6Q-06

# F0最大値とアクセント成分最大値を用いたプレゼンテーション音声の重要性強調判定*

小島 淳嗣†, 伊藤 克亘†, 花泉 弘†

## 1 まえがき

音声認識の発展に伴い, 言語情報だけでなく, パラ言語情報の認識が課題となっている [1]. 中でも, 強調の認識は重要な課題である. 強調は, 話者が聞き手にもっとも訴えたい部分で行われるため, 重要な情報を含んでいる [2]. 例えば, 語の重要性 [3] などの情報を含む. この情報は, プレゼンテーション (以降, プレゼンとする) の要約 [4] や対話における重要な箇所へのアノテート [5] など様々な場面で応用できる.

本稿では, 強調習得の練習システム構築のために強調語を判定する手法を提案する. 検出対象とする語は名詞とする.

## 2 日本語における単語の強調

我々は, 日本語の文中での単語の強調方法に関する知見を得るため, 文献 [2, 6–8] を調査した. 調査の結果より, アクセントの高さ, 語の強さ, 語の直前のポーズ長, 語の直後のポーズ長, 話速, 語の文中の位置を特徴量として用いる.

得られた知見を元に強調を定量化する. アクセントの高さ, 語の強さ, ポーズは我々が提案した手法で定量化する [9]. 具体的には, アクセントの高さは F0 に対し, 藤崎モデル [10] を仮定して, 各単語のアクセント成分の最大値 (A) を推定して求める. また,F0 最大値 (F) も用いる. 語の強さは, 語の対数パワー最大値 (Po) を求める. ポーズは, 語の直前, 直後のそれぞれのポーズの長さ (Pab,Paa) を求める.

話速は,2 つの方法で定量化する.1 つめの方法は, 単語の継続長 (Wl) であり, 2 つめの方法は, 単語内の 1 音節あたりの継続長 (Sl) である.

強調したい単語が文頭に近い場所に位置しているかを表すため, 発話中の単語の位置を定量化する. 本稿では, 発話長を 1 とする単語の開始位置 (Ws), 終了位置 (We) を求める.

強調単語を含む発話の分析から得られる単語の特徴量を図 1 に示す (経済学という学問です, という発話の分析結果).

## 3 強調発話コーパスの作成

提案手法の学習および評価のため, 強調された単語を含む発話を収集した. 我々は, すでにこの発話をテレビから収集し, コーパスとして整備する試みをしている [9]. 我々は, 新たに 303 発話整備した.1 発話の長さは 4.22s

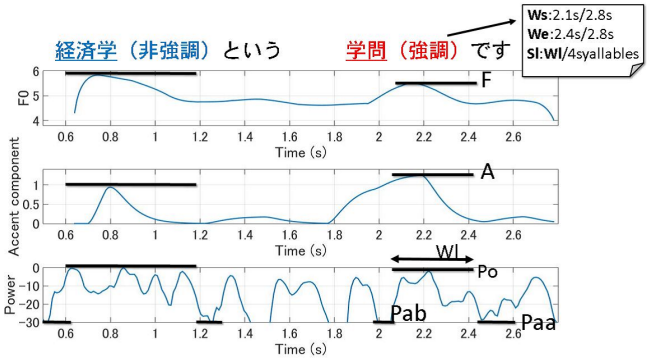


**図 1.** 音響特徴量の例

〜17.8s であった. 発話の長さの平均は 6.23s であった.

このコーパスは, 強調練習時の強調検出のための学習データに用いられる. そこで, 受聴者による強調の知覚の一致度を評価した. 強調ラベルを付けた作業者とは別の 3 人が, コーパスからランダムに選ばれた 50 発話を聞き, 強調している語にマークを付ける.3 人には, はじめにラベルを付けた作業者が強調に聞こえた語は教えない. マークの数は指定しない. また, 発話は何度聞き直してもよいとした.

評価実験の結果を表 1 に示す. 実験の結果, はじめにラベルを付けた作業者によって付けられた強調ラベルに対し,2 人以上が同じラベルをつけた語の割合は,0.91, 全員が同じラベルをつけた語の割合は 0.50 となった. 文頭に強調された単語がある時は, 全員が強調ラベルを付けていた.

**表 1.** 強調の知覚実験結果

| 人数 (人) | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 単語数 (語) | 6 | 9 | 10 | 10 | 35 |

## 4 実験

提案法による強調検出の有効性を検証するため,3 で収集した発話を利用して評価実験を行った. 評価には単語ごとに強調/非強調のラベルが付いた 978 単語 (強調 489 単語, 非強調 489 単語) を用いた. ラベルの情報は, 作業者 1 名が付与したものを用いる. 評価尺度としては, 適合率, 再現率, 分類精度を用いる.

$$\text{適合率} = \frac{\text{検出された強調単語数}}{\text{検出された単語数}}.$$

$$\text{再現率} = \frac{\text{検出された強調単語数}}{\text{強調単語数}}.$$

$$\text{分類精度} = \frac{\text{検出された強調単語数} + \text{検出された非強調単語数}}{\text{全単語数}}.$$

*Detecting an important emphasis in presentation speech using maximum F0 and maximum accent components.: *Atsushi Kojima (Hosei Univ.) et al.*

†法政大学大学院 情報科学研究科

比較手法は, [9] で提案された全ての特徴量 (A+F+Pab+Paa+Po) とする. F0 とアクセント成分はフレーム長 30ms, フレームシフトは 10ms で計算する. パワーはフレーム長 50ms, フレームシフトは 10ms で計算する. 特徴量の計算には hann 窓を用いる. 識別器には, ロジスティック回帰モデル [11] を用いる. そして, 学習したモデルを用いて求めた確率が 0.5 以上であれば強調とする.

強調検出に用いた特徴量と適合率, 再現率,F 値, 分類精度を表 2 に示す. 先行研究の特徴量に提案した特徴量を加えても性能は同程度だった. また, 全ての特徴量を組み合わせた時の精度 0.69 は, 英語で強調を検出した先行研究の精度 0.70 [3] と同程度となった.

さらに, 特徴量ごとに強調検出の実験をした. 検出に用いた特徴量と適合率, 再現率,F 値, 分類精度を表 3 に示す. F 値より, 最も有効な特徴量はアクセント成分である. 次いで F0 とパワーが同程度有効である. 単語継続長と単語開始位置は, 単語直前のポーズ長よりも有効である. また, 単語終了位置は単語直前のポーズ長と同程度有効である.

さらに, 学習データの数と強調検出の性能について評価する.978 単語 (強調 489 単語, 非強調 489 単語) から学習データをランダムに取り出し, 全単語から強調検出する実験を 1000 回行う. 学習データを 100 単語から 200 ずつ増やして 900 単語まで増やす. 学習データの数と強調検出の性能の関係を図 2 に示す. エラーバーは標準偏差を表す. 100 単語から 500 単語までの範囲で, 適合率, 再現率, 分類精度が上昇した. しかし,500 単語のデータを使って学習した時の適合率, 再現率, 精度は 900 単語を使って学習した時とあまり変わらなかった.

単語の文中の位置がポーズ長より強調検出に有効であることは, ある話題に関する発話が続く前に「今から何を話すのか見出しを先に示す」[8], という単語の文中の位置が知覚的に有効であるという知見と合致する.

また, 単語の継続長に比べて 1 音節あたりの継続長が強調検出に有効でなかったことに関して, 強調単語の 1 音節あたりの継続長の平均と非強調単語の 1 音節あたりの継続長の平均について調査した. その結果, それぞれ 0.45s と 0.41s となり有意差がなかった (危険率 0.05).

**表 2.** 強調検出の実験結果

| Feature | All | A+F+Pab+Paa+Po |
|---|---|---|
| 適合率 | 0.69 | 0.68 |
| 再現率 | 0.71 | 0.72 |
| F 値 | 0.70 | 0.70 |
| 精度 | 0.69 | 0.69 |

**表 3.** 特徴量ごとの強調検出の実験結果

| feature | A | F | Po | Wl | Ws | Pab | We | Sl | Paa |
|---|---|---|---|---|---|---|---|---|---|
| 適合率 | 0.62 | 0.59 | 0.65 | 0.61 | 0.53 | 0.64 | 0.52 | 0.51 | 0.49 |
| 再現率 | 0.71 | 0.61 | 0.56 | 0.51 | 0.55 | 0.44 | 0.53 | 0.49 | 0.21 |
| F 値 | 0.66 | 0.60 | 0.60 | 0.56 | 0.54 | 0.52 | 0.52 | 0.50 | 0.29 |
| 精度 | 0.64 | 0.59 | 0.63 | 0.60 | 0.53 | 0.60 | 0.52 | 0.51 | 0.50 |

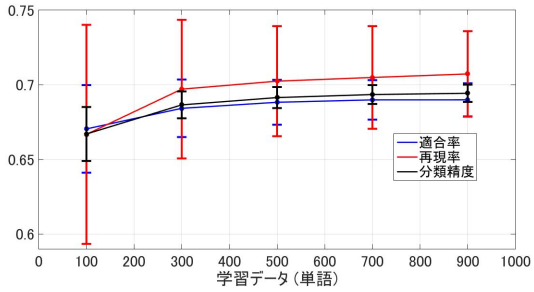


**図 2.** 学習データと強調検出の性能 (青線：適合率 赤線：再現率 黒線：分類精度)

学習データの数については, 500 単語を使って学習した時の性能と 900 単語を使って学習した時の性能が変わらなかったことから, データを増やすだけでは性能は大きく向上しないと考える. そのため, 先行研究に習って特徴量を選択するか特徴量を増やす必要があると考える.

## 5 あとがき

本稿では, 強調習得の練習システム構築のために発話から強調語を検出する手法を提案した. そのために, 音声学, 日本語教育学から得た日本語の単語の強調方法の知見を定量化することで強調検出に有効な特徴量を明らかにした. 具体的には, アクセント成分最大値,F0 最大値, 対数パワー最大値に加え, 発話中の単語の位置, 単語の継続長が有効であった. 強調検出の評価実験では, 精度 0.69 と, 英語における検出精度と同等の結果となり, 提案法の有効性が示された. 今後の課題は, データ整備については, 強調ラベルを付ける被験者の人数の検討である. 強調検出に関しては, 新たな特徴量の追加と提案した特徴量から有効な特徴量を選択する手法の検討となる.

## 参考文献

[1] 前川他, “音声はパラ言語情報をいかに伝えるか,” 認知科学, 9(6), pp. 46-66, 2002.
[2] 中条, “日本語の音韻とアクセント,” 勁草書房, 1989.
[3] V. K. R. Sridhar, et al., “Detecting prominence in conversational speech: pitch accent, givenness and focus,” in *Proc. SP*, pp. 453-456, 2008.
[4] R. Francine, et al., “The use of emphasis to automatically summarize a spoken discourse,” in *Proc. ICASSP*, pp. 229-232, 1992.
[5] Z. Malisz, et al., “Acoustic-phonetic realisation of polish syllable prominence: a corpus study of spontaneous speech,” Rhythm, melody and harmony in speech. Studies in honour of Wiktor Jassem, Vol.14/15, pp. 105-114, 2002.
[6] 郡他, “講座日本語と日本語教育 2 日本語の音声音韻,” 明治書院, 1989.
[7] 中川他, “初級文型でできる日本語発音アクティビティ,” アスク出版, 2010.
[8] 外山, “「伝わる話し方」のための 10 のルール：『実践日本語表現法』の授業現場から大学生の口頭表現を考える,” 愛知淑徳大学論集, 文学部・文学研究科篇 32, pp. 55-64, 2007.
[9] 小島他, “プレゼンテーション中の単語強調習得のための強調推定,” 日本音響学会秋季研究発表会, pp. 335-338, 2015.
[10] H. Fujisaki, et al., “A model for synthesis of pitch contours of connected speech,” Annual Report, Engg. Res. Inst., University of Tokyo, vol. 28, pp. 53-60, 1969.
[11] P. McCullagh, et al., “Generalized Linear Models,” New York:Chapman Hall, 1990.