

擬態語・擬音語音声入力を用いた LINE 風スタンプ描画システムの動作分析

雨宮香織^{†1} 相川清明^{†2}

擬態語・擬音語による音声入力を用いてスタンプを描画させる LINE 風スタンプ描画インターフェースを作成した。描画システムは音声認識による音素列情報と音声認識で用いていない大局的音響特徴を併用する。この2種の情報の使用法により2種類の検索方法を比較する。2種類の方法について様々な言語表現に対する動作について分析した。

Analyzing the performance of LINE-like stamp retrieval systems using onomatopoeic voice input

KAORI AMEMIYA^{†1} KIYOAKI AIKAWA^{†2}

This report describes the performance of LINE-like stamp retrieval systems using onomatopoeic voice input. The system uses phoneme recognition results and acoustic features which were not used in speech recognition. This report compares two variation of systems for various onomatopoeic expressions.

1. はじめに

最近、コミュニケーションを行う為のツールとして LINE などのアプリケーションが利用されている[1]。これらのアプリケーションの機能の中には、文字を使用したチャットの他に、スタンプと呼ばれるイラストによる情報表現が使用されている。これらのスタンプを音声入力によって検索し、表示するインターフェースを作成してきた。

スタンプには決まった言語表現は無い為、擬態語・擬音語を用いて音声入力を行い、音響特徴分析や音素列検索を利用してスタンプを検索する方法が行われてきた[2][3]。

擬態語・擬音語は個人により表現や発声のバリエーションがあり、スタンプを正しく検索させるにはバリエーションに対応させる必要がある。本報告では、擬態語・擬音語による音声入力を用いてスタンプを描画させる LINE 風スタンプ描画インターフェースを作成した。そのシステムの様々な言語表現に対する動作について分析し、多くの人がスタンプを正しく検索できる方法について提案をする

2. 従来の方法

2.1 従来のスタンプ検索方法

従来の擬態語・擬音語音声入力を用いてスタンプを描画する方法では、擬態語・擬音語音声の複数の大局的テンプレートと対応しているかをベクトル空間法による類似度を用いて音響特徴を分析し、ベクトルの類似度が重複してしまった場合にのみ音素列検索によって再度検索が行い、スタンプを判断する方法である。

2.2 問題点

従来システムでは、音響特徴分析ではスタンプを決定できない場合以外は音素列検索は行われない為音響特徴分析で誤ったスタンプが表示されることがあった。更に、音響特徴分析で正しいスタンプが選ばれても類似度が重複してしまい、音素列検索で誤ったスタンプが表示されてしまう事もあった。

擬態語・擬音語にはスタンプと同様に決まった発声方法は無い為、個人によって表現や発声のバリエーションがある。また、擬態語・擬音語を話す際にあまり抑揚を付けずに音声入力を行う人もいる。この事から音響特徴分析だけでは結果に差が出る場合があった。

2.3 方法

音響特徴分析を中心にスタンプを判断していた検索方法を改良した。音素列検索も音響特徴分析と同じタイミングで行い、それぞれの結果を使いスタンプを判断する。音響特徴分析と音素列検索、両方の結果を用いることで個人の表現や発声のバリエーションに対応し、スタンプを正しく判断することが可能である。

3. 音響特徴分析と音素列検索

被験者である男女10名に用意した10種類の擬態語・擬音語を5回ずつ音声入力してもらい、音響特徴分析と音素列検索をそれぞれ動かして各擬態語・擬音語の正答率を調べ、擬態語・擬音語によってどのような特徴があるかを比較した。音響特徴分析にはベクトル空間法による類似度を用いる。音声入力してもらった擬態語・擬音語は表1のとおりである。

表1 音声入力評価用の10種類の擬態語・擬音語

Table1 Onomatopoeic words for system evaluation

どんっ	きらっ	すぱっ
ざーざー	ちくちく	しとしと
がーん	しゃきーん	どっかーん
しーっ		

3.1 各擬態語・擬音語の音響特徴分析について

各擬態語・擬音語の音響特徴分析について、実験を行った結果、個人により表現や発声のバリエーションに差が少ない擬態語・擬音語と、差が大きい擬態語・擬音語があった。結果は表2、表3のとおりである。

表2 表現や発声の差が少ない擬態語・擬音語

Table2 Onomatopoeic words of small utterance variation.

どんっ	ちくちく	がーん
しーっ	すぱっ	きらっ

表現や発声の差が少ない擬態語・擬音語は「どんっ」「きらっ」「すぱっ」「ちくちく」「がーん」「しーっ」の6種類であった。これらの擬態語・擬音語は、ほとんどの被験者が同じ発声方法で音声入力をした。

表3 表現や発声の差が多い擬態語・擬音語

Table3 Onomatopoeic word with large utterance variation.

ざーざー	どっかーん	しゃきーん
------	-------	-------

表現や発声の差が大きい擬態語・擬音語は「ざーざー」「どっかーん」「しゃきーん」の3種類であった。

これらの擬態語・擬音語は被験者全員が同じ発声方法でなく、個人により様々な発声方法が見られた。表現や発声の差が多かった擬態語・擬音語に対して雑音性とスペクトル変化をプロットし、違いを分析する。また、雑音性の分析にはゼロクロスカウントを用いている。

3.1.1 「ざーざー」の表現・発声方法の違い

「ざーざー」は2種の発声方法があった。第1の発声方法は、「ざ」の部分強調し、声のトーンを上げるような発声方法である。10名の被験者のうち7名がこの発声方法であった。第2の発声方法は「ざあざあ」のような発声方法で、後に続く「あ」の部分も強調し、声のトーンを下げるような発声方法である。10名の被験者のうち3名がこの発声方法であった。

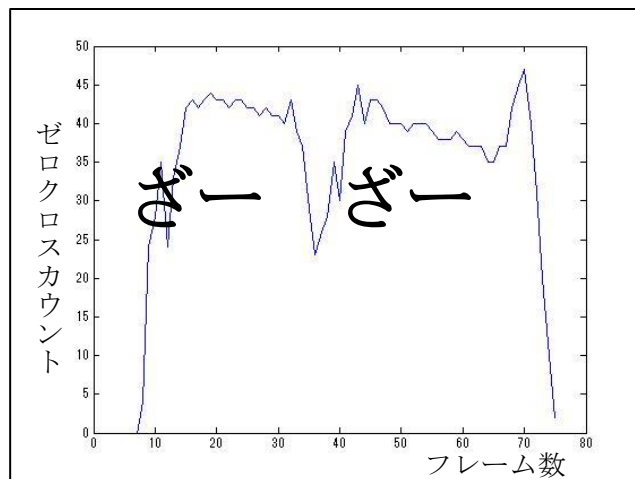


図1 第1の発声方法のゼロクロスカウント

Figure1 An example of zero cross count with the first utterance style.

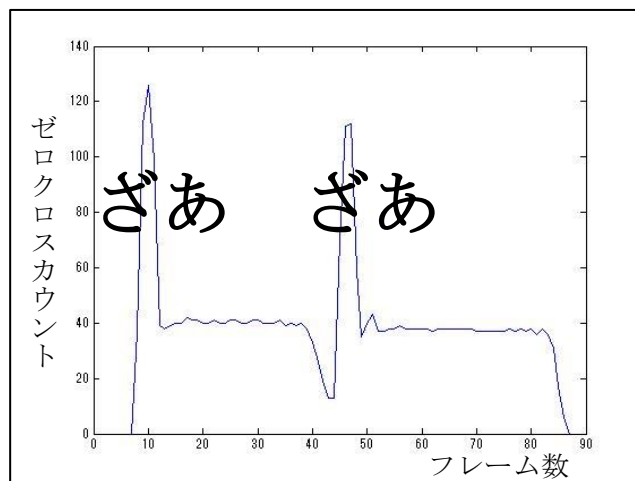


図2 第2の発声方法のゼロクロスカウント

Figure2 An example of zero cross count with the second utterance style.

図1と図2のゼロクロスカウントを比較すると、第1の発声方法の声のトーンを上げるように発音する「ざーざー」では、全体的にゼロクロスカウントは高い結果に対して、第2の発声方法の声のトーンを下げるように発音する「ざーざー」では、「ざ」の部分のゼロクロスカウントが強調して高くなり、音引きの部分のゼロクロスカウントの変化は少ない。

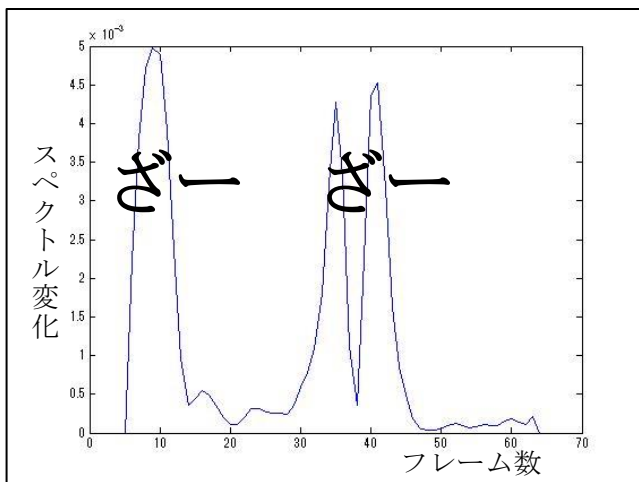


図3 第1の発声方法のスペクトル変化

Figure3 An example of spectrum change with the first utterance style.

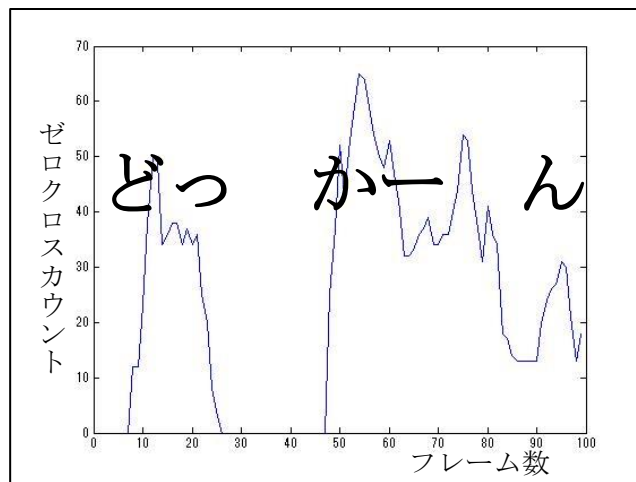


図5 第1の発声方法のゼロクロスカウント

Figure5 An example of zero cross count with the first utterance style.

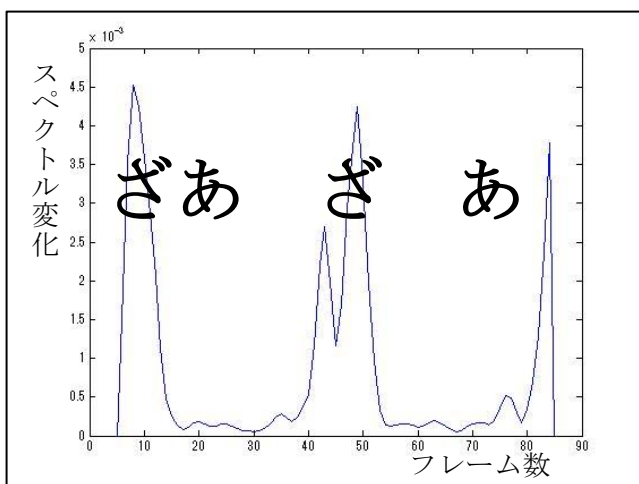


図4 第2の発声方法のスペクトル変化

Figure4 An example of spectrum change with the second utterance style.

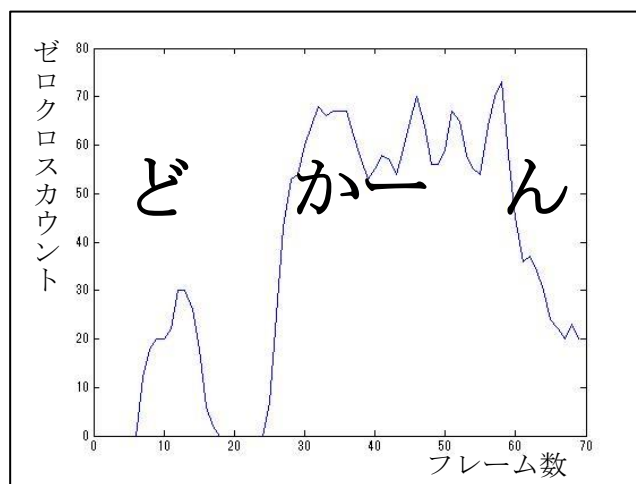


図6 第2の発声方法のゼロクロスカウント

Figure6 An example of zero cross count with the second utterance style.

図3と図4のスペクトル変化を比較すると、両方ともスペクトル変化は同じように見えるが、第2の発声方法の声のトーンを下げるように発音する「ざーざー」では、後半の部分のスペクトル変化が大きい。

3.1.2 「どっかーん」の表現・発声方法の違い

「どっかーん」は2種の発声方法があった。第1の発声方法は「どっ」の部分強調する発声方法である。10名の被験者のうち8名がこの発声方法であった。第2の発声方法は「どかーん」と発音するように「かーん」の部分強調する発声方法である。10名の被験者のうち2名がこの発声方法であった。

図5と図6のゼロクロスカウントを比較すると、第1の発声方法である「どっ」の部分強調した発声方法では、「ど」の部分のゼロクロスカウントの変化が大きい結果に対して第2の発声方法の「どかーん」と発音したような発声方法では、「ど」の部分のゼロクロスカウントの変化は大きく、その後の「かーん」の部分のゼロクロスカウントも継続して大きい結果となった

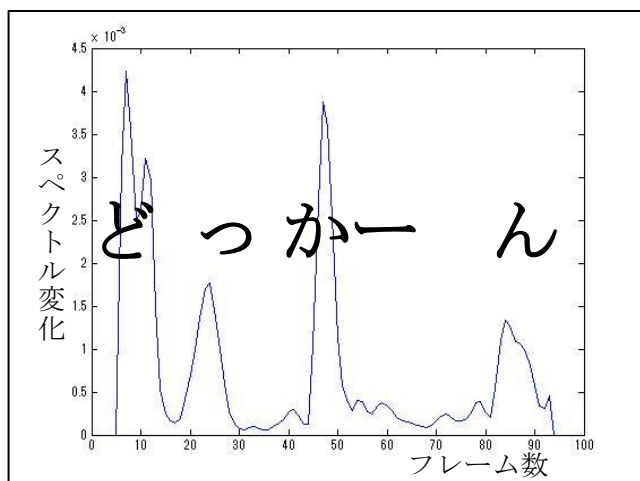


図7 第1の発声方法のスペクトル変化

Figure7 An example of spectrum change with the first utterance style.

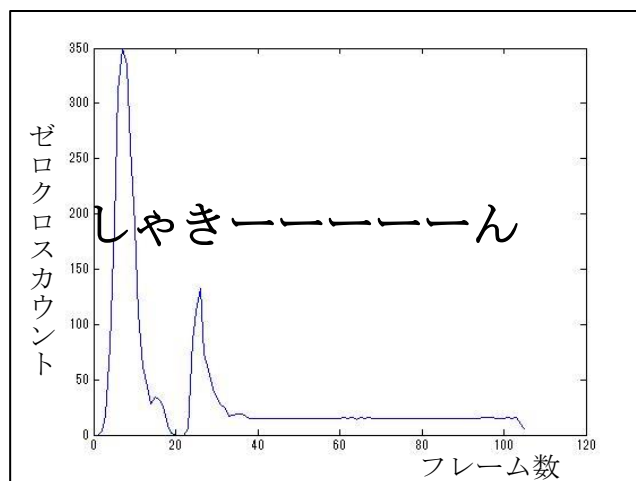


図9 第1の発声方法のゼロクロスカウント

Figure9 An example of zero cross count with the first utterance style.

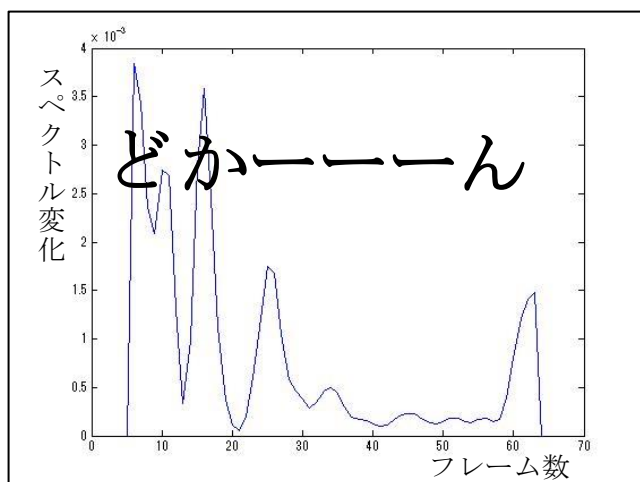


図8 第2の発声方法のスペクトル変化

Figure8 An example of spectrum change with the second utterance style.

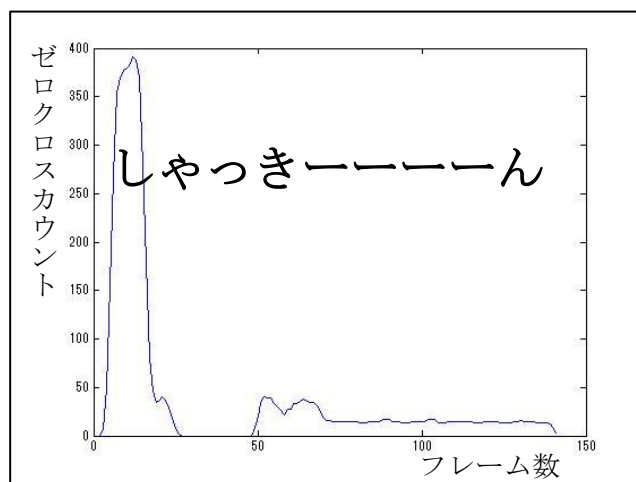


図10 第2の発声方法のゼロクロスカウント

Figure10 An example of zero cross count with the second utterance style.

図7と図8のスペクトル変化を比較すると、変化に大きな差が出た部分は前半であった。「か」の位置に大きな差が出ている。

3.1.3 「しゃきーん」の表現・発声方法の違い

「しゃきーん」は2種の発声方法があった。

第1の発声方法は「しゃきいん」などのように「きーん」の部分強調する発声方法である。10人の被験者のうち5名がこの発声方法であった。第2の発声方法は「しゃつきーん」のように、「しゃ」の部分強調させて発声する発声方法であった。10人の被験者のうち5名がこの発声方法であった。

図9と図10の雑音性を比較すると、第1の発声方法である「きーん」の部分強調した発声の雑音性では、「き」の部分で雑音性が多いことに対して、第2の発声方法である「しゃつきーん」と発声したような雑音性では「き」の部分はゼロクロスカウントの変化が少ない。

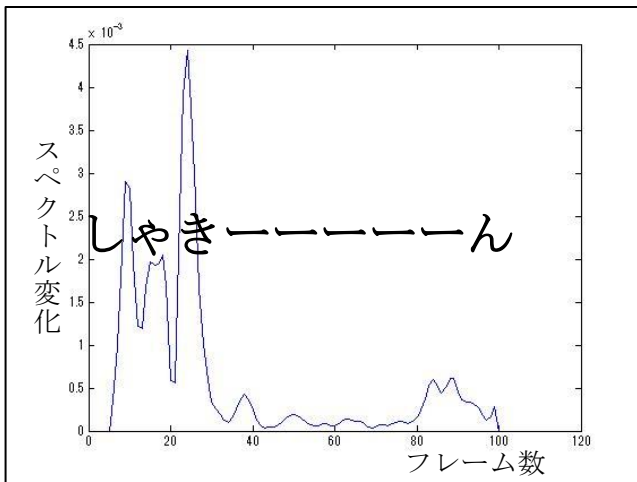


図 11 第 1 の発声方法のスペクトル変化
 Figure11 An example of spectrum change with the first utterance style.

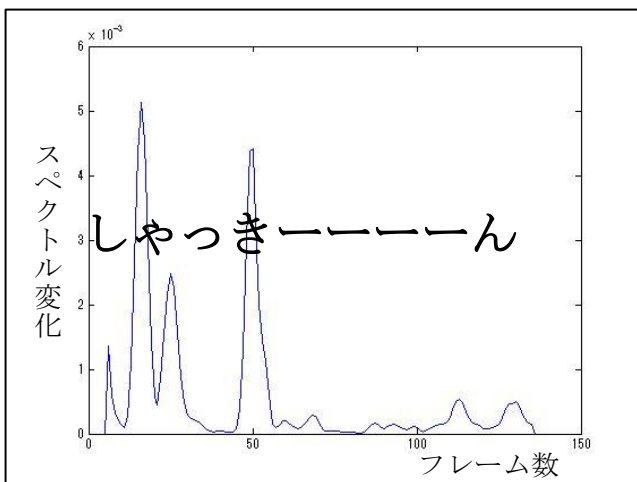


図 12 第 2 の発声方法のスペクトル変化
 Figure12 An example of spectrum change with the second utterance style.

図 11 と図 12 のスペクトル変化を比較すると、「どっかーん」の場合と同様、変化に大きな差が出た部分は前半であった。「か」の位置に大きな差が出ている。「きーん」の部分を強調した発声のスペクトル変化では「しやき」の部分のスペクトル変化が前半に多く見られているが、「しやつきーん」と発声したようなスペクトル変化では「しや」と「き」のスペクトル変化は前半、中盤と別れている。

3.2 各擬態語・擬音語の音素列検索について

各擬態語・擬音語について音素列検索を行った結果、音素列検索で正しく判断されにくく、誤った判断がされてしまう擬態語・擬音語があった。音素列検索で正しく判断されにくい擬態語・擬音語は表 4 のとおりである。

表 4 音素列検索で正しく判断されにくい擬態語・擬音語

Table4 Search by phoneme is Difficult onomatopoeia

きらっ	ちくちく	しーっ
-----	------	-----

音素列検索で正しく判断されにくい擬態語・擬音語は「きらっ」「ちくちく」「しーっ」の 3 種類であった。「きらっ」は音素列検索では 40% の確率で「いらっ」という擬態語・擬音語で認識されてしまう結果になった。「ちくちく」は音素列検索では 90% の確立で「しくしく」という擬態語・擬音語で認識されてしまう結果になった。「しーっ」については音素列検索では被験者全員が正しく検索することが出来なかった。また、音素列検索では擬態語・擬音語を音声入力する際に抑揚を強く付けすぎてしまうと、正しい判断がされにくい結果となった。特に「きらっ」は抑揚を付けやすい擬態語・擬音語である為、個人によって音素列検索の結果に差が生じた。

4. システムの特徴・描画方法

上記分析に基づき、2 種の方法を提案する。音響特徴分析では、従来の研究と同様にベクトル空間法による類似度を用いる。メインプログラムには MATLAB R2014a を利用し、音声認識には Julius を用いている。

4.1 音響特徴分析と音素列検索の類似度を比較する方法

第 1 のスタンプ検索・描画方法では、音響特徴分析による類似度と音素列検索による類似度を比較し、可能性が高い方のスタンプを表示する方法となっている。擬態語・擬音語にアクセントをつけて音声入力をする人は主に音響特徴分析により判断される事が多く、擬態語・擬音語にあまりアクセントを付けずに音声入力する人は音素列検索により判断される事が多い事になる。

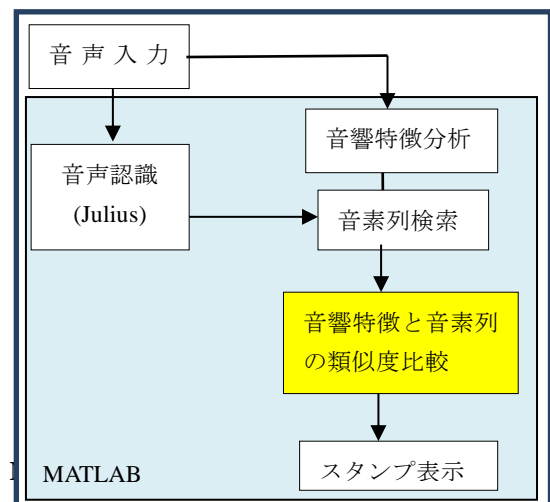


図 13 音響特徴と音素列の類似度を比較する方法

Figure13 Stamp retrieval system1.

4.2 スタンプ選択画面を用いた方法

第二のスタンプ検索・描画方法は、スタンプの選択画面を作成し、音響特徴分析、音素列検索ともにスタンプが得られた場合には、それぞれ結果がスタンプ選択画面に表示され、利用者を選択してもらい、スタンプを判断する方法と

なっている。

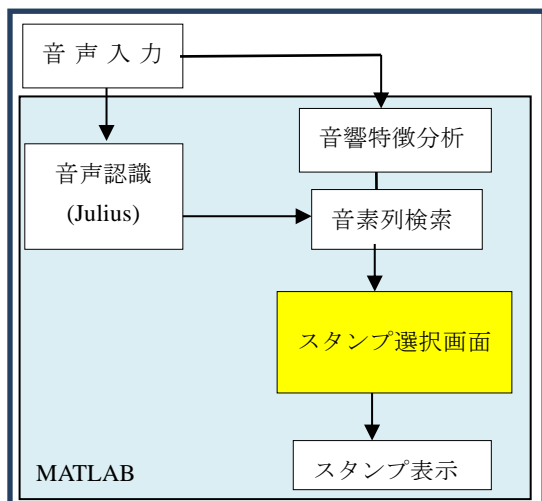


図 14 スタンプ選択画面を用いた方法
 Figure14 Stamp retrieval system2.

5. 実験

5.1 実験方法

被験者である男女 15 名に 2 つの検索方法を使用してスタンプを表示してもらい、それぞれの検索方法で、10 種類の擬態語・擬音語を 3 回ずつ音声入力してもらい、目的のスタンプ得られた割合によりスタンプの正答率を求めた。

5.2 実験結果

2 つの検索方法を用いた結果、スタンプの正答率が高い結果となったのは第 2 の検索方法であった。実験の結果は表 5、表 6 のとおりである。

表 5 音響特徴と音素列の類似度を比較する方法の結果

Table5 Result of Stamp retrieval system1

擬態語・擬音語	正答率
どんっ	58%
きらっ	73%
すばっ	56%
ざーざー	60%
ちくちく	64%
しとしと	49%
がーん	84%
しゃきーん	84%
どっかーん	78%
しーっ	78%

ここでの正答率は表示されたスタンプが正しければスタンプは正しく検索された結果とする。正答率は、10 種類中 9 種類が 50%以上という結果になった。全体的に正答率は上がったが、「どんっ」「すばっ」「しとしと」の正答率はあまり高くない結果になった。「どんっ」と「すばっ」と「きらっ」、「ちくちく」と「しとしと」は音響特徴分析での類

似度が重複しやすく、誤った判断がされやすい結果となった

表 6 スタンプ選択画面を用いた方法の結果

Table6 Result of Stamp retrieval system2

擬態語・擬音語	正答率
どんっ	94%
きらっ	94%
すばっ	94%
ざーざー	96%
ちくちく	85%
しとしと	98%
がーん	100%
しゃきーん	98%
どっかーん	96%
しーっ	98%

ここでの正答率は選択されたスタンプが正しければスタンプは正しく検索された結果とする。正答率は、10 種類全てが 50%以上という結果になった。また、「がーん」は全員がスタンプを表示できる結果となった。

6. まとめ

本報告では、音響特徴分析や音素列検索を用いて、擬態語・擬音語による音声入力を用いてスタンプを描画するインタフェースを作成した。個人によって表現や発声のバリエーションがある擬態語・擬音語がある中で、2 種の検索方法を使用した結果、音響特徴による判断と音素列による判断がともに信頼性が高い場合に両方表示して選択する方式を採用することにより、従来 64%であった全体の平均正答率を 95%に改善できた。これは個人の擬態語・擬音語による表現や発声のバリエーションに対応できているためと考えられる。

謝辞 本研究は科学研究助成金 24500131 の援助を受けた。

参考文献

- 1) LINE 公式ホームページ
<http://line.me/ja/>
- 2) 鈴木 教子、相川 清明，“音声入力を用いた感性特徴による図柄パターン検索”，2010-SLP-85, No.4, pp. 1-6, (2011-02)
- 3) 鈴木教子，相川清明，“音声入力による図柄パターン検索における未知語処理”，日本音響学会講演論文集，3-10-3, pp75-76, 2011-19.
- 4) 富田観月，相川清明，“擬態語・擬音語音声入力によるスタンプ描画インタフェース”，日本音響学会講演論文集，2-Q4-12, pp.144-145, 2014-3.
- 5) 雨宮香織，日置優介，相川清明，“音声入力による描画インタフェースの類似度の検討”，日本音響学会講演論文集，2-Q-13,
- 6) 鹿野清宏，伊藤克亘，河野達也，武田一哉，山本幹雄，“音声認識システム”，オーム社。