様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 ９ 日現在

機関番号：３２６７５
研究種目：基盤研究(C)
研究期間：2012～2014
課題番号：２４５００２１５

研究課題名（和文）出現頻度の偏った母集団の希少事象の認識のための通信路符号化モデルの構築

研究課題名（英文）Design of channel coding model for recognition of infrequent events in time series

研究代表者
伊藤　克亘（ITOU, Katunobu）

法政大学・情報科学部・教授

研究者番号：３０３５６４７２

交付決定額（研究期間全体）：（直接経費）　　4,100,000円

研究成果の概要（和文）：統計モデルでは希少な事象に対する性能は低くなりがちである。本研究では、意味のある希少な事象として、音メディアにおける表現に着目した。ここでは、表現を文脈から期待される標準的な特徴からの逸脱によって生起するものであると仮定する。これに基づき、楽器の演奏表現や声優の演技表現、音声の強調表現など様々な表現コーパスを整備した。この研究を通し、表現といわれるものの一部は、高い技術を持った生成者が意図的に行う逸脱である。したがって希少ではあるが、局所的(例えば一曲内など)には、ある程度繰り返して生成されることがわかり、階層的確率モデルで表現するのが適切であることがわかった。

研究成果の概要（英文）：Recognition performance of statistic models fall for infrequent events. In this study, as significant infrequent events, expressions in acoustic media were focused on. Expressions are assumed as deviation from expected features by listeners in their situations or contexts. Based on this assumption, various expression corpora, such as violin performance, voice acting, and emphasis for presentation speech, were constructed. In conclusion, some class of expression was assumed to be purposely generated by high-skilled performers. Expressive deviation was infrequent, however repeatedly generated locally, such as in a single song or a single lecture. Such expression was well-described by hierarchical probabilistic models such as HDP-HMM.

研究分野：音声・音響・音楽情報処理

キーワード：表現　逸脱　階層モデル　音声　音楽　演技音声

１．研究開始当初の背景

現在、主流である確率モデルを用いた音声認識では、音響信号x を観察したときに最大の尤度を持つ単語列w を様々な候補から探索する。これは、符号理論の観点からは通信路符号化における最尤復号化法とみなせる。この枠組では、音響モデルP(xjw) と事前確率P(w) の積が大きくなるものが選択されることになる。

しかし、音・言語メディアを対象とする認識問題においては、有効な事前確率を利用できている問題はまれであり、音響モデルだけで認識する(事前確率を均一と見なす) のが一般的である。

これらの問題に対し、希少な事象を大量に観測した結果がポアソン分布に従う例が多いことなどを参考にして、個別の事象を別々にモデル化するのではなく、共通の構造(例えばポアソン分布) を活用して希少事象の事前確率を導入する点が特色である。

希少な事象は、統計モデルに基づく手法では、軽視されがちである。なぜならば、評価用データでも出現頻度が小さいため、その部分の認識性能が悪くても全体の性能に与える影響が小さいからである。

しかし、情報検索のように、希少であることに価値を見い出すアプリケーションは多数ある。例えば、音声認識結果を検索する場合には、希少な固有名詞などが正しく認識できていることが望まれる。

本研究の独創的な点は、このような希少な事象の認識性能自体を向上しようという点である。

それを実現するために、これまでは、現象の事後的な説明のみに利用されていた方法を予測(事前分布) として活用する。例えば、ジップの法則にしたがう母集団のうち、希少事象については、ポアソン分布にしたがって偏って出現するものがあることが知られている。その場合に、低頻度の事象をまとめた総出現回数(例えば、出現順位がN 位よりも大きな事象全てをまとめた場合) が比較的大きければ、そのままでも事前分布として有効であると期待される。

そうでない場合については、アプリケーションごとに利用できる周辺情報で出現予測を補強する可能性について検討する。

２．研究の目的

音・言語メディアでは、出現頻度の多いものが全体に占める割合が多いような現象(ジップの法則)がよく見られる(単語の使用頻度や、登場人物の登場回数など)。これらの現象では、希少な事象(低頻度なもの) の出現回数は非常に小さい。したがって、放送コンテンツの認識など、実データを学習データとして利用しなければならない応用では、希少な事象に対する性能は低くなる。本研究では、このような希少な事象の認識性能を高性能化することを目的とする。

本研究では、一般的に価値がある希少な事象として、主として音メディアにおける「表現」に着目する。表現とは、文脈から期待される標準的な特徴からの逸脱によって生起すると考える。具体的な対象としては、音楽の楽器演奏における楽譜で決定的には指定されない演奏表現、声優の演技、講演における単語の強調表現などである。

３．研究の方法

(1)　放送コンテンツの収録とコーパス設計、構築

話者ダイアライゼーション(音声コンテンツをセグメントに分割してそれぞれのセグメントに話者ラベルを付与するタスク)の研究のため、受聴者が音声だけで話者情報を推定しているコンテンツであるラジオ番組を収録し、話者ラベルを付与したコーパスを作成した。50 時間のデータから、153 名の話者の 1779 セグメント(164 分)を整備した。これらを、モデルの学習データおよび性能評価用

のデータとする。

音声の表現としてプレゼンテーションにおける重要単語の強調表現を収集するために、解説、情報、報道番組 60 番組 から 4 番組を選び、118 発話を収集した。

また、声優の音声演技表現を分析するために、アニメーション番組で物語が同じで俳優が異なるものを声優セットごとに 50 エピソードほど収集した。それぞれの声優セットに対し、600 分、500 分程度整備した。

(2)　音楽コンテンツの収録とコーパス設計、構築

奏法によって音響的な特徴が大きく変化する擦弦楽器を代表してバイオリンの様々な奏法を含む独奏 5 フレーズをプロ奏者が演奏したデータを収録した。これに対し、手動で楽音の開始時刻を付与して、評価データとして用いる。

ポピュラー音楽のプロ歌手の音色を詳細に観察するためにカラオケ音源が存在するコンテンツを整備した。カラオケ音源を減算することにより、歌声だけを抽出し歌声コーパスとし、観察用や評価に用いる。

(3)　信号処理・統計的手法を用いたモデル化

表現すなわち、逸脱という現象を扱うには、大量の学習データを収集することが困難である。その問題点を克服するために、特徴量と確率モデルを工夫する。理論担当の分担者が分担をはずれることになったため、当初とは方法論を変更した。(詳細は成果の項を参照のこと)

４．研究成果
(1)　音楽情報処理

擦弦楽器の発音現象に関して、カオスを用いたモデル化の可能性が示唆された。擦弦楽器の音色に関しては、従来は、線形システムに微分方程式を組み合わせて事変とするようなモデル、もしくは、確率過程としてモデル化されていた。整備した演奏コーパスに基づいて、分析合成を行い、様々な主観評価実験を行った結果を仔細に検討した。その結果、音色に対する知覚的な影響が非常に大きい発音部は、カオスでモデル化するのが最適であるとの仮説を得た。この仮説をもとに予備実験を行った結果、より高品質な合成を得られた。

演奏表現の処理のために、逸脱と表現の関係を様々なレベルで明らかにした。特に、擦弦楽器の様々な演奏表現に対応できるような音符内状態推定の手法と音声波形の音量軌跡をアーティキュレーションとよばれる局所的な音量変動による音符内での表現と、ダイナミクスとよばれる音符間にわたる大域的な表現に分離する手法を確立した。これらの研究を通じて、演奏表現ごとにさまざまに変化し、従来はうまく扱えなかった希少現象を無限混合正規分布でモデル化し、扱えるようにした。これらの研究により、印象的な表現(「よい」表現)は、高い技術を持った演奏者が意図的に行う逸脱であり、希少ではあるが、ある程度繰り返して用いられる(生成される)ことがわかった。多彩な表現のプリミティブはやはり多彩であり、学習データによる学習は不向きである。しかし、逆に同一の曲内では、局所的に同一のプリミティブが出現する。その現象を sticky HDP-HMM で表現することで、学習データなしに音符内での状態推定を可能にした。

グルーブと呼ばれる、ポピュラー音楽のリズムにおける逸脱現象についてモデル化を試みた。しかし、グルーブでは、確かに、周期的なリズムからの逸脱はあるが、その逸脱

は周期的に生じるものであり、ゆらぎも小さいため、本課題で対象としている逸脱とは異なる種類の逸脱であるとの結論を得た。

(2) 音声情報処理

話者識別に関しては、ラジオ番組の話者を区別し話者ラベルを付与する、ダイアライゼーションに取り組んだ。一定の性能を示す手法は実現したが、希少現象が扱えないことよりも、話者識別においては、たいていのタスクで学習データが十分ではないことが問題であることが明らかになった。

音声における表現という現象を解明するために、アニメーションの同一キャラクターを演じる異なる声優間での話者変換の研究を行った。第三話者コーパスを利用した変換という新しい手法を提案することで、異なるテキストを学習データとした話者変換を実現した。この研究では、一般的なイメージとは異なり、声の高さや発声方法などを大きく変化させる「表現」自体に、声優の個性がでるわけではなく、主として性質に個性が表れており、性質を変換することで一定の話者変換性能が達成された。

講演などにおける強調現象については、日本語の講演では、うまく強調できているデータはそれほど多くはない。そこで、強調が上手い話者のデータを対象に強調表現を分析したところ、3 つのタイプの強調表現が観測された。一つは、強調したい単語のアクセントの部分の高さを通常よりも高くするタイプ。他の二つは、強調したい単語を強く発生する、その単語の前後にポーズをおくというものであった。これらは、組み合わせて用いるというよりは、使い分けている印象であった。しかし、収集したデータ量が不十分だったため、どのように使い分けているかは不明である。アクセントに関しては、学習データからアクセント成分を自動で分離し、アクセント成分の強さをモデル化することで、任意の発話音声の任意の単語を強調しているように聞こえるように修正することに成功した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 4 件）

① 小泉悠馬，伊藤克亘，"音量軌跡の遷移型状態空間表現に基づくダイナミックとアーティキュレーションへの分解"，電子情報通信学会論文誌，査読有，J98-D (3)，492-500，(2015)

② 小泉悠馬，伊藤克亘，"連続励起振動楽器を対象としたノート内セグメンテーション"，電子情報通信学会論文誌，査読有，J-97D (3)，584-592，(2014)

③ 小泉悠馬，伊藤克亘，"擦弦楽器の意図表現合成のための奏法モデル",情報処理学会論文誌，査読有，54 (4)，1319-1326，(2013)．

④ Yuma Koizumi, Katunobu Itou, "Performance expression synthesis for bowed-string instruments using Expression Mark Functions", POMA, 査読有, Vol. 15, pp. 035003, (2012).

〔学会発表〕（計 18 件）

① Yuma Koizumi, Katunobu Itou, "Intra-note segmentation via sticky HMM with DP emission", ICASSP, 査読有, 2144-2148, (2014), 2014.5.7, Florence(Italy).

② 木村拓也，伊藤克亘，"調波音・打楽器音分離手法を用いたギター・ベースギターの自動採譜",情報処理学会全国大会，査読無，5R-3，(2014)，2014.3.13，東京電機大(東京都足立区)．

③ 山上泰志，伊藤克亘，"演奏動画を用いた作曲支援のためのバイモーダル型 TAB 譜採譜システム",情報処理学会全国大会，査読無，5R-4，(2014)，2014.3.13，東京電機大(東京都足立区)．

④ 木立真希，伊藤克亘，"呼気量のモデル化に基づく歌唱修正システム",情報処理学会全国大会，査読無，5R-6，(2014)，2014.3.13，東京電機大(東京都足立区)．

⑤ 安田沙弥香，小泉悠馬，伊藤克亘，"ラジオ放送話者ダイアライゼーション",情報処理学会全国大会，査読無，5S-6，(2014)，2014.3.13，東京電機大(東京都足立区)．

⑥ 塩出萌子，小泉悠馬,伊藤克亘，"中間話者コーパスを用いたアニメーション演技音声のための話者変換",情報処理学会全国大会，査読無，6S-8，(2014)，2014.3.13，東京電機大(東京都足立区)．

⑦ 谷山拓未，伊藤克亘，"旋律の特徴抽出による作風類似メロディ生成",情報処理学会全国大会，査読無，1R-1，(2014)，2014.3.11，東京電機大(東京都足立区)．

⑧ 蠍川内努，伊藤克亘，"歌声を用いた DTM 向け演奏表現パラメータの入力",情報処理学会全国大会，査読無，1R-4，(2014)，2014.3.11，東京電機大(東京都足立区)．

⑨ 小泉悠馬，伊藤克亘，"ディリクレ過程を出力する Nest 型 HMM を用いた音符内状態推定",日本音響学会講演論文集，査読無，985-988，(2014)，2014.3.10，日本大(東京都千代田区)

⑩ 小泉悠馬，伊藤克亘，"連続励起振動楽器を対象とした音量軌跡のダイナミクスとアーティキュレーションへの分解法"，情報処理学会研究報告(MUS)，査読無，1-6，(2014)，2014.2.24，筑波大(東京都文京区)

⑪ 小泉悠馬，伊藤克亘，"奏者の意図したテンポ変動の推定に基づく演奏録音の自動伸縮修正法"，FIT 論文集，査読有，12(2)，19-24，(2013)，2013.9.6，鳥取大(鳥取県鳥取市)．

⑫ Yuma Koizumi，Katunobu Itou，"Expressive oriented time-scale adjustment for mis-played musical signals based on tempo curve estimation"，DAFx，査読有，1-7，(2013)，2013.9.3，Maynooth(Ireland)．

⑬ 上野涼平，小泉悠馬，伊藤克亘，"音楽知識を利用したハーモナイザー",情報処理学会全国大会，査読無，5R-5，(2013)，2013.3.8 東北大(宮城県仙台市)．

⑭ 荒木大誉，伊藤克亘，"歌詞とモチーフを入力とした作曲モデルによる作曲支援システム",情報処理学会全国大会，査読無，5R-6，(2013)，2013.3.8 東北大(宮城県仙台市)．

⑮ 森田花野，小泉悠馬，伊藤克亘，"教則本を利用したギターフレーズの難易度推定",情報処理学会全国大会，査読無，4R-2，(2013)，2013.3.7 東北大(宮城県仙台市)．

⑯ Kentaro Hirayama，Katunobu Itou，"Discriminant analysis of the utterance state while singing"，ISSPIT-2012，査読有，49-54，(2012)，2012.12.14，Ho Chi Minh(Vietnam)．

⑰ 平山健太郎，伊藤克亘，"筋電センサーを用いた歌声分析のための喉頭音源分析",FIT 論文集，査読無，11(2)，179-180，(2012)，2012.9.4，法政大(東京都小金井市)。

〔その他〕
ホームページ等
http://cis.k.hosei.ac.jp/info/faculty/digital/itou.html

６．研究組織
(1)研究代表者
伊藤　克亘（ITOU Katunobu）
法政大学・情報科学部・教授
研究者番号：30356472

(2)研究分担者
西島 利尚（NISHIJIMA Toshihisa）
法政大学・情報科学部・教授
研究者番号： 70211456

廣津 登志夫（HIROTSU Toshio）
法政大学・情報科学部・教授
研究者番号： 10378268