

プレゼンテーションにおける強調習得の支援*

小島 淳嗣 (法政大学 情報科学部), 伊藤 克亘 (法政大学 情報科学部)

1 まえがき

プレゼンテーション (プレゼン) を改善する際に, 見やすいスライドの作成法や内容, 発表の構成などは経過が残るため修正しやすいが, 話し方は, どこが悪いのかがわかりにくいいため, 改善しにくいという問題がある.

この問題に対し, これまで, プレゼンの話し方の練習に関する研究として, 基本周波数 (F_0) の標準偏差から発話の抑揚を評価し, 抑揚があるかないかをフィードバックする試み [3] がある. 本研究では, 抑揚の中でも特に強調を重視した練習システムを提案する. 強調は, プレゼンに含まれるキーワード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.

本稿では, 発話中の単語の強調に関する定量的な評価方法を提案する. さらに, これを利用した練習システムで, 発表者の練習時の発話を分析し, 単語の強調に判定される確率を練習者にフィードバックして, 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

まず, 本稿では, 強調 (focus/prominence) とは, 話者が聞き手に重要性を示すために発話の特定部分を際立たせること, と定義する. この際立たせ方に関して, 音声学では, 高さ, 長さ, 大きさ等の韻律の内, イントネーション (高さの変化) に関して頻繁に述べられる [5].

さらに, ポーズとの関連についても, 頻繁に述べられる [6]. ただし, 高さで強調していることが必須で, ポーズ単体では強調できない. このことから, 強調に重要なのは高さとポーズである. これらを用いて, 強調を定量的に評価する.

そのために, まず, アクセントが明瞭に聞こえるコーパスを収集し, 分析し, 強調判定に有効な特徴量を決定する. また, 練習時の目標にできるように, 強調が非強調かを識別するための基準を設ける. この基準は, 学習データから学習しておく.

2.2 強調発話の分析

学習データは, テレビ番組から収集した. 収集の際に, 正しく特徴量を計算するため, BGM, 共演者の声が入らない区間だけを切り出した. データは, 解説番組, 情報番組, 報道番組, 音声学の付属 CD から収集した. その結果, 113 発話収集した.

アクセントを明瞭に発音するには, 語に存在する High から Low に落ちる直前 (アクセント核) までになるべく高さを上げる必要がある. しかし, 話し声 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減少する. そのため, 観測 F_0 を時間構造で分割して,

評価すると, 文の前半と文の後半に位置する語を必ずしも平等に評価できない. そこで, F_0 が, 単語の局所的な変化に対応する成分と, 息の影響による全体にわたる変化に対応する成分からなるとする, 話し声 F_0 生成過程モデルである藤崎モデル [7] を用いる. このモデルの前者の成分に着目して, 文の位置に依存しない高さの特徴量を得る.

2.3 強調パラメータ計算

藤崎モデルのパラメータを逆問題として, F_0 から求める. パラメータ推定手法には, 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 [8] を参考にする. ただし, 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. 最後に, 分析区間 150ms で区分的に 3 次曲線でフィッティングする. 最終的に得られた, アクセント成分の最大値を強調判定の特徴量とする.

ポーズ長推定には, パワーとゼロクロスを利用する [10].

例えば, 図 1 は, 強調単語の直前にポーズを入れることによって強調している例である. "凝縮" が強調されている単語である.

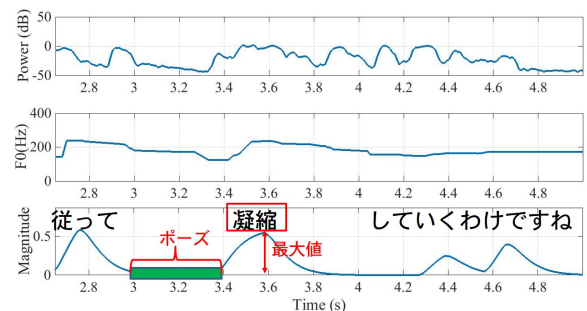


図 1. 強調する単語の直前にポーズが挿入される例 上段: パワー 中段: F_0 下段: アクセント成分とポーズ

2.4 強調判定

アクセント成分のみを特徴量とした強調の判定の有効性を評価する. 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評価データをロジスティック回帰モデルで判別分析し, 正しく判別できるかで評価する.

学習用, 評価用別々の単語で, それぞれ強調 18 単語, 非強調 80 単語で評価した.

再現率 66.7%, 適合率 96.0% となった.

さらに, 上記以外の, 文頭に強調がある単語を除くアクセント成分の最大値, 直前, 直後のポーズ長を特徴量した, 強調の判定の有効性を評価する. 学習用, 評価用別々の単語で, それぞれ強調 11 単語, 非強調 11 単語で評価した.

再現率 77.8%, 適合率 77.8% となった.

* Computer aided training for acquisition of focus on presentation.: Atsushi Kojima (Hosei Univ.) et al.

本システムでは、強調されていない単語を強調していると判別すべきではないので、適合率を重視する。そのため、高い適合率を算出した、アクセント成分の最大値のみを特徴量とした判別が有効であると考えられる。

3 強調発話習得システム

本システムの入力は、プレゼン時に強調したい単語を含む発話である。解析では、まず、入力発話に対し、 F_0 とポーズ長、ポーズ位置を計算する。次に、得られた F_0 から、藤崎モデルのパラメータを推定し、さらに単語ごとに強調に判別される割合を計算する。出力は、単語ごとの強調に判定される確率、入力と見本のピッチ曲線、注意点である。

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、特徴量の分析には、高速に動作する手法、さらに F_0 に関しては、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[11]。

練習時のGUI(Graphical User Interface)の詳細について説明する。図2は、そのインターフェースを示す。インターフェースの上部は、パワーで、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、元の発話と見本のピッチ曲線で再合成した見本の発話は、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。

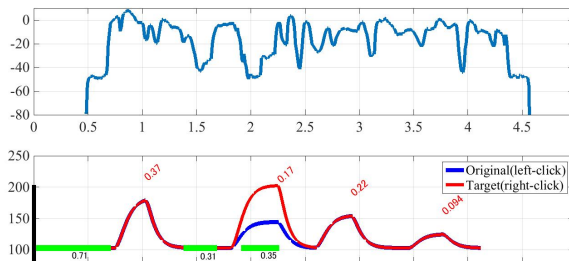


図 2. 練習時のインターフェース

4 評価

4.1 実験方法

提案システムでの練習により、発話中で強調したい単語が強調して発話できるようになったかを評価する。具体的には、発声訓練等の特別な訓練を受けたことのない、学生3人(以降、被験者)が2~3分のプレゼンをする。まず、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、システムで練習し、再びプレゼンする。プレゼン中の強調単語は5個とした。最後に、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、単語ごとに3段階で強調のスコアを付ける(3:かなり強調されている)。さらに、客観評価では、学習データから学習されたモデルを用いて、練習前と練習後の

単語が強調に判別される割合を見る。これにより、強調が習得できているか評価する。

4.2 実験結果

主観評価の結果を述べる。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。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う。検定には、ウィルチのt検定を適用し、

有意水準 $\alpha=0.05$ とした結果、有意差が認められた。

また、被験者のコメントでは、練習時に、ピッチ曲線ととも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下げることを意識できるようになったとあった。よって、強調に判定される確率の提示が強調習得に有効であるといえる。

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を、t検定を適用し、有意水準 $\alpha=0.05$ とした結果、有意差が認められた。客観的にも強調が習得できていると言える。

5 あとがき

本稿では、プレゼンにおける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、 F_0 に着目し、藤崎モデルを利用して強調に判定される確率を計算した。さらに、実際に練習ができるインターフェースを作成した。また、システムを用いてプレゼンで強調できるようになったかを主観、及び客観評価し、本手法が強調習得の支援に有効であることを示した。

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] 栗原, 他: “プレゼン先生: 音声情報処理と画像情報処理を用いたプレゼンテーションのトレーニングシステム”, WISS 第14回インタラクティブシステムとソフトウェアに関するワークショップ, pp.59-64, 2006
- [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] 杉原満: “音声表現から見る共通語の韻律理論”, 2011
- [6] 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
- [7] H.Fujisaki, “Vocal Physiology: Voice Production, Mechanisms and Functions”, Raven Press,1988
- [8] 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
- [9] 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,
- [10] Lawrence Rabiner, et.al: “Theory and Applications of Digital Speech Processing”, pp.605-607, 2010
- [11] 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.