の状況のもと、音成分のゲインと位相を調整するだけでなく、時間軸方向や周波数軸方法に木目細かく自在に成分移動させることによって、より良いミキシングが可能であるかを調査し、その有効性を示すのが本研究の目的である。あわせて、研究者用に公開している音源データベースを拡充して整備しより使いやすくし、利用者からくる問い合わせや要望に応じていくのも本研究の目的である。

3. 研究の方法

まず、時間周波数平面上で成分を移動することの有効性を調べた。これは、時間周波数平面上での成分移動を信号処理に知見を持った技術者が手動で行い、そもそも成分移動が有効であるか否かを事前に調べておくために行った。この実験で有効性を確認した後、指標に基づいて自動で成分移動を行う方法に取り組み、指標としてどのようなものが良いかを検討し、提案する手法を使って成分移動を行った後、それを

被験者に聴いてもらって成分移動の有効性を確認するという研究方法をとった。音声データベースについては研究で使いながら不十分なところを拡充し、また、利用者からの意見を聞いて改良していくという方法をとった。

4. 研究成果

(1) まず、時間周波数平面上で成分を移動することの有効性を調べた。データとしては本研究で構築している音源データベース SRV-DB を用いた。特に今回はカーナビ音声を収録し、右や左などの重要語句が聞きとりやすいかをチェックした。カーナビの場合には大きな時間移動を行うとナビのタイミングを逃してしまう可能性があるため、僅かな時間移動によって明瞭度を改善する必要があるが、それが可能か否かも焦点となる。また、車内でリスニングしている BGM 音楽に対してスペクトル的に衝突しないナビ音生の時間周波数成分であったほうが聞き取りやすいのではないかと考えられたので、これについても調べた。実験に用いた信号の時間周波数面を図1に示す。左が時間移動で右が周波数移動である。図1の時間周波数平面に対して、MOS値(主観評価量)で評価した結果を図2に示す。成分移動することで聞き取りやすさが向上したことがわかる。

(2) 次に指標に基づいて自動で成分移動を行う方法に取り組み、指標としてどのようなものが良いかを検討した。研究当初は時間

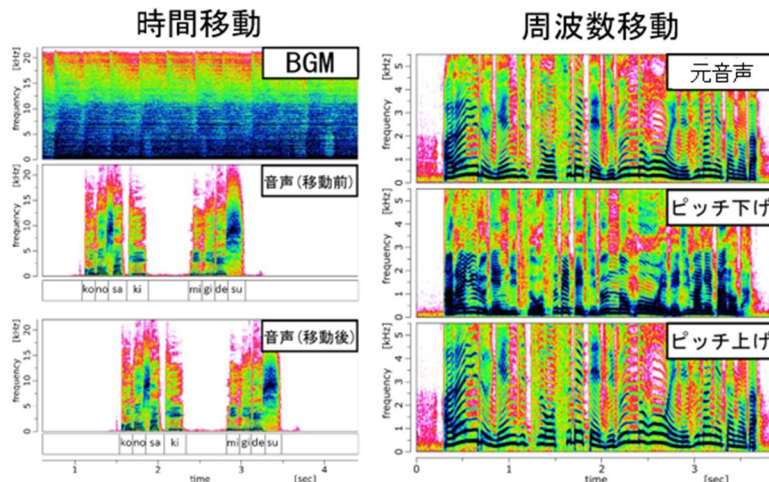


図1 時間移動/周波数移動の実験に用いた信号の時間周波数平面

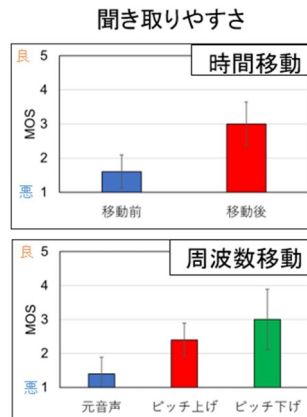


図2 成分移動の有効性

周波数平面の各点で周波数重み付けSNRを用いる手法を検討したが、最終的には、音声知覚にはスペクトル遷移が極大となる部分が重要という研究成果[1]を根拠として、遷移極大付近でのSNRを利用するという方法を提案するに至った。図3に音節毎の指数に基づいて音節を時間移動する例を示す。

この指標の有効性を精密に確認する目的で、音声データベースに「一文字違いの語句の発声音声」のカテゴリを追加して収録した。収録した音声のリストの一部を表1に示す。例えば、「貝に」と「紀伊に」は、の冒頭の「か」と「き」の1文字の違いである。このような音声を考えて

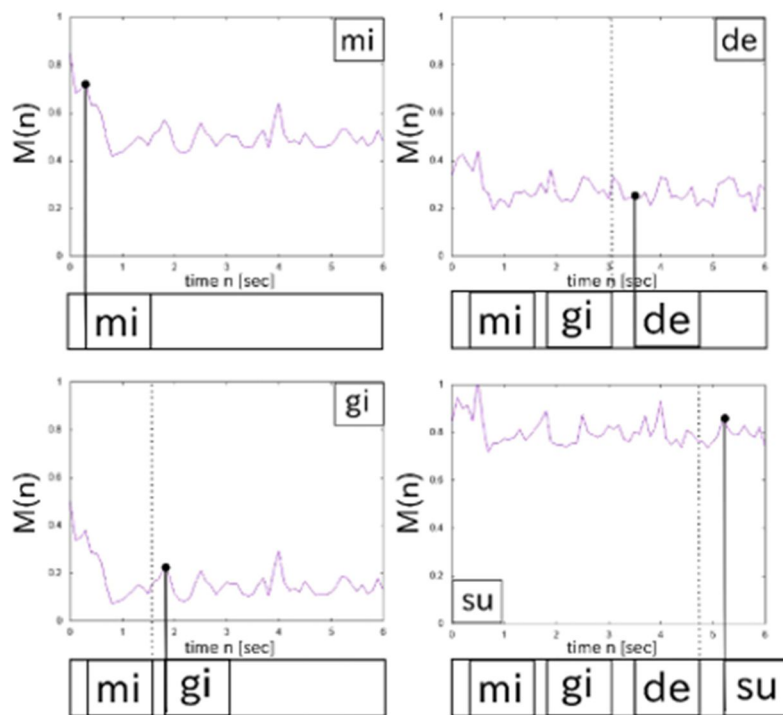


図3 音節毎の指数に基づいて音節を時間移動した例。

1. 発話のプロフェッショナルによるオリジナル原稿（一文字違い文章）の読み上げ

話者名：PF02					同話速一括ダウンロード
自然な話速 (5.01)	選択	選択	選択	選択	この行をダウンロード

2. 発話のプロフェッショナルによるオリジナル原稿（カーナビ文章）の読み上げ

話者名：PF02				同話速一括ダウンロード
自然な話速	選択	選択	選択	この行をダウンロード

3. 発話のプロフェッショナルによる編集手帳（読売新聞）の読み上げ

				同話速一括ダウンロード
自然な朗読 (自由話速)	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
6.73 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
8.00 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
9.51 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
11.31 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
13.45 [モーラ/秒]	ダウンロード	ダウンロード	ダウンロード	この行をダウンロード
同話速一括ダウンロード	この列をダウンロード	この列をダウンロード	この列をダウンロード	一括ダウンロード (約80MB)

図4 公開している音源データベース SRV-DB の一部（冒頭部分）

第1回収録	メロディアス 00 00 やや ち 音楽：大脇 渉	やや効果音的 00 00 やや ち 音楽：大脇 渉
男性：VM00 女性：VF00	試験/ダウンロード	試験/ダウンロード
自然な話速 (7.50)		

図 5 公開している音源データベースの一部

リストを作り、プロの発話者（アナウンサー）に朗読してもらって編集し音声データベース SRV-DB に加えた。また、音声の強度の大小によってどの音声であるかを被験者から見抜かれないように、音量バランスも精密に調整してデータベースに加えている。この音声を使って聴き取り正解率のテストを行ったところ、相関係数が 0.39 となり、この指標で成分移動を行ったほうが聴き取り正解率が上昇するという結果を得ることができた。

(3) 構築して公開している音源データベースの一部を図 4 と図 5 に示す。図 4 が上に述べたカーナビ音声と 1 文字違い音声である。音楽(BGM)との混合の研究のために図 5 のように音楽も音源として追加し、様々な研究に応じられるように配慮している。

<引用文献>

[1] 古井貞熙, 音声知覚研究とその時間情報処理への応用, 日本音響学会誌, 42 巻, 12 号, 1986 .

表 1 一文字違い発声の例

単語リスト(27単語)		
貝に	ご飯	秋である
紀伊に	五品	雨期である
杭に	五分	駅である
刑に	五篇	四季である
恋に	五本	席である
差異に	五点	劇である
鯛に	五年	時期である
地位に	五万	簿記である
二位に	五問	
壘に		

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 若林拓弥、高橋弘太
2. 発表標題 音響信号混合のための疑似複素正弦波による信号表現法とその評価
3. 学会等名 日本音響学会2019年春季研究発表会
4. 発表年 2019年

1. 発表者名 木下寛紀, 高橋弘太
2. 発表標題 大規模な会場におけるライブ演奏のためのスマートミキサー
3. 学会等名 電子情報通信学会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

研究者用に公開している音源データベース (SRV-DB) http://www.it.cei.uec.ac.jp/SRV-DB/

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----