

## 音響的特徴の分析に基づく

## 美しい日本語発話の習得支援

## Acquisition support of fine Japanese utterances based on analysis of acoustic feature

辻 亮<sup>\*1</sup>  
Ryo Tsuji桐山 伸也<sup>\*2</sup>  
Shinya Kiriyama北澤 茂良<sup>\*2</sup>  
Shigeyoshi Kitazawa<sup>\*1</sup> 静岡大学大学院情報学研究科  
Shizuoka University, Graduate School of informatics<sup>\*2</sup> 静岡大学情報学部  
Shizuoka University, Faculty of Informatics

We aim at the construction of the support system to which the learner can acquire the nasal consonant and the voiceless of vowel which are the indispensable elements to fine Japanese utterance. The technique for judging the quality of the nasal consonant and the voiceless of vowel automatically has been developed by using recognition scores of the HMM acoustic model.

## 1. はじめに

近年、日本語の正しい話し方が出来ない人が増え、その重要性についても軽視される傾向にある。現に、正しい話し方が求められるアナウンサーでさえも、美しいとされる発話ができないのが現状である。そこで本研究室ではなおざりにされつつある美しい日本語の保護を目標とし、2003年度よりボイストレーニングの講師を招き、学生が実際にボイストレーニングを受け、その収録風景などの収録をおこなってきた。そしてこれらの資料を生かし、美しい発話の習得のための支援として、ボイストレーニングのコンテンツを作成してきた。[太田 04]

本稿ではボイストレーニングのコンテンツの一部として、PC上でボイストレーニングが受けられるような支援システムの検討をおこなう。具体的には、音響モデルを使った鼻濁音・無声化識別を提案し、先行研究との比較実験をおこなう。そしてその結果をふまえて、先行研究で使われた波形情報と連携させた手法を提案する。

## 2. 先行研究との違い

同研究室で行った先行研究[中島 04]ではボイストレーニング項目の母音の無声化・鼻濁音に対して音響的な分析を行って、ユーザーの習熟度を自動判定するものであった。

先行研究では音響的に分析がおこないやすい鼻濁音・母音の無声化部分に焦点を絞っていたのだが、波形分析を行った鼻濁音に関しては、話者性などに対して影響を受けやすいという欠点がある。さらに無声化に関しても既存の音響モデルを使っていた。

本稿では以前のシステムの識別精度を上げるために、鼻濁音・母音の無声化に対応できるような音響モデル作成を考える。そして発声した音声を分析・評価してくれるようなシステムの構築を目指す。これにより先行研究における識別率の向上や母音の無声化・鼻濁音などの音響特徴を考慮した音声データベースへの利用など音声認識・合成分野でも幅広く応用できると思われる。

連絡先：辻亮，静岡大学大学院情報学研究科，  
cs1057@s.cs.inf.shizuoka.ac.jp

表1. 先行研究との比較

	無声化	鼻濁音
従来手法	・既存の音響モデル ・shとshnなど対応するものとの尤度比較	・波形情報の利用 ・子音と母音のパワー、パワー比、振幅比など
提案手法	・非無声化、無声化の音節単位でのモデル作成 ・無声化の「sh」モデルと有声の「sh」モデルなど ・対応するものとの尤度比較	・音響モデルを使用 ・鼻濁音の「ga」モデルと非鼻濁音の「ga」モデル ・対応するものとの尤度比較

## 3. 鼻濁音と無声化の音響的特徴

## 3.1 鼻濁音

本稿で分析する鼻濁音というのは、語中の『ga,gi,gu,ge,go』に対応する子音部分が、軟口蓋音に変化する現象で、また語頭の場合では鼻濁音にはならないという特徴をもったものである。本来『g』の音素は破裂音であるが、鼻音化すると『m,n』の子音の特徴に類似し、滑らかに母音に移行する。

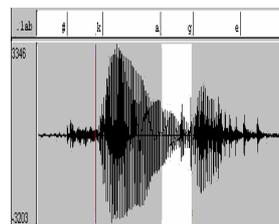


図1、濁音の波形「影」

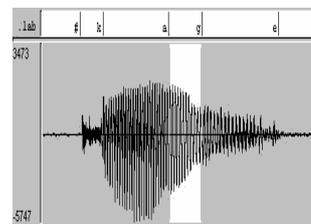


図2、鼻濁音の波形「影」

## 3.2 無声化

無声化とは母音『i,u』が『k,s,t,h』などの子音に挟まれたときに、その母音部分が息を出すだけにとどまるものである。

その結果、子音の特徴しか現れなかったり、母音部分が弱くなったりする現象を観測することができる。

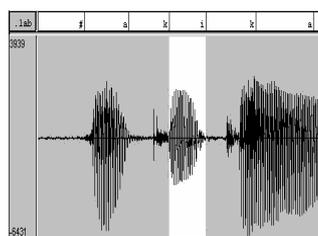


図3、有聲の波形「空き缶」

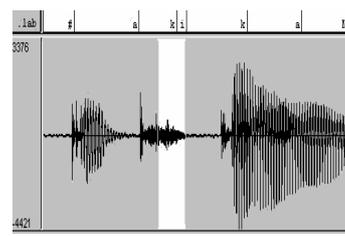


図4、無声化の波形「空き缶」

## 4. 音響モデルによる習熟度自動判定

### 4.1 自動判定の手法

昨年に同研究室で行ったボイストレーニングの際に収録した音声データ(4話者)を音響モデル評価に用いた。この音声データはボイストレーニングの講師の方が無声化・鼻濁音化できているかの判定をしてくださったものである。その中から一人当たり30単語程度、約120単語の評価用データを使って、作成した音響モデルの識別性能を分析する実験を行った。評価内容は、評価用データから該当する音節部分だけを切り出して識別させる音節単位での識別と、単語単位の評価用データを、そのまま識別させる単語単位での識別の2種類である。

識別方法としては、無声化判別をする場合なら HTK (Hidden Markov Model Toolkit) [HTK 05] に付属している認識コマンドを利用して、2種類の無声化・非無声化モデルそれぞれに出力される尤度の大小を元に判定を行うようにした。

### 4.2 音響モデルの構築

本稿では ATR のデジタル音声データベースのセット B[ATR 05] の男性話者5名、女性話者4名から、判別したい鼻濁音(5種類)・無声化(9種類)・非鼻濁音(5種類)・非無声化(9種類)の音節部分を文法規則に基づき切り出した。そしてHTKを用いて音響モデルの作成・最適化を行った。以下に音響モデルの分析条件を示す。

表2. 音響モデルの分析条件

項目	パラメータ
標本周波数	16kHz
フレームシフト	10msec
特徴ベクトル	mfccの26次
HMM状態数	5
フレーム幅	25msec

### 4.3 自動判定性能の評価

音節単位の識別において鼻濁音に対する認識率が約81%、無声化に対しては約85%になった。

以上の識別実験では評価データを手動で正確に切り出したものであったが、システムとして自動識別を行うには無声化・鼻濁音部分を含む単語単位の音声データを用いた識別が非常に重要になってくる。

そこで単語単位での識別結果を算出し、無声化の場合、最大で約63%、鼻濁音の場合は最大で約68%という結果を示した。

また先行研究の手法を再現して無声化の識別をおこなった。その結果、識別率が約50%であり、今回の手法の有効性が示された。また鼻濁音に関しては約70%であり、今回の手法では識別率向上には至らなかった。

## 5. システム試用実験

### 5.1 システムの設計

音響モデルを用いて、対象の音韻区間における尤度とフレーム数の平均尤度の2種類を取り出した。そして両者が鼻濁音と示す結果となれば、鼻濁音出来ていると判定し、後者のみが鼻濁音と示していれば、鼻濁音一步手前と判定、前者のみが鼻濁音と示していれば鼻濁音の傾向があると判定、両者が鼻濁音だ

と示していなければただの濁音と、4段階の判定をユーザーにフィードバックするようにした。

### 5.2 先行研究との比較実験

今回は音響モデルによって、どのくらい鼻濁音・無声化を識別できるかを実証する目的で、システムを tcl/tk で実装した。

評価実験については本研究室の学生5名の協力を得て、鼻濁音・無声化の識別を実際にシステムを使っておこなってもらった。

実験内容としては先行研究で使用したシステムと本研究でのシステムの使いやすさの違いと、識別手法の違いについてそれぞれ5段階でのユーザー5名による主観評価である。

図5・6より、識別手法の違いにおいて安定感・信頼度の2項目について評価してもらったが、鼻濁音においてはそれぞれ2.4、3.2と先行研究に比べて1ポイント、0.2ポイント低くなる結果となった。また無声化においてはそれぞれ3.6、3.6と先行研究に比べて、0.8ポイント、1ポイント高い結果が得られた。

使いやすさの項目においては4.4ポイントと高い評価を得ることが出来た。しかし、機能面においては自身の発声した音声波形情報の表示など、まだまだユーザーから追加して欲しい機能などの要望が多く見られた。

これらを踏まえてユーザー自身が必要としている情報が取得できるような支援システムを構築していく。

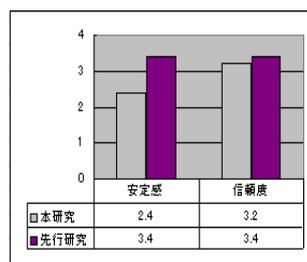


図5. 鼻濁音における評価

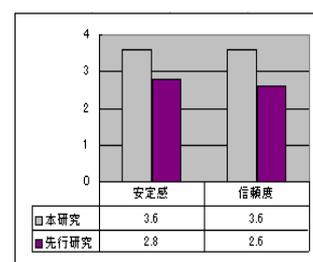


図6. 無声化における評価

### 5.3 考察

実際に自動判定システムを利用してもらうことで、ユーザーが欲している情報・機能がどういふものかという有益な情報を得ることが出来た。

判定項目の鼻濁音に関しては識別率に差がほとんど無いものの、ユーザー評価には大きく差が出る結果になった。これは判定する際の識別結果の偏りが影響しているためと考えられる。つまり、明らかに鼻濁音ではないのに鼻濁音と判定されるとき、ユーザー心理にシステムに対する不信感が増大するためだと言える。このため、ユーザーが納得してシステムを利用できるような認識手法の調整・検討が急務といえる。

識別手法に関しては、音響モデルの尤度だけではなく、そのほかの情報を使うことでのハイブリッドな方式の必要性がみえてきた。そのために音響モデルの調整はもちろんのこと、波形情報と連携させる手法の確立が識別において重要になってくると思われる。

## 6. 認識手法の改善

先行研究では使っていた既存の音響モデルを使い、鼻濁音識別の核ともいえる、以下の5種類のパラメータ値にそれぞれ閾値を設けた。そしてそれを満たしたパラメータ値の個数の大小で鼻濁音の判定をおこなっていた。

- lrcv 子音と母音の平均振幅比
- prcv 子音と母音のパワー比
- prtc 子音パワーが全パワーに占める割合
- prncv 単位長辺りの子音と母音のパワー比
- prntc 単位長あたりの子音と母音の全パワー比

そこで、従来手法で使っていた既存のモデルに対して本稿で作成した鼻濁音・濁音の音響モデルに置き換えることを検討した。

従来手法に適用するための音響モデルの選択は、提案手法を用いて尤度スコアを参照し、より正確なセグメンテーションをおこなうことができた音響モデルを使用することにした。

また識別率の判定基準については5つのパラメータ値において、2個以上が閾値を満たしたときには鼻濁音とみなし、2個未満の場合は濁音というように設定した。

その結果、最適な音響モデルを使用したほうが識別率において約73%と既存のモデルを利用した先行研究と比べて識別精度の向上がみられた。そのほかにも各パラメータ平均値の変化については適応した音響モデルを使用したほうが、鼻濁音を示すパラメータ値が、鼻濁音データでは全般的に上昇し、逆に濁音データでは値の上昇を抑え、誤った判断をする可能性を減らしていたことがわかった。

これらのことから、正確なセグメンテーションと鼻濁音・濁音の波形識別は高い精度を目指す上で密接な関係があるといえる。

## 7. まとめ

鼻濁音の識別は音響モデルを最適化し波形情報を付加することで、従来手法以上に各パラメータの平均値が最適化され精度の向上と安定性が得られることが分かった。

また無声化についても音響モデルの学習をおこなうことで、識別率の向上がみられることが確かめられた。

今後は閾値を最適化するための検討や、音声境界についての検討、さらに重みの概念を導入することによりパラメータの最適化をおこない、識別率の向上を見込める認識手法の構築を目指す。

また無声化に対しても音響モデルによる識別だけでなくピッチの変化などの波形情報を付与することを検討する。

またシステム面についても、ノイズキャンセル機能を付いたヘッドセットの利用の検討し、雑音下における性能変化を考慮していかなければならない。

## 謝 辞

本稿作成に際して多大な協力をいただきましたボイストレーニング講師・大月美保子先生、実験に協力していただいた研究室の皆様には深く感謝いたします。本研究は科研費・若手研究(B)16700172の支援による。

## 参考文献

[中島 2004] 中島梓,北澤茂良「日本語発音に関する習熟度自動判定」静岡大学情報学部情報科学科卒業論文,(2004)

[HTK 05] HTK Speech Recognition Toolkit

<http://htk.eng.cam.ac.uk/>

[太田 2004] 太田有美,北澤茂良「マルチモーダルボイストレーニング支援システム」静岡大学情報学部情報科学科卒業論文,(2004)

[ATR 05] ATR 技術リエゾンセンタ

<http://www.red.atr.co.jp/index.html>