「プレゼンテーションの強調習得」支援システム
Computer aided training for acquisition of focus on presentation

小島淳嗣
Atsushi Kojima
法政大学情報科学部ディジタルメディア学科
E-mail:atsushi.kojima.2v@stu.hosei.ac.jp

Abstract

This paper presents a presentation practice system, which makes a presenter vocalize with an emphasis on important words. According to a knowledge of phonetics, a height of an intonation mountain on an emphasized word becomes higher. Based on this idea, an emphasis is evaluated quantitatively by Fundamental Frequency (F0) in a speech. Specifically, firstly F0 is interpolated to become a continuous value. And estimates the parameters model for the generation process of F0 of a speech Secondly, a component caused an accent within the estimated parameters on a word is regarded a maximum value as a feature for judge for an emphasis or not. A standard on a model for judging is learned from training data. We evaluated word that subjects chose in advance was whether to listen emphatically using a three grade evaluation. As a result, a mean value of an utterance after a training was higher than after by 1 point. From this result, Feedback that judge for emphasis is efficient for acquire emphasis in a presentation.

1 はじめに

プレゼンテーション（プレゼン）は，大学，企業，試験，学会などで幅広く実施されている．それゆえ，良いプレゼンを行うための指南書 [1][2] も多く，能力向上への関心も高い．これらの教材を参考にプレゼンを見直す際に，スライドの作成法や内容の構成などは経過が残るため修正しやすい一方，話し方はどこがどう悪いのかがわかりにくいため，改善しにくい問題がある．そこで，理解しやすい話し方の特徴を調査するため，テレビに頻繁に出演する著名な解説者の 30 分程度の長さのプレゼンを聴取した．その結果，2 分に 1 回程度重要な点が強調されていることが分かった．これによって，解説者は聴衆（この場合，視聴者）に重要な点を伝達していると考えた．強調は重要な点を聴衆にとって明瞭にする．それゆえ，プレゼンの目的である聴衆の理解達成に重要である．本研究では，これを取り入れることを考えて，プレゼンの強調習得を支援するシステムを提案する．

これまで発話の強調された区間を抽出し，要約を行うために．強調されているかどうかを，文ごと [3] や時間構造で区切って [4] 定量的に評価する試みがなされてきたものの，単語の強調については十分に検討されていなかった．そのため，従来手法では，どの単語が強調されているかを特定できない．本稿では，発話中の単語が強調できているかを定量的に判定する方法を提案する．さらにこれを利用した練習システムを構築し，発表者のプレゼン時を想定した発話を分析し，単語ごとに強調か強調に聞こえない（非強調）かをフィードバックすることで強調習得を狙う．

2 プレゼンテーションにおける強調

2.1 強調の判別

本稿では，強調を聞き手に重要性を示すために発話の特定単語を際立たせることと定義する．発表者が強調できているかを判定するに当たり，どのような特徴に着目すれば良いか知るため，日本語を対象とした強調の仕方について音声学の文献を調査した．その結果，高さに関してはアクセントや音調との関連が明らかになった．文献によると強調時には，音調の山が高くなる [5]，プロミネンスのある語のアクセントの型が音調として明瞭に実現される [5]，音調の上昇幅が大きくなる変化はプロミネンスと呼ばれてきた語の強調法の一部に相当する [6]．さらに，ポーズに関しては，強調したいところのまえで，区切って，少し休む [7]，フォーカスのある語の直前，あるいは直後，あるいはその両方にポーズを置き，よりいっそうことばを目立たせてる [5]．また，大きさについてはプロミネンスが置かれた部分は，強く発声される [8]．

これらの知見より，強調判定のためにアクセントの明瞭さとポーズと強さに着目する．アクセントは，声の高さによって表される．よって，これが反映される物理量として F_0（基本周波数）を分析する．さらに，ポーズは音声がない背景雑音のみ含まれる時間長を分析する．強さは大きさの物理量であるパワーを分析する．

F_0 から得られた特徴量に加え，ポーズ長，パワーを特徴量とした回帰分析を用いて単語が強調されているか判定する．判定する基準は，学習データを用いて学習する．そのために，アクセントの明瞭さに相当する特徴量が明らかになっていないため，音声学の知見からアクセントの明瞭さが表れる音調の山の高さを評価する．この時，F_0 を分析する際の問題点を説明して，これを改善する．さらに，強調判定に有効な特徴量が明らかになっていないため，特徴量を選択する強調判定実験を行う．

2.2 強調発話の分析

学習データは，強調があり，プレゼンの形態に近い演技や対話でない発話が望ましい．単独話者の発話解析などで多く用いられる既存の講演音声コーパスや，放送大学のテレビ講義も検討したが，聴取した限り強調していることがなかったため対象

Supervisor: Prof. Katunobu Itou

から除外した．よって，コンスタントに収集できることもあり，放送大学を除くテレビ番組からデータを収集した．番組のジャンル問わず 60 番組程度聴取した結果，解説，情報，報道番組（経済，政治，時事，教養関係），通販で，10 番組に出演する話者 5 人が強調していることが分かった．ここから，特徴量を計算する際に邪魔になる BGM や効果音，共演者の声が発話に被ることが多かったり，番組の殆どがビデオで構成されていて話している区間が短く，収集効率の悪い番組を除外した．その結果，4 番組の 4 人の話者の発話を学習データとすることにした．収集した後，整備する際には，番組を聴取して強調している単語を含む 1 発話ごとに切り出した．結果，118 発話収集した．1 発話の長さは 4 s〜15 s 程度だった．

音調の山の高さを評価する際に，1 単語単体で発声した時とは異なり，話し声の F_0 は，息の減少により呼気段落の初めから終わりに向かってゆるやかに減少するという特徴がある．そのため，話し始めてから次の息継ぎまで発話した文の前半と後半に位置する単語を必ずしも平等に評価できない問題がある．そこで，話し声 F_0 生成過程モデルの藤崎モデル [9] を利用する．これは，話し声 F_0 が単語の局所的な変化に相当するアクセント成分と，息の減少による全体に渡る変化に相当するフレーズ成分の畳み込みからなるとするモデルである．観測 F_0 から，それぞれの成分を逆問題として推定し，前者の成分に着目することで文の位置に依存せずに，音調の山の高さを評価し，アクセントの明瞭さを表す特徴量を得る．

2.3　強調パラメータ計算

藤崎モデルのパラメータを，F_0 から推定する．推定したいアクセント成分は，矩形波で表される運動指令に対するステップ応答になる．よって，アクセント成分を推定するには，まずアクセント指令を推定することになる．アクセント指令のパラメータは，指令開始時刻，指令終了時刻，指令の振幅の 3 つとなる．これらのパラメータと観測 F_0，アクセント成分との関係を図 1 に示す．ここでは，「ひまわり」という単語を発声した音声の解析結果を示す．F_0 から分析されたアクセント成分は，指令の開始時刻から上昇し始めて，F_0 が最大値をとる指令終了時刻にアクセント成分も最大値になり，減少していることが分かる．

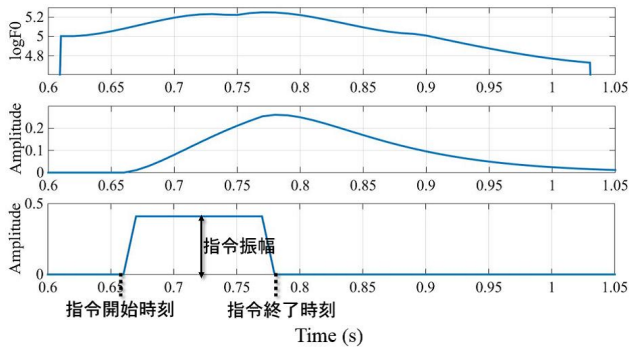

図 1. 観測 $\log F_0$ から推定されるアクセント指令のパラメータと生成されるアクセント成分　上段：$\log F_0$　中段：アクセント成分　下段：アクセント指令

日本語のアクセントは高低アクセントである．単語の各音節は高低のどちらかで発音され，下降する音節が先頭から数えて何拍目かで単語のアクセントが決まる．下降する音節は 1 箇所で，アクセント核と呼ばれ，これが見られる単語は起伏型，見られない単語を平板型と呼ぶ．図 2 に，起伏型の単語の例「ひまわり」のアクセントを示す．アクセント核は「ま」である．図 3 に，平板型単語の例「三日月」のアクセントの例を示す．音節の上の直線が高低のアクセントを示しており，曲線は音調を示している．よって，日本語の単語の音調を曲線で表すと山

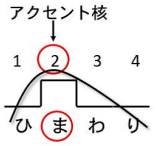

図 2. 「ひまわり」（起伏，2 型）のアクセント

図 3. 「三日月」（平板型）のアクセント

のような形になり，藤崎モデルのアクセント成分はこれをモデル化している．

藤崎モデルのパラメータ推定には，モデル提案者らが手動でパラメータを推定した経験から提案した手法 [10] を参考にする．手法は，前処理として無音区間や無声区間などが F_0 を持たないために，有声区間以外のみ値を持つ離散値となっている F_0 を連続値へ補間する．そして，アクセント指令のパラメータは，時間方向に一次微分した連続値の F_0 の正と負の極値の位置がそれぞれ指令開始時刻と指令終了時刻，それぞれの極値の平均を指令の振幅としている．ただし，推定の際に本手法と異なる点として有声区間は，外れ値の影響を考慮して幅 50ms のメディアンフィルタで平滑化する．無声区間は，有声区間以外のピークを発生させないため，線形補完を行う．さらに，話者に依存する基底周波数は，はずれ値の影響と自動化を考慮した手法 [11] を用いて，有声区間の F_0 の平均と標準偏差を計算し，平均から 3 倍した標準偏差を引くことで計算する．最終的に推定された，アクセント指令から生成されたアクセント成分の最大値を強調判定の特徴量とする．強さの推定には，パワーを利用し，最大値と平均値を特徴量とする．ポーズ長推定には，パワーとゼロクロスを利用し，音声認識等で音声区間の推定などにも用いられる手法 [12] を用いる．手法は，背景雑音のみが含まれる区間から，これらの平均と分散を用いて閾値を計算して音声区間を探索する．これを用いて計算された，強調したい単語の直前と直後のポーズ長を特徴量とする．

これまで述べた特徴量を強調された発話から推定した結果の例を示す．図 4 は，強調したい単語を大きく，アクセントを明瞭に発声し，さらに直前にポーズを挿入して強調している例である．発話の中で，「凝縮」という単語が強調されている．上段のパワーの図では，「凝縮」の直前にポーズが挿入され，パワーの値が発話の中で最大値になっていることが分かる．中段の $\log F_0$ の図では，フレーズ成分の影響で，「凝縮」が発話の中で最大値になっていないことがわかる．下段の F_0 から推定したアクセント成分では，「凝縮」の場所が最大値になっており，強調されていることが明瞭に分かる推定結果となっている．

2.4　強調判定のための特徴量選択

強調判定の有効性を評価することで，特徴量を選択する．具体的には，あらかじめ強調か非強調かをハンドラベリングした学習データを用いて，判別モデルを構築し，評価データを判別分析して，正しく判別できるかを再現率と適合率で評価する．モデルは，ロジスティック回帰モデルを用いる．データは，強調 13 単語，非強調 13 単語，これとは別の評価用の強調 10 単

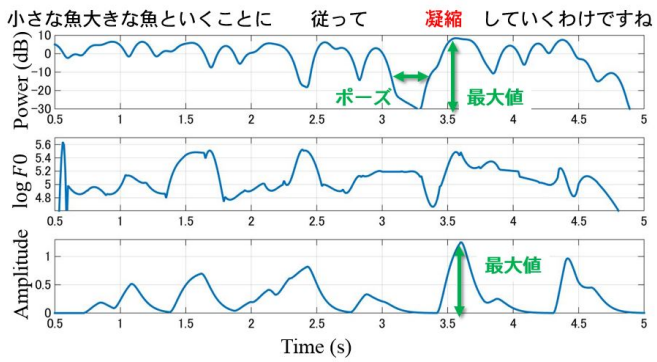

図 4. 単語が強調された発話の分析例　上段：パワーとポーズ　中段：logF_0　下段：アクセント成分

語，非強調 10 単語を用いる．

アクセント成分の最大値のみを特徴量とした強調の判定の有効性を評価する．その結果，再現率 1.00，適合率 0.91 となった．文頭に強調がある単語を除くアクセント成分最大値，強調している単語の直前，直後のポーズ長を特徴量とした強調の判定の有効性を評価する．その結果，再現率 0.73，適合率 0.80 となった．アクセント成分最大値と単語内のパワーの最大値を特徴量した，強調の判定の有効性を評価する．その結果，再現率 0.90，適合率 0.90 となった．アクセント成分最大値と単語内のパワーの平均値を特徴量した，強調の判定の有効性を評価する．その結果，再現率 0.90，適合率 0.82 となった．

提案システムでは，強調に聞こえない単語を強調していると判別するべきではないので，適合率を重視する．よって，最も高い適合率を算出したアクセント成分の最大値を特徴量とした判別が有効であると考えて，判別する際の特徴量として採用する．

なお，アクセント成分の最大値のみを特徴量とした判別が，単語の前後のポーズ長も含んだ重回帰より，高い結果となった理由として，分析したデータに，強調しようとしていないポーズが含まれていたからだと考える．つまり，強調するためのポーズと思考が発話に追いついていない場合のポーズが混同しているからである．さらに，強調を目的としていないポーズとして係り受けが考えられる．これらが含まれてしまったため，ポーズによる強調の判定結果が悪かったと考えられる．また，アクセント成分最大値とパワーの判別では，パワーが大きくする要因が強調以外ではなかったため，ポーズより結果が良かった．しかし，プレゼンでは普段より声が大きくなっているために大小で差をつけて強調することが難しく，アクセント成分最大値には及ばない結果になったと考えられる．

3 強調発話習得システム

システムの入力は，ユーザが強調したい単語を強調しようとして発声した発話である．解析では，入力発話に対し，F_0 を計算する．次に，F_0 から得られた特徴量から判別分析を行って，単語ごとに強調と判定される確率を計算する．出力は，単語ごとの強調と判定される確率，入力と見本のピッチ曲線，見本発話となる．見本発話は，分析から得られた強調したい単語のアクセント成分最大値から，声が裏返らずに強調に聞こえるように 0.1 刻みでユーザが手動で上げた値で，入力発話を変換することで生成される．強調判定のための学習データは，強調が 68 単語，非強調が 124 単語である．素早くフィードバックして練習者のやる気を削がないように，F_0 の分析には，高速に動作する手法，さらに，後に強調しているように変換して再合成するため，分析合成系のための手法を用いる [13]．入力発話と強調判定のための学習データの分析条件は，サンプリング周波数が 16 kHz，分析窓長が 25 ms，フレーム周期が 10 ms である．練習時の GUI(Graphical User Interface) の詳細について

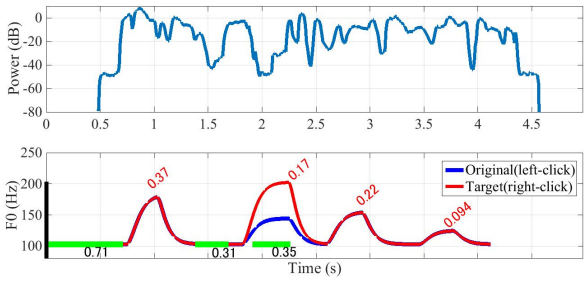

図 5. 練習時のインターフェース　上：パワー軌跡　下：見本と入力のピッチ曲線と強調判定確率

説明する．図 5 は，出力をフィードバックするインターフェースである．上部は，対数パワー軌跡で，無音区間などを確認する．インターフェースの下部の青い曲線と赤い曲線はそれぞれ入力発話と見本のピッチ曲線である．その上に示された赤字の数字が強調に判定される確率である．さらに，下の緑のラインはポーズを示す．また，入力した発話と見本発話は，それぞれ左クリックと右クリックで聞き比べることができる．

4 評価

4.1 実験方法

提案システムを用いた練習により，プレゼンで強調したい単語が強調して発声できるようになったかを評価する．発声訓練等の訓練を受けたことのない，学生 3 人（以降，被験者 A,B,C）が 2~3 分のプレゼンをする．テーマは，各自の研究の概要とした．強調したい単語はそれぞれのプレゼンに 5 個づつ含まれる．被験者は，各自のプレゼンに含まれる強調したい単語をあらかじめ選択しておき，プレゼンをする．次に，システムで練習する．練習は，強調したい単語が強調と判定されるまで行う．そして，再びプレゼンする．

主観評価では，練習前と練習後の発話それぞれに対し，発表者とは別の 3 人（以降，評価者）が音声を聞きながら，書き起こしの紙に単語ごとに 3 段階で強調のスコアを付ける（1:強調に聞こえない 2:少し強調 3:かなり強調）．さらに，客観評価では被験者が選択した強調したい単語が含まれる文を分析して，練習前と練習後の単語が強調に判定される確率を計算する．

4.2 実験結果

評価の結果を述べる．3 人の被験者の選んだ単語の練習前，練習後のスコアの平均と強調判定の確率を表 1，表 2，表 3 に示す．

表 1. 練習前後の平均評価スコアと判定確率 (被験者 A)

強調単語	前	後	前	後
分散	1.3	2.6	0.25	0.67
LSH	1.0	2.0	0.24	0.74
最近傍探索	1.0	2.0	0.34	0.59
局所鋭敏	1.7	2.4	0.33	0.50
画像特徴量	1.0	2.0	0.1	0.73

表 2. 練習前後の平均評価スコアと判定確率 (被験者 B)

強調単語	前	後	前	後
icp	1.3	2.3	0.41	0.72
キーフレーム	1.3	2.3	0.36	0.65
mpeg4	1.3	1.6	0.09	0.21
フレーム	2.0	1.6	0.53	0.67
位置合わせ	1.3	1.0	0.55	0.58

表 3. 練習前後の平均評価スコアと判定確率 (被験者 C)

強調単語	前	後	前	後
藤崎モデル	1.4	3.0	0.23	0.48
F_0	1.0	2.3	0.06	0.72
プレゼン	1.0	1.6	0.29	0.42
アクセント指令	1.0	3.0	0.26	0.62
強調	1.6	3.0	0.25	0.71

練習前のスコアの平均が 1.20，練習後のスコアの平均が 2.20 となった．これは，評価で定めた 3 段階で 1 段階評点が上がることを意味している．よって，練習前にスコアが 1 で強調に聞こえなかったとしても，練習後には強調して聞こえるようになっている．さらに，主観評価の結果を練習前，練習後のスコアの検定を行った．検定には，平均の差の t 検定を適用し，有意水準 α=0.05 とした結果，有意差が認められた．次に，客観評価の結果を述べる．練習前の平均は 0.29，練習後の平均は 0.60 となった．さらに，強調に判定される確率が練習前と後でそれぞれ 0.5 未満か 0.5 以上かを，平均の t 検定を適用し，有意水準 α=0.05 とした結果，有意差が認められた．よって，客観的にも強調が習得できていると言える．

被験者のコメントでは，練習時にピッチ曲線と共に強調に判定される確率が提示されることによって強調したい単語以外を抑えて発声することを意識できるようになった，とあった．さらに，評価者のコメントでは，練習前の発話は助動詞や接続詞が強調されて聞こえることが多かったが，練習後ではそれがなくなり，単語が強調されて聞こえるようになった，とあった．よって，強調の判定が強調習得に有効だったといえる．しかし，改善がされなかった単語として，「mpeg4」があり，主観評価では上昇したものの，客観評価では 0.21 で，非強調と判定されている．これは，「mpeg4」が平板型のアクセントであるからだと考えられる．平板型アクセントは，起伏式のアクセントより高くなりにくい傾向があること [5] が述べられている．そこで，平板型アクセントの単語の強調の仕方を調査するために，収集したデータに含まれている平板型の単語 3 つ（りそな，みずほ，cvcc）が強調されている発話を聴取したところ，全音節を強く発声していることがわかった．よって，これを意識することで改善が考えられる．さらに，これをシステムで強調と判定するためには，平板型の単語と起伏型の単語を判別する必要がある．そのため，F_0 からアクセント型を推定する試み [14][15] を利用して，単語を判別してから，別の判定モデルを用いることで，強調判定が行えると考える．

さらに，主観評価で満点だったが，客観評価で非強調に判定されてしまった単語があった．そこで，それらの単語（「藤崎モデル」,「プレゼン」）が含まれている発話を分析したところ，強調したい単語の強調判定確率が 0.5 に満たなかったものの，最大となっており，他が抑えられていることが分かった．本システムでは，単語ごとに絶対的に強調を評価しているが，評価の仕方を相対値にすることで判定が改善できると考える．

また，システム評価では被験者が選択した単語のみで練習したが．今後は被験者が練習した発話を別の話者が練習する等の実験や，アクセント型を網羅したコーパスを用いて大規模な実験を行うことで，強調しにくい単語や判定できない単語を明らかにできると考える．さらに，これを利用してユーザの習得の程度によっては，事前に練習文で練習するなどができると考える．

5 おわりに

本稿では，プレゼンでの強調習得を支援するシステムを構築した．強調を定量的に評価するために，特徴量としてアクセント成分の最大値を用いた．判定には，ロジスティック回帰分析を利用し，強調に判定される確率を計算した．さらに，これを利用して練習ができる GUI を作成し，判定確率を提示した．システムを用いて強調したい単語を強調できるようになったかを主観評価したところ，1.0 上昇し，本手法が強調習得の支援に有効であることを示した．今後は，平板型アクセントの単語に対する，強調の仕方単語の改善，判定，大規模なシステムの評価が今後の課題である．

参考文献

[1]『科学』編集部: “プレゼンテーションのコツ”，科学同人，1994

[2] 橋本和則：” Powerpoint マル勝プレゼンテーション術 ”，翔泳社，2001

[3] Francine R, et al:”THE USE OF EMPHASIS TO AUTOMATICALLY SUMMARIZE A SPOKEN DISCOURSE”, Acoustics, Speech, and Signal Processing, ICASSP-92., pp.229-232,1992

[4] Barry Arons:” PITCH-BASED EMPHASIS DETECTION FOR SEGMENTING SPEECH RECORDINGS” Proceedings of International Conference on Spoken Language Processing (September 18-22, Yokohama, Japan), vol. 4, pp1931-1934, 1994

[5] 杉藤美代子：” 講座日本語と日本語教育” , 明治書院，pp.316-339,2006

[6] 杉原満:” 音声表現から見る共通語の韻律理論”，2011

[7] 中川千恵子，他:”初級文型でできる にほんご発音アクティビティ”, アスク出版，pp.88,2010

[8] 中条 修:”日本語の音韻とアクセント”，勁草書房,pp.125-126,1989

[9] H.Fujisaki,”Vocal Physiology: Voice Production, Mechanisms and Functions” ,Raven Press,1988

[10] 成澤修一，他：“音声の基本周波数パターン生成過程モデルのパラメータ自動抽出法の評価,” 情報処理学会音声言語情報処理研究会, pp. 1-6, 2003

[11] 浅野泰史:” 音声合成のための文節単位での感情程度を考慮した統計的韻律制御”，東京大学大学院修士論文， pp.16-17,2006,

[12] Lawrence Rabiner, et.al:” Theory and Applications of Digital Speech Processing”,pp.605-607 ,2010

[13] M. Morise,et.al:“Fast and reliable F0 estimation method based on the period extraction of vocal fold vibration of singing voice and speech,” AES 35th International Conference, CD-ROM, London UK, Feb. 11-13, 2009

[14] 石井，他：” ピッチ知覚を考慮した日本語連続音声のアクセント型判定”，電子情報通信学会技術研究報告. SP, 音声 101(270), pp.23-30, 2001

[15] 佐々木，他：” 帯域制限ケプストラム法を用いた日本語単語アクセント型の判定”，日本音響学会春季講演論文集, pp.255-256, 2000