合成音声利用による中国語声調の学習システム

早稲田大学　比企静雄，今泉一哉，砂岡和子，楊　立明

hiki@waseda.jp, imaizumi@ruri.waseda.jp, ksunaoka@waseda.jp, youl@waseda.jp

要旨　標準中国語の４種の声調に対応する声の高さの変化を弁別できる能力を習得するための，合成音声利用による学習システムを試作した．コンピュータ・プログラムで音声を分析再合成する手法を利用して，弁別するべき元の音声サンプルの対の音響分析的な特徴を段階的に内挿・外挿して，声の高さの変化の対比の程度を任意に調節した音声サンプルの刺激を用意する．そして，CAI によって，応答の履歴に応じて次の最適な刺激を選んで提示して，習得の目標に到達するまでの提示回数が最少になるようにしている．

１．言語情報を伝える声の高さの変化

聴覚的な声の高さ（ピッチ，物理的には音声の基本周波数）の変化によって伝わる音声の情報のうちには，多くの言語に共通な，強調や抑揚のための，句や節や文の区分にまたがる変化に加えて，日本語の単語アクセントや中国語の音節声調のような，音節の間の渡りや音節の区分の中での変化がある[1].

日本語の肯定や疑問の文の抑揚の発話での声の高さの変化は，基本周波数の変化を全音階的音階でとると，その速さは，1音節の平均的な長さの0.2秒あたりで1全音くらいにゆっくりである．これに対して，日本語の単語アクセントで，アクセントがある音節から次の音節への渡りでの基本周波数の下降は，0.2秒あたりで2全音以上にはやくなる．

標準中国語での4種類の声調（第1声：高くて平ら，第2声：上昇，第3声：低い平ら，第4声：下降）の基本周波数の変化の速さは，上昇は0.2秒あたり1全音ぐらい，下降は0.2秒あたり3全音にもなる．

２．声の高さの変化の弁別に必要な分解能

実際の発話で，これらの基本周波数の変化の特徴を認識できる変動の許容範囲があり，それらを相互に弁別するのに必要な分解能が必要になる．中国語の声調を伝える声の高さの変化パタンの許容範囲は，音響分析と聴取実験を組み合わせ求めると，それぞれ相互の中間の近くにまで拡がっている[2]（図1）．

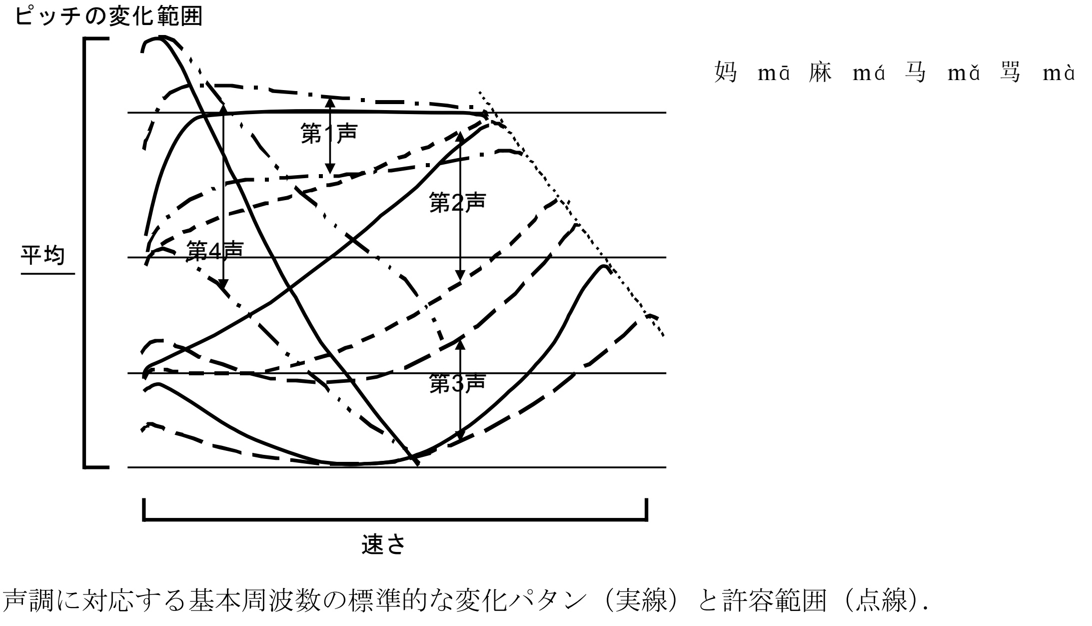

図1．4種の声調に対応する基本周波数の標準的な変化パタン（実線）と許容範囲（点線）．

声の高いと低いの対比は，第1声と第2声と第3声の相互が，1全音以下の狭い許容範囲の差しかないので，それに対する分解能が必要になる．（ただし，第1声や第3声は，それぞれ特異な音色を伴っているので，それも弁別の手がかりになっている．）

声の高さの上昇のうちの対比では，第1声と第2声との相互の許容範囲の差が狭くて，0.2秒あたり1/2全音以下の上昇を区別する必要がある．

（なお，日本語の単語アクセントのための声の高低の対比の相対的な差は，実際の発話では1全音ぐらいに狭まる場合もあるので．この許容範囲に対して，やはり1全音以下の分解能が必要になる．）

３．分析再合成による刺激の音声サンプル

このような標準中国語の４種の声調に対応するような声の高さの変化を弁別できる能力を習得するためには，従来は，音節や単語や文を編集して，それを実際に教師が発音した音声サンプルの教材を用意し，生徒はこれらを繰り返し聴き比べながら，それぞれの声調の標準的な声の高さの変化パタンの特徴を把握していくような方法がとられていた．

それに対してこの学習システムでは，コンピュータ・プログラムによって音声を分析して再合成する手法を利用して，弁別するべき元の音声サンプルの対の間で，音響分析的な特徴を段階的に内挿または外挿して，声の高さの変化の対比の程度を任意に調節した音声サンプルの刺激を用意する（図 2）.

音声の分析再合成の計算機プログラムは，河原秀樹氏（和歌山大学工学部）らによって，ATR人間情報通信研究所で開発されたSTRAIGHTに，モーフィングの手法が組み合わさったものである[3]．音源の基本周波数や強度や持続時間などの音響的パラメータを制御することができるので，この学習システムで必要とするような声の高さの制御パラメータによって，録音した音声を修正するための高度な機能を備えている．

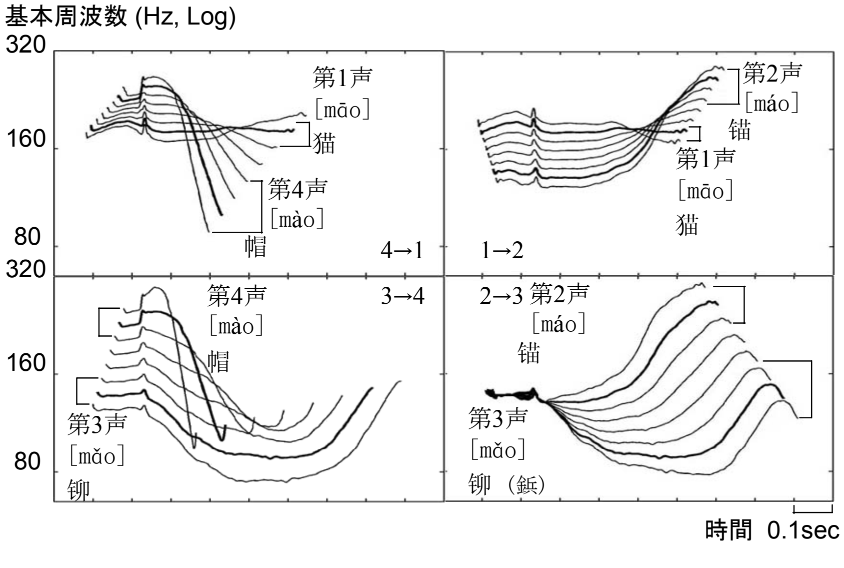

図2．第1声と第2声，第2声と第3声，第3声と第4声，第4声と第1声の間を，声の高さの変化パタンを内挿・外挿した例．元の音声サンプルは太い線，内挿・外挿した音声サンプルは細い線．コ印はそれぞれの声調としての許容範囲.

４．声の高さの変化パタンの制御パラメータ

刺激となる合成音声の声の高さの時間的な変化パタンを，聴覚的に必要な分解能で制御するパラメータは，次の4種で表現する．

1) 平均の声の高さ：女性と男性のそれぞれについて，普通の平均に対して，高いと低いの2レベルを設け，必要に応じて細分化する．

2) 声の高さの変化する範囲：普通の範囲に対して，広いと狭いの2レベルを設け，必要に応じて連続的に変える．

3) 声の高さの変化パタンの相互の対比：普通の変化パタンの相互の対比から，各パタンの許容範囲を保ちながら，それらの差を順次に縮小して近づける．

4) 話す速さ：普通の話す速さに対して，速いレベルを設け，必要に応じて連続的に変える．

５．刺激の提示と弁別の応答の様式

これらの刺激の音声サンプルは，被験者に聴覚で提示するだけでなく，とくに，初期の段階で弁別が非常に難しい場合には，聴覚的な刺激の差を把握しやすくして練習するために，その波形や声の高さの変化パタンを，聴覚提示と同時に視覚表示するように選ぶことができる（図3）．

刺激に対する弁別は，同定か対比較かは課題によるが，それらの応答は，口頭や筆記で，またはタッチスクリンに表示された漢字と音声記号や，模式的な声の高さの変化パタンを指差して報告することもできる．

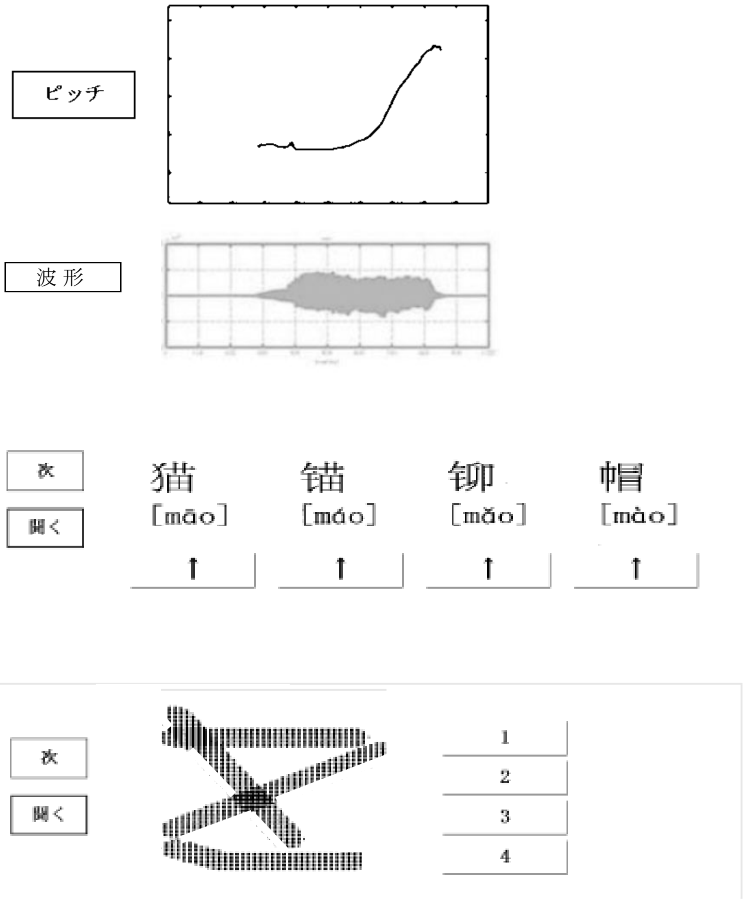

図3．タッチスクリンの表示の例．上から，ボタンで選ぶ声の高さの変化パタンと音声の波形，漢字と音声記号での応答ボタン，4種の声調の模式的な声の高さの変化パタンでの応答ボタン．

６．CAIのプログラムの構成

あらかじめ，単語表の編集と，録音した音声サンプルから刺激の音声サンプルの分析再合成をして，プログラムの中に保存しておく．CAIのプログラムでは，単語表の選択，音声サンプルの提示，弁別の応答の記

録，修正した音声サンプルの選択の過程を繰り返す．

誤答が続いたあとでは，次の刺激の弁別が易しくなるように，正答が続いたあとでは難しくなるように，音声サンプルの声の高さの変化パタンを修正したものを，次の刺激として自動的に選択して提示する．性別や声の高さの平均や変化範囲や，変化パタンの対比の程度や，話す速さなどの，多種類の制御パラメータの組み合わせがあるが，その提示の順序は，保存してある被験者の応答の履歴に応じて選択し，次の最適な刺激を効率良く提示することによって，習得の目標に到達するまでの提示回数が最少になるようにしている．

７．声調の声の高さの変化パタンの対比の枠組み

標準中国語の声調のための声の高さの変化は，高くて平らと低くて平らとの対立や，それらと上昇と下降の対立などが，複雑に組み合わさっている．通常の発話では，音節が連続し，さらに発話の速さや強さが変わるので，声の高さの変化はさらに複雑に変形する．そのような中から 4 種の記号を取り出すためには，それぞれの声調の聴覚的な不変量を把握しておかなければならない．

この学習システムでは，音声の分析再合成の手法によって，声の高さの変化の，高い／低いや上がり／下がりなどの基本的な対比から習得を始められるように，刺激の音声サンプルを任意に用意できる．したがって，声の高さの変化の対比の枠組みを，音響分析的な特徴に対する聴覚心理的な応答の特性から原理的に組み立てて，音節の連続への拡張に備えることができるという利点がある．

８．単語表の編集のための資料

音声サンプルの単語の表を編集するためには，単語を構成する音節の子音と母音の種類や，声調のつながりや連続した発声での変形など，多様な要因を配慮しなければならない．

そこでこのような単語表の編集のための資料として，中国語の機械翻訳のために編纂されたデータベース [4] から，単語の品詞別，単語を構成する音節数別，単語中での音節位置別に，音節の種類順，声調の種類順に配列しなおしたデータベースを作成してある．また，単語の日常生活での使用頻度だけでなく，品詞別に単語中での音節位置による声調の出現頻度の偏りや，遷移確率の偏りの統計資料も調べてある[5]．これらを参照して，聴覚的な弁別の困難な程度を推定して，単語表を系統的に編集することができる．

補足　当初はこの学習システムは，人工聴覚を装用した中国の聴覚障害児を対象にしていたが[6]，今回は，外国語としての中国語の音声を学習する日本人学生に，活用の範囲を拡張する可能性を検討した．

文献

[1] 比企静雄，人工聴覚による音声情報の伝達の可能性：諸言語への普及に伴って生じる問題，ヒューマンサイエンス（早稲田大学人間総合研究センター紀要），12 巻，2 号，pp. 16－22，2000.

[2] C-K. Chuang and S. Hiki, Acoustical features and perceptual cues of the four tones of standard colloquial Chinese, J. Acoustical. Society of America, Vol. 52, No. 1(Part 1), p. 146, 1972.

[3] 松井久美，川原英紀，STRAIGHT による感情モーフィング音声の知覚と変換関数の構造について，電子情報通信学会　信学技報 EA2003-18, SP2003-18 (2003-04),　pp. 63-68, 2003. 4.

[4] The Grammatical Knowledge-base of Contemporary Chinese, edited by S-W. Yu, Tsinghua University Press, China, 1988 (in Chinese).

[5] S. Hiki, L-m. Yang, K. Sunaoka and Y. Tokuhiro, Statistical properties of the Chinese four tones: Occurrence frequency and transition probability, Proceedings of the 18th International Congress on Acoustics, ICA2004, April 2004, Kyoto, Japan, pp. IV-3317-3320.

[6] K. Imaizumi, S. Hiki, L-m. Yang, M. Shiroma and Y. Fukuda, CAI system for testing and training discrimination ability of change in voice pitch utilizing analysis-synthesis speech samples, Proceedings of the 18th International Congress on Acoustics, ICA2004, April 2004, Kyoto, Japan, pp. I-849-852.