JAIST Repository
https://dspace.jaist.ac.jp/

Title	アクセント核に着目した対比強調効果を有する音声の特徴抽出
Author(s)	大谷，泰博
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15886
Rights	
Description	Supervisor: 赤木 正人，先端科学技術研究科，修士（情報科学）

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修 士 論 文

アクセント核に着目した
対比強調効果を有する音声の特徴抽出

1710042　　大谷 泰博

主指導教員　赤木正人
審査委員主査　赤木正人
審査委員　鵜木祐史
党建武
吉高淳夫

北陸先端科学技術大学院大学
先端科学技術研究科 [情報科学]

平成31年2月

Abstract

Speech communication of humans is rich in expressions. It includes not only linguistic information but also para and non-linguistic information. However, synthesized speech cannot fully convey para-linguistic information yet. Emphasis is one of the important elements of para-linguistic information to convey intentions of speech contents. According to phonetics, speech emphasis is a part that makes differences with other parts. The emphasized part is made outstanding from other parts. Speech emphasis is realized by making voice size, length, and prominence. Humans can perceive not only presence/absence of emphasis but also degrees of emphasis from actual emphasized speech. However, humans cannot fully do from synthesized speech. Human can perceive strength of the speaker's intention from degrees of emphasis. Perceiving strength of intention from synthesized speech make speech communication using synthesized speech rich. Thus, it is necessary to synthesize emphasized speech that can convey degrees of emphasis to listeners. Pitch is the most important prosodic attribute for perceiving para-linguistic information. Fundamental frequency (F0) contours are one of the acoustic features elated to pitch. Many previous studies have synthesized emphasized speech focusing on F0 contours. In Japanese, pitch decreases rapidly from accent nucleus to next mora. Mora is the relative length of the sound which becomes the unit of strength and intonation. Accent nucleus is a mora just before the pitch decreasing. In addition, the peak of pitch decreases after the decreasing of the accent nucleus (catathesis or down step). In the case of emphasized speech, this phenomenon is hindered. Thus, we focus on the two features of F0 contours which related the variation of pitch in Japanese: decreasing of F0 from the accent nucleus and difference between accent nucleus of emphasized part and other accent nuclei in sentences. Also, we hypothesize that these features are important for synthesizing emphasized speech.

This study aims to clarify relationships between these features and degree of emphasis in order to evaluate the hypothesis. To clarify relationships, degrees of emphasis from the recorded voices are evaluated. In addition, decreasing of F0 contours from the accent nucleus of emphasized word to next mora are analyzed. F0 contours are expressed by using F0 at the barycentric point of the vowel (point pitch). Features in F0 contours are represented by calculating the difference of point pitch. In order to discuss relationships, it is necessary to know the degree of emphasis of each stimulus. In addition, it is necessary to know segment information of speech stimulus in order to extract point pitch from the F0 contours. Thus, a listening test is carried out to evaluate the degrees of emphasis of stimuli. A listening test was carried out to evaluate whether the stimuli are useful for analysis and the degrees of emphasis of the stimuli. This test is named as the experiment

1. Tokyo dialect utterances were used as the stimuli. Each stimulus was recorded with instruction of emphasizing one of the three noun words or non-emphasizing all words. Ten native Japanese students with normal hearing were participated in the experiment 1. The listening test was performed in a soundproof-room. The stimuli were randomly presented to the listener via a headphone. They were asked to evaluate not only presence/absence of emphasis but also degrees of emphasis in four steps (1 to 4). Degrees of emphasis were averaged in each stimulus.

It was necessary to segment speech stimulus to obtain the point pitch. Speech stimuli were segmented manually by using result of analysis obtained by using Praat. Segmentation was based on the knowledge of spectrogram. Speech stimuli were segmented into vowel, voiced consonant, and unvoiced consonant portions.

In order to clarify the relationships, it is necessary to analyze the two features and discuss the relationships. Point pitch, which was the value of F0 at the time of energy barycentric point, was extracted from F0 contour to analyze the two features. The F0 contour of each voice is obtained by using STRAIGHT(V40_005b) with frame length 40 ms, frame shift 1 ms and boundary of F0: 80 Hz - 600 Hz. Point pitch is extracted from the F0 contours. This study focuses on two amounts of decays: variation from accent nucleus to next mora and difference between accent nuclei in sentences. Then, relationship between degree of emphasis and features is compared to clarify the relationships. In order to discuss the relationship between degrees of emphasis and amount of decay, Amount of decay and Amount of growth are calculated. Amount of decay is decrease of point pitch from accent nucleus to next mora. Amount of growth is increase of point pitch from mora before accent nucleus to the accent nucleus. The degree of emphasis is a value derived from the result of the listening test. From the result, degrees of emphasis increase with amount of decay increase. On the other hand, the amount of growth does not change even if the degree of emphasis varies. In addition, point pitches do not change regardless of the presence/absence of emphasis or the change in degree of emphasis. From two results, it is considered that point pitch at accent nucleus of emphasized word increases as the degree of emphasis increases. In addition, point pitch at mora of one before accent nucleus increase according to increasing of point pitch at accent nucleus. Therefore, the presence or absence of emphasis changes when the amount of decay change.

In order to discuss the relationship between the degrees of emphasis and amount of decay, Amount of decay is calculated from point pitches of accent nuclei. Amount of decay is the difference between first and second word, difference between second and third word or difference between first and third word respectively. When the word is unaccented word, point pitch with the highest value is selected as point pitch of accent nucleus. From the results, difference of point pitches of accent nuclei between emphasized word and next word is more increasing. In addition, difference between emphasized word and before

word is more decreasing. However, amount of decrease does not change even if degree of the emphasis changed. Therefore, the presence or absence of emphasis changes when the amount of decay change. A listening test is conducted to evaluate whether people can perceive emphasis from stimuli or not, when modifying the F0 contour according to the two findings: amount of decay on emphasized word and difference of accent nucleus in sentences. This test is named as the experiment 2. Three kinds of stimuli are used in the experiment 2. The first stimuli are synthesized by using F0 contours analyzed with STRAIGHT. The second stimuli are synthesized by using F0 contour obtained from the point pitch. These are used to clarify whether the quality of synthesized speech using F0 contour obtained from point pitch is suitable. The third stimuli are synthesized by using F0 contours obtained from the point pitch, which are manipulated according to the two findings. The experiment 2 was carried out under the same procedure as the experiment 1. The evaluation results are averaged in each stimulus. From the experiment 2, we clarify that participants of the experiment 2 can perceive emphasis from synthesized speeches which were synthesized by using manipulated F0 contour. Therefore, the hypothesis is important for synthesizing emphasized speech. In addition, there are variations of degree of emphasis when amount of decay from accent nucleus varies. Therefore, it is considered that the variations may affect degrees of emphasis in human perception and relationship between degrees of emphasis and F0 contours may clarify by modeling variations.

目 次

第 1 章　序論 **1**
1.1 研究背景 1
1.1.1 ヒトの音声コミュニケーション 1
1.1.2 対比強調 1
1.1.3 先行研究 2
1.2 研究の目的 3
1.3 本論文の構成 3

第 2 章　アクセント核に着目した F0 軌跡 **5**
2.1 はじめに 5
2.2 本研究の着眼点 5
2.2.1 アクセント核前後の変化 5
2.2.2 ダウンステップ 7
2.3 着眼点の分析法 9
2.4 方法論 10

第 3 章　音声刺激の評価・処理 **11**
3.1 はじめに 11
3.2 音声刺激 11
3.3 実験 1 12
3.3.1 実験方法 12
3.3.2 実験結果 14
3.4 セグメンテーション結果 15
3.5 まとめ 16

第 4 章　点ピッチの変化に着目した強調の程度の変化の比較と評価 **17**
4.1 はじめに 17
4.2 強調単語のアクセント核の前後の F0 の変化量と強調の程度との関係の比較 17
4.3 文章中の各単語間における点ピッチの比較 23
4.4 実験 2 分析結果の評価 28
4.4.1 実験刺激 28
4.4.2 実験参加者 29

4.4.3 実験方法と実験装置 29
4.4.4 実験結果 29
4.5 まとめ 31

第 5 章 結論 **32**
5.1 本研究で明らかになったこと 32
5.2 今後の課題 32
5.2.1 強調の程度との関連 32
5.2.2 使用データの個人性 33
5.2.3 アクセント核が存在しない単語 33
5.2.4 強調に関連するその他の物理量の検討 33

参考文献 **34**

研究業績一覧 **37**

謝辞 **38**

図 目 次

1.1 論文の構成 4

2.1 アクセント核前後の F0 軌跡 6
2.2 ダウンステップの模式図 8
2.3 対比強調音声の F0 軌跡 9

3.1 実験に使用した GUI 13
3.2 Praat を用いたセグメンテーション 15

4.1 強調単語のアクセント型が 0 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係 19
4.2 強調単語のアクセント型が 1 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係 20
4.3 強調単語のアクセント型が 2 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係 21
4.4 強調単語のアクセント型が 3 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係 22
4.5 強調単語のアクセント型が 0 型の場合の強調の程度と各アクセント核の F0 の差の関係 24
4.6 強調単語のアクセント型が 1 型の場合の強調の程度と各アクセント核の F0 の差の関係 25
4.7 強調単語のアクセント型が 2 型の場合の強調の程度と各アクセント核の F0 の差の関係 26
4.8 強調単語のアクセント型が 3 型の場合の強調の程度と各アクセント核の F0 の差の関係 27
4.9 合成に使用した F0 軌跡 28

表 目 次

3.1 使用音声に含まれる名詞単語 12
3.2 使用音声の強調の有無の回答 14

4.1 0 型アクセント単語が強調された場合の音声の強調の有無の回答 30
4.2 1 型アクセント単語が強調された場合の音声の強調の有無の回答 30
4.3 2 型アクセント単語が強調された場合の音声の強調の有無の回答 30
4.4 3 型アクセント単語が強調された場合の音声の強調の有無の回答 30

第1章　序論

1.1　研究背景

1.1.1　ヒトの音声コミュニケーション

我々が普段行っているヒトとヒトとの音声コミュニケーションは表現豊かなものであり，電話越しなどの発話者の視覚的情報がない場合でも，音声から発話者の性別や感情といった個人性や心的態度などを認識できる．ヒトとヒトとの音声コミュニケーションの過程で発せられる音声信号は言語情報だけでなく，話者の年齢や性別，感情，個人性といった情報や発話の意図や話者の心的態度に関する情報を伝達することができる．ヒトが生成するこれらの情報は言語的情報，パラ言語的情報，非言語的情報に三分することができる [1]．機械が言語的情報だけでなく非言語的情報やパラ言語的情報の伝達や認識ができると，ヒトと機械が円滑なコミュニケーションを行うことができる．ヒトと機械が音声信号を用いたコミュニケーションがより円滑になると，ヒトが機械の複雑な操作をする必要がなくなるといった利点が存在し，多くの研究が人と機械の円滑なコミュニケーションの実現を目指している．

近年の機械の発展により,，ヒトと機械の間の表現豊かな音声コミュニケーションは実現可能になりつつある．音声認識技術の発展により，機械とヒトのコミュニケーションが画面越しのコミュニケーションだけでなく，Apple の「Siri」の音声アシスタント技術のような機械がヒトの音声を認識し，合成音声と画面情報で応えるような音声コミュニケーションをとることが可能になってきた．また，合成音声技術の発達により，ヒトが音声コミュニケーションの際に生成するような非言語的情報・パラ言語的情報も伝達可能な表現豊かな音声信号の合成が可能になってきた．これらの技術の発展によって，ヒトと機械のコミュニケーションがより表現豊かになり，円滑にコミュニケーションを行うことが可能になりつつある．しかし、音声コミュニケーションに利用される合成音声では，音声発話中に局所的に表れる表現（強調部分や句末音調部）については不十分な箇所が残っている [2] [3].

1.1.2　対比強調

ヒトが発話する自然音声は，伝えたい内容に合わせて声の大きさや抑揚，速さなどが一定ではなく，伝えたい内容の箇所の声の高さや大きさ，速さなどを相対的に変化させてい

る．このような音声発話中に局所的に表れる強調表現は現在の合成音声では表現が不十分である．これらの局所的な強調は発話者の意図というパラ言語的情報伝えるための重要な要素の 1 つである．音声学によると、音声による強調は理解や印象付けのために，声の大きさ・長さ・卓立の形成を行うことによって，一部が他から際立った箇所が作られることである．その類型として，前後の語群の差を伝える対比強調と強意強調の二つが示されている [4]．また，局所的な強調はプロミネンスや強調（focus）とも呼ばれる．本研究では，局所的な強調を対比強調と定義する．また対比強調効果のある音声を対比強調音声と定義する．

音声の対比強調は焦点の表出や並列構造や依存構造の曖昧性の解消，スケール合意などの機能を果たすといわれている [3] [5].

パラ言語情報は離散的であると同時に連続的な変化が生じ，表出に関わる特徴を調節することによって，意図の程度も表現することができる [1] [6]．そのため，対比強調の実現においても，強調の有無だけでなく任意の強調の程度（弱い強調や強い強調）を実現する必要がある．福岡らは，パラ言語情報の強調順位について，強調部分を高く長く発話した音声，強調部分を長く発話した音声，強調していない音声，発話速度の速い音声の順に強調の順位付けが行われると報告した [7].

1.1.3 先行研究

Ladd らは，発話意図と感情の知覚実験の結果，パラ言語的情報に対し，F0 が最も重要であることを明らかにした [8]．郡らは，音声の強さと長さが対比強調の実現に果たす役割を，音の高さが果たす役割と比較した結果，強さ及び長さが対比強調の実現に果たす役割は副次的であり，基本周波数 (F0) が重要であると報告した [9]．武田らは，対比強調の対象を文節全体だけでなく文節の一部の強め等の様々なタイプの対比強調について特徴を解析し，対比強調の主要な生成要因が抑揚，声の大きさ，発話速度，間等の「韻律」にあることを報告した [10].

音声合成については，白井らは，F0 について，2 つまたは 3 つの韻律語からなる文章において一つの韻律語を強調した場合の基本周波数変化パターンの特徴を解析し，音声合成のための単語の強調表現の規則化を行った [11]．武田らは報告した 4 つの韻律に関連する音響特徴の規則化を行い，対比強調音声合成を行った [12]．浜田らは，武田らが報告した 4 つの韻律特徴を制御できるインターフェースの設計を検討し,「弱め」「普通」「強め」の 3 段階の強調の程度を合成した [13] [14]．また，近年注目を集めている HMM（隠れマルコフモデル）に基づく音声合成技術 [15] [16] を用いて，森實らはスペクトル包絡，基本周波数，音素時間長などの音声特徴量をモデル化し強調音声の合成を行った [17]．現在の合成音声技術では強調の有無について実現されており，任意の強調の程度の表現は三段階でのみ実現されている

現在の音声合成技術では，任意の強調の程度を表現するためには多量のデータが必要である．また，合成音声は学習用データベースの特性を超えるものではなく，任意の強調の

程度を表現することが難しい．対比強調の有無に関しての韻律的特徴の解析はこれまでに行われているが，任意の強調の程度の音声合成を表現するためには，強調の程度に関連した特徴の解析が必要である．

1.2　研究の目的

本研究では対比強調の実現に F0 が重要であるとし，強調の有無や程度の違いによる F0 の変化を解析する．日本語ではアクセント核から次のモーラへ F0 が急激に減少し，アクセント核からの F0 の減少後の単語の 0F のピークは減少する特徴がある．そこで、アクセント核前後の F0 の変動やアクセント核の減少後の単語の F0 ピークの減少量の違いが強調音声に重要であるという仮説を立てた．本研究は仮説を評価するために，音声ごとにこれらの特徴と強調の程度との関係を比較し明らかにすることを目的とする．関係を明確にすることができれば，変動量や減少量をモデル化することで音声合成への応用が期待できる．任意の強調の程度を付加するための特徴が明らかになれば，伝える内容が複数あり， 内容に順位をつける必要がある音声合成への応用も期待できる．また，明らかにした特徴量の変動量を手がかりに，機械が音声から強調の程度という情報を認識できることが考えられる．

1.3　本論文の構成

本論文は，5 章で構成される．図 1.1 に本論文の構成を図示する．1 章は序論であり，本論文で対象とする研究課題と研究の目的を述べる．2 章は着眼点と方法について述べる．3 章は使用する音声データについてを述べる．4 章は音声の強調度と着眼点の関係とその評価について述べる．5 章は結論を述べる．

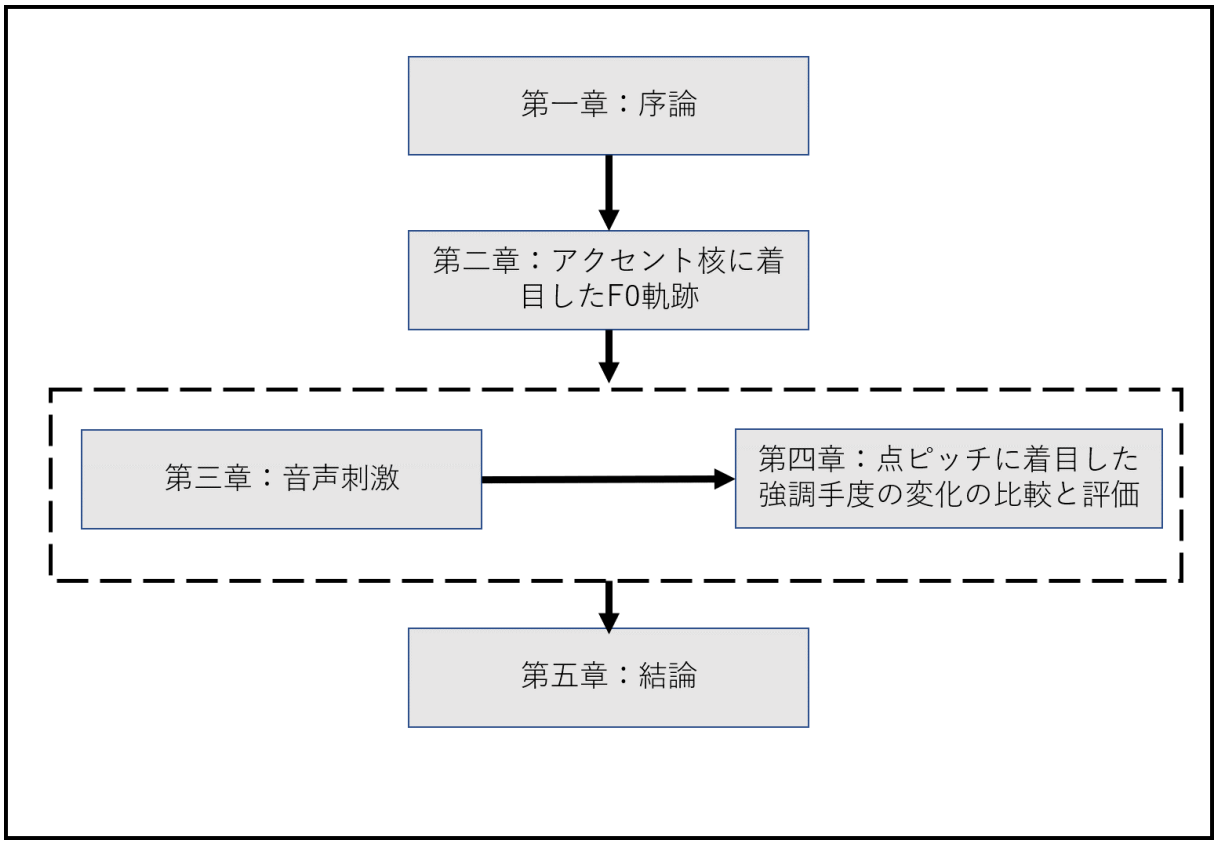
第一章：序論
第二章：アクセント核に着目したF0軌跡
第三章：音声刺激
第四章：点ピッチに着目した強調手度の変化の比較と評価
第五章：結論

図 1.1: 論文の構成

第2章　アクセント核に着目したF0軌跡

2.1　はじめに

対比強調の有無に関して関連する音響特徴については知られているが，強調の程度に関連している音響特徴やその変化が明らかではない．本研究では，対比強調に重要であるピッチに関連した日本語の特徴の音響特徴の変化が音声の対比強調に重要であると仮説を立てた．この章では，本研究で着目するピッチに関連した日本語の特徴について述べ，研究の目的のための方法論を述べる．

2.2　本研究の着眼点

2.2.1　アクセント核前後の変化

日本語において，対比強調される単語のF0のレンジが大幅に上昇することが知られている [20]．音響学的に音の高さが「高」から「低」に移るとき，高さが下がる直前の「高」のモーラにアクセント核があることがわかっている [18]．モーラとは，日本語音声における韻律の基本単位である．アクセント核の後ではF0の急激な下降が現れる．対比強調された単語のアクセント核の存在するモーラのF0の値が大幅に大きくなることから，アクセント核後のF0の急激な下降量も多くなる．図2.1にアクセント核前後のF0軌跡を示す．赤線は対比強調された場合のF0軌跡を，青線は対比強調されていない場合のF0軌跡を示す．発話内容は「ハナ¬ヨメカラ」で「¬」はアクセント核を示しており，モーラ「ナ」からF0の下降が見られる．対比強調された場合のアクセント核の存在するモーラとそれ以外のモーラではF0の増加量が異なり，F0の増加量は音声ごとに異なっていた．F0レンジの下限は個人ごとにほぼ一定であり，下限に比べ上限は変動幅が大きいことがわかっている [19]．よって，単語が対比強調されると強調単語のアクセント核前後のF0の増減量も多くなる．対比強調の知覚には後続する韻律句のF0が影響することがわかっている．従って本研究では，アクセント核前後のF0の増減量が強調の程度に関連があると考える．

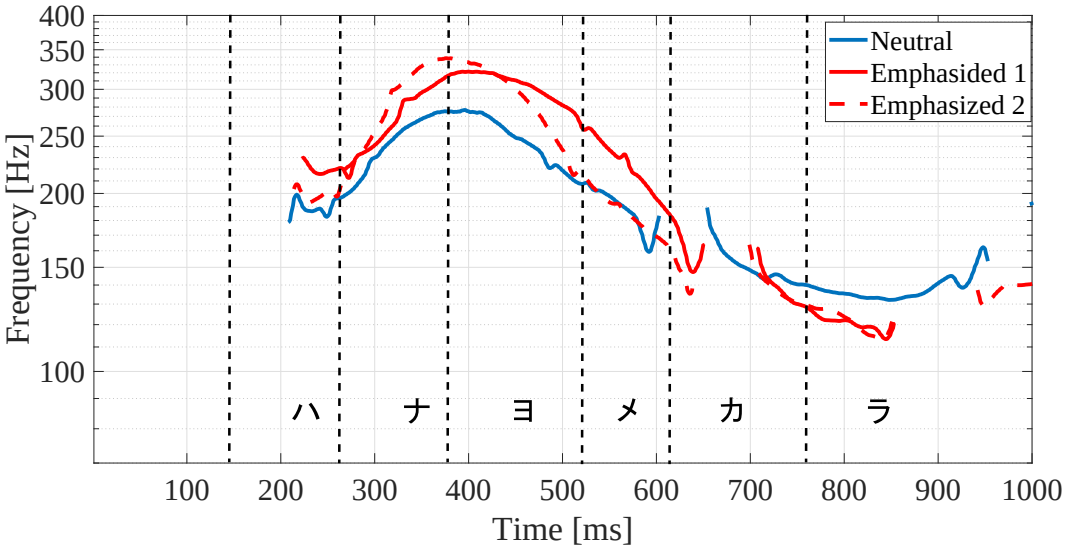

図 2.1: アクセント核前後の F0 軌跡

2.2.2 ダウンステップ

東京方言では，アクセント核による下降があると，それに後続するアクセント句の F0 ピーク値を反復的に低下させる現象がある．この現象はダウンステップまたはカタセシス（catathesis）と呼ばれる [20]．アクセント句はアクセント核がたかだか 1 個のみからなる句 [18] で，通常句頭の F0 の上昇によって定義づけられる．発話の後方に位置する単語が強調された対比強調音声の場合，この現象の効果が阻害され，後方に位置するアクセント句の F0 のピーク値が低下しないといった特徴がある．

図 2.2(a) にダウンステップの効果を模式した F0 軌跡を，図 2.2(b) にダウンステップの効果が阻止された F0 軌跡を示す．図 2.2(a)(b) はそれぞれ同じ発話内容の平静音声と対比強調音声の F0 軌跡を表す．発話内容は (a)(b) とも「花嫁から若者に料金を渡してほしい」である．対比強調音声は「料金」に強調が置かれている音声である．2 番目のアクセント句である「若者に」はアクセント核が存在しないため，図 2.2(a)(b) とも「若者に料金を渡してほしい」の発話間ではダウンステップは観察されない.「花嫁から」のアクセント核による下降によって生じるダウンステップは図 2.2(a) には観察されるが，図 2.2(b) には観察されず，対比強調されているアクセント句の「料金を渡してほしい」の F0 のピーク値が「花嫁から」のピーク値より大きくなっている．

後続するアクセント句の F0 ピーク値の低下量や，対比強調された場合のアクセント句のピーク値の上昇量は音声ごとによって異なっていた．そのため本研究では，ダウンステップの影響による F0 のピーク値の低下量が対比強調の程度に関連があると考える．

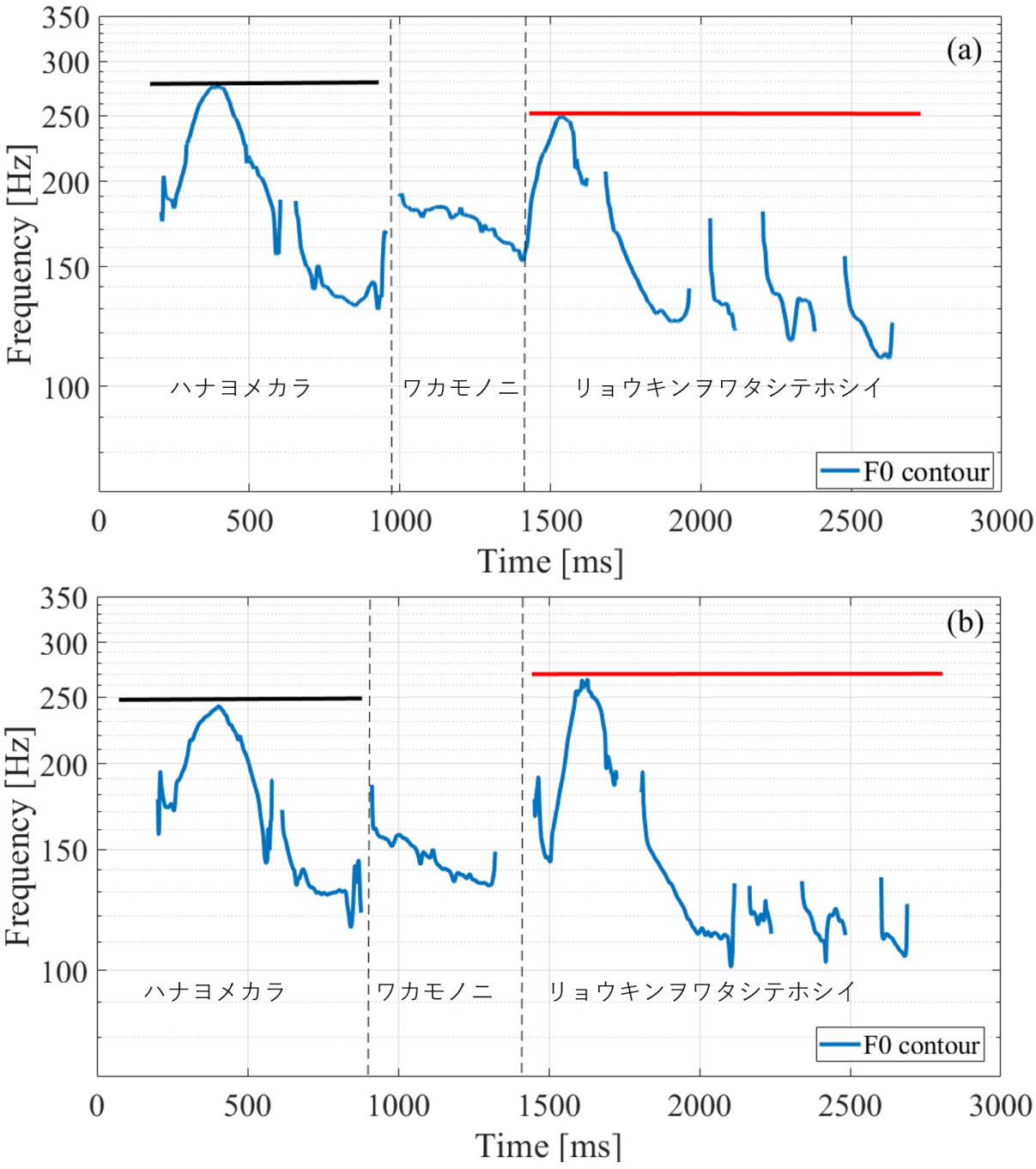

図 2.2: ダウンステップの模式図，(a) ダウンステップ，(b) 対比強調によるダウンステップの阻止

2.3　着眼点の分析法

本研究の目的のために，アクセント核の前後の F0 の変動量やダウンステップの影響による F0 のピーク値の低下量を分析する必要がある．F0 軌跡について母音のエネルギー重心点の時間位置における F0 の値が安定している．その値を点ピッチとし，点ピッチの間を直線内挿することで F0 軌跡を表現することができる [21] [22]．本研究の着眼点である F0 のアクセント核の前後の増加量・減少量やアクセント核ごとの F0 の差を，各モーラの点ピッチの差を計算することで求める．点ピッチを用いることで，モーラの声の高さと F0 の対応が取れ，少ない計算量で着眼点の変化量を求めることができる．各音声の F0 は STRAIGHT(V40_005b) [23] を用いて推定した (サンプリング周波数 44100 Hz，フレーム長 40 ms，フレームシフト 1 ms，F0 の推定範囲: 80 Hz - 600 Hz)．

図 2.3 に本研究の着眼点を示した対比強調音声の F0 軌跡の図を示す．縦軸は F0 の値を，横軸は時間を示す．緑枠で囲まれた箇所は強調が置かれた単語の F0 軌跡を示す．赤丸は各点ピッチを示す．矢印 (a) はアクセント核後の F0 の減少量を，(b) は各アクセント核の点ピッチの差を示す．図 2.3(a)(b) で示した F0 の増減量や差を対応する点ピッチの差分をとることで求める．

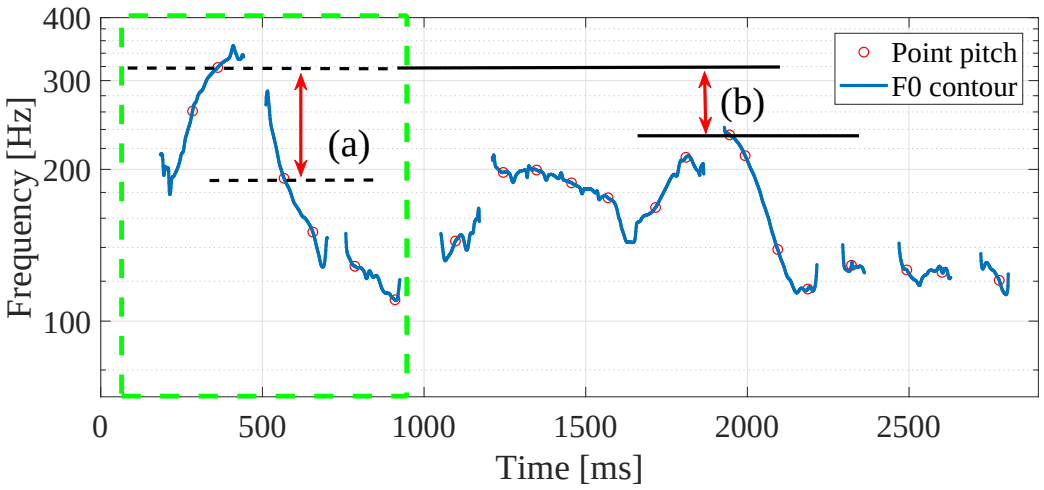

図 2.3: 対比強調音声の F0 軌跡

2.4　方法論

対比強調の音声学的実現には，任意の強調の程度を表現することが必要である．任意の強調の程度を表現するためには, 強調の程度に関連した特徴の解析が必要である．

本研究は，アクセント核からの下降やダウンステップといったF0軌跡の特徴に着目し，アクセント核前後のF0の増減量やダウンステップの影響によるF0のピーク値の低下量が強調の程度と関連すると仮説を立てた．本研究の目的は，各音声の着目したF0軌跡の特徴と強調の程度との関係を比較し，明らかにして仮説を評価することである．本研究の目的のために以下のように分析，評価を行う．

音声刺激の評価・処理：各音声刺激に対比強調が含まれているか評価するため，強調の程度を明確にするために主観評価実験を行う．また，対比強調音声分析行うための処理としてセグメンテーションを行う．セグメンテーションにはPraatを用いた分析結果を手がかりに行う．

点ピッチの変化に着目した強調程度の変化の比較と評価：アクセント核前後のF0の変動量や音声中のアクセント核ごとの差を，点ピッチの差を計算することによって求める．各音声刺激のF0情報はSTRAIGHTを用いて抽出する．各音声刺激の強調の程度と求めた値との関係を比較することでF0軌跡の特徴と強調の程度との関係を分析する．仮説の評価を，分析結果に従って合成された音声を用いた主観評価実験によって行う．

第3章　音声刺激の評価・処理

3.1　はじめに

本章では，対比強調音声の分析をするために，使用する音声刺激の評価・処理を行う．強調の程度と着眼点の F0 軌跡の変化の関係を明らかにするために，使用する音声の強調の有無や程度の情報を明らかにする必要がある．加えて，着眼点の F0 軌跡の変化を分析するために，音声のセグメント情報が必要である．

本章では実験 1 として，音声の強調の有無や程度について明らかにするために聴取実験を行った．また各音声に対してセグメンテーションを行った．

3.2　音声刺激

本研究で使用する音声は，複数のアクセント句から構成され，強調部と非強調部が存在する必要がある．また，アクセント核に着目するため，各アクセント句に含まれる単語のアクセント型について統一のとれているデータが必要である．

本研究では，強調の説明とともに「その部分を間違いなく相手に伝えようとする」表現を設定された文章に対して行ってくださいという教示とともに録音された音声を用いる．音声はサンプリング周波数 44,100 Hz 量子化ビット数 16 bit で録音された．音声の文章の構造は 3 個のアクセント句から構成されており，強調ありの教示で録音された音声は 3 つのアクセント句のうちのいずれかが強調されていた．各アクセント句に 4 モーラの名詞単語が含まれていた．単語のアクセント型は，アクセント核がないものを 0 型，アクセント核の位置が単語の先頭から数えて 1 モーラ目にあるものを 1 型，2 モーラ目 3 モーラ目にあるものをそれぞれ 2 型，3 型と分類された．音声のアクセント型は平板型，頭高型，中高型，尾高型 [20] [24] の 4 種類が存在した．単語はアクセント型を考慮したうえで，強調されても違和感のない親密度が高いものが選ばれた．使用する音声は 3 つのアクセント句と 4 種類のアクセント型の組み合わせた 64 種類の文章が存在した．各文章で 4 種類の強調の種類（強調なし，強調箇所が 1 番目のアクセント句，2 番目のアクセント句，3 番目のアクセント句）の存在したため，音声刺激の総数は 256 種類であった．使用音声に含まれる名詞単語を表 3.1 に示す．

表 3.1: 使用音声に含まれる名詞単語

		単語の箇所		
		第 1 単語	第 2 単語	第 3 単語
アクセント型	0 型	チチオヤ	ワカモノ	イレモノ
	1 型	ニイサン	チョウチョウ	リョウキン
	2 型	ハナヨメ	ナコウド	エハガキ
	3 型	ベンゴシ	センセイ	ベントウ

3.3 実験 1

本研究で用いる音声の強調の有無や程度について明らかにし，各音声刺激の強調の有無と程度のラベル付けを行うために聴取実験を行った．

3.3.1 実験方法

実験参加者

北陸先端科学技術大学院大学の 20 代の成人男女 10 名（男性 7 名女性 3 名）が実験に参加した．すべての被験者は正常聴力を有していた．

実験方法と実験装置

実験には本研究で使用する 256 個の音声を用いた．実験刺激を被験者にランダムで提示した．実験刺激の提示制御は PC(Windows 8, MATLAB 2015a) で行い，オーディオインターフェース (Fireface UCX) で DA 変換したのちアンプ (STAX SRM-1/MK-2pp) で信号増幅し，ヘッドホン (STAX SR-404) で提示した．各刺激は 1 度だけ被験者に提示した．実験は防音室で行われた．実験では GUI を用いて，各刺激に対して被験者に強調の有無を回答するよう求めた．加えて，強調があると回答した場合には強調の程度を 4 段階 (1〜4) で回答するよう求めた．実験は休憩を含め，60 分程度で行われた．図 3.1 に実験に使用した GUI を図に示す．

図 3.1: 実験に使用した GUI

3.3.2 実験結果

各音声の強調の有無の回答の結果を表 3.2 に示す．実験結果より，対比強調の教示で録音された音声は対比強調がありの回答率が 97.9%，対比強調なしの教示で録音された音声の対比強調がなしの回答率は 82.3%であった．対比強調なしの教示で録音された音声から対比強調が知覚された箇所の多くは文頭の第 1 単語と文末の第 3 単語であった．各音声の強調の程度は，各音声の強調の程度の回答の平均値を用いた．よって，今回使用する音声刺激の対比強調の有無が明らかになった．また，各音声の強調の程度が明らかになった．

表 3.2: 使用音声の強調の有無の回答

		回答箇所			
		強調なし	第 1 単語	第 2 単語	第 3 単語
対比強調の教示箇所	強調なし	82.3%	8.9%	0.8%	8.0%
	第 1 単語	2.2%	96.9%	0.9%	0%
	第 2 単語	0.2%	0.6%	99.1%	0.2%
	第 3 単語	1.7%	0.5%	0%	97.8%

3.4　セグメンテーション結果

アクセント核前後の F0 の増減量や音声中の各アクセント句のアクセント核の F0 の値の差を点ピッチを用いて求めた．各点ピッチを計算するために，使用する音声刺激はセグメントされていなければならない．セグメンテーションは Praat を用いた分析結果を使用し，発話内容に対応する音韻のスペクトログラム特徴 [25] に基づいて手動で行った．図 3.2 にセグメンテーション例を示す．

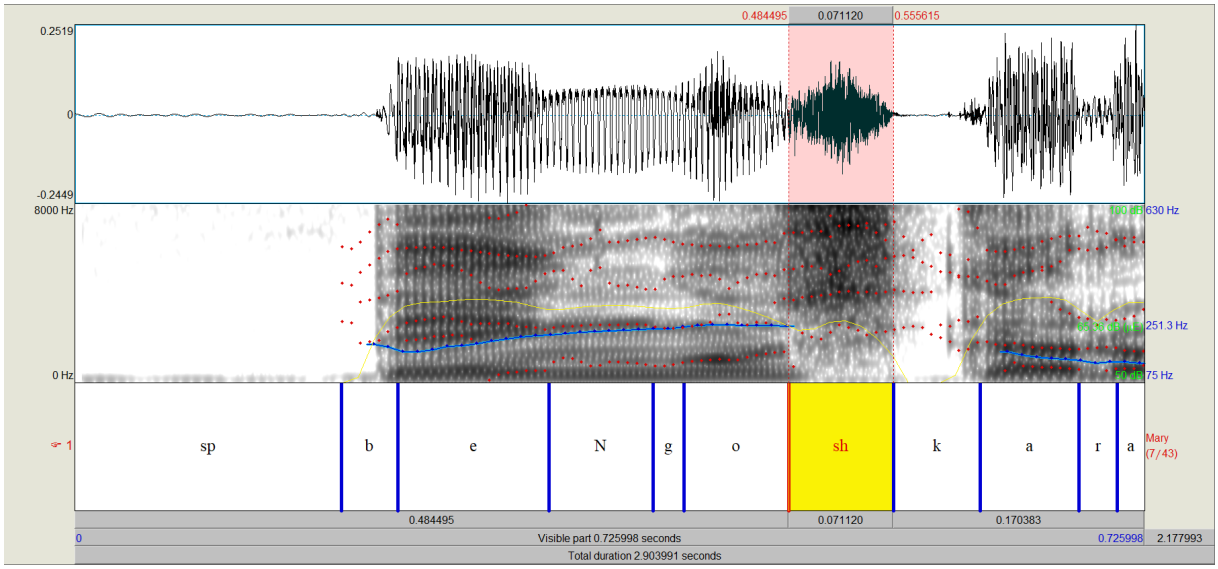

図 3.2: Praat を用いたセグメンテーション

3.5　まとめ

本研究を進めるにあたって，複数のアクセント句から構成されており，各アクセント句に含まれる単語のアクセント型について統一のとれた音声データが必要である．使用する音声は強調の有無の教示で録音されたアクセント型について統一のとれたデータを用いる．使用する音声の強調の有無や程度の情報を明らかにするために実験 1 が行われた．実験 1 より各音声刺激の強調の有無や強調程度の情報が明らかになった．また，各音声に対しセグメンテーションを行うことでセグメント情報が得られた．よって，本研究で使用する音声刺激は対比強調音声の分析をするために有用であるとした．

第4章　点ピッチの変化に着目した強調の程度の変化の比較と評価

4.1　はじめに

強調の程度と本研究の着眼点（アクセント核の前後のF0の増加量・減少量，ダウンステップの影響によるF0のピーク値の低下量）との関係を明らかにするために，各音声の着眼点のF0の変動量と強調の程度の関係を求め，比較する必要がある．各音声の関係を求めるために，本研究の着目点のF0変動量を計算する必要がある．橋本らは，母音部の振幅包絡から定まるエネルギー重心の時間位置における基本周波数の値（点ピッチ）が安定していることから，各モーラの点ピッチによってF0軌跡の変化パターンを表現することを提案し，これを音声合成に適応した．橋本が提案した点ピッチの差を計算することで，アクセント核の前後のF0の増加量・減少量，ダウンステップの影響によるF0のピーク値の低下量を求めることができる．本章では，点ピッチを用いて着眼点のF0変動量の計算を行い，各音声の強調の程度と着眼点の関係を比較することで強調の程度と本研究の着眼点の関係を明らかにする．

4.2　強調単語のアクセント核の前後のF0の変化量と強調の程度との関係の比較

各音声の強調の程度とアクセント核の前後におけるF0の増加量・減少量の関係を比較する．各音声の強調の程度は実験1の結果から得られた値を示す．アクセント核の前後のF0の増加量 A_g・減少量 A_d はアクセント核の点ピッチ，アクセント核前後のモーラの点ピッチの差から計算する．

$$A_d = \ln x_2 - \ln x_1 \tag{4.1}$$

$$A_g = \ln x_2 - \ln x_3 \tag{4.2}$$

x_1 はアクセント核の前のモーラの点ピッチの値を，x_2 はアクセント核の存在するモーラの点ピッチの値を，x_3 はアクセント核の次のモーラの点ピッチの値を示す．図4.1から図4.4に強調単語がそれぞれ0から3型のアクセント型の場合の各音声の強調の程度とF0の増加量・減少量の関係を比較した図を示す．横軸はアクセント核の前後のF0の増加量減少

量を示している．縦軸は実験 1 で得られた各音声刺激の強調の程度を示している．0 型の単語にはアクセント核が存在しないため，対比強調のない音声の F0 のピークが存在するモーラをアクセント核とみなし，その前後の F0 の増加量・減少量と強調の程度の関係を比較した．アクセント型が頭高型の単語の場合，アクセント核の前のモーラの点ピッチが存在しないため A_g は計算できなかった．また，文頭の強調された単語のアクセント型が 3 型単語のとき，アクセント核後のモーラが無声化していたため A_d は計算できなかった．

図 4.1(a)(c)(e) より，0 型アクセント単語が強調される場合，A_g の値は強調の位置が文頭の場合は 0 付近の値であった．また，A_g の値の違いによる強調の有無の違いは見られなかった．しかし．単語が文中の場合は，F0 軌跡はアクセント核を持つ単語と同様に F0 増加が見られ，A_g の値もアクセント核を持つ単語と同様の値を示した．また，A_g の値の違いによる強調の有無の違いが見られた．図 4.1(b)(d)(f) より，A_d は 0 付近の値を示し，F0 の大きな変化はなく，平板型の特徴を示していた．また，A_d の値の違いによる強調の有無の違いは見られなかった．

図 4.2 より，A_g の違いによる強調の有無や程度の違いは見られなかった．しかし，対比強調音声の A_d の値は平静音声の値の 1.6 倍であり，A_d の変化量の増加が示された．図 4.3(b)(d)(f)，図 4.4(c)(e) より，強調単語が 2 型アクセント単語や 3 型アクセントの場合も同様に，対比強調音声の A_d の値は平静音声の値の 1.6 倍であり，A_d の変化量の増加が示された．

図 4.3(a)(c)(e) より，A_g は強調単語が 2 型の場合には，対比強調音声の A_g の値は平静音声の値の 1.3 倍であり，A_g の変化量の増加が示された．図 4.3(a)(b)(d) より，強調単語が 3 型の場合では，A_g は単語が文頭の場合は 2 型アクセント単語と同様の結果であった．しかし，単語が文中にある場合の対比強調音声の A_g の値は平静音声の値を下回っており，A_g の変化量の減少が示された．

したがって，強調単語がアクセント核を持つ単語の場合，アクセント核からの F0 の減少量が変化すると強調の有無が変化する．また，F0 の減少量が異なると強調の程度に違いがあった．アクセント核までの F0 の増加量について，アクセント型が 2 型 3 型の場合は増加量が変化すると強調の有無が変化する．また，アクセント核のない単語の場合，強調単語が文中に位置する場合，アクセント核までの F0 の増加量が変化すると強調の有無が変化する．

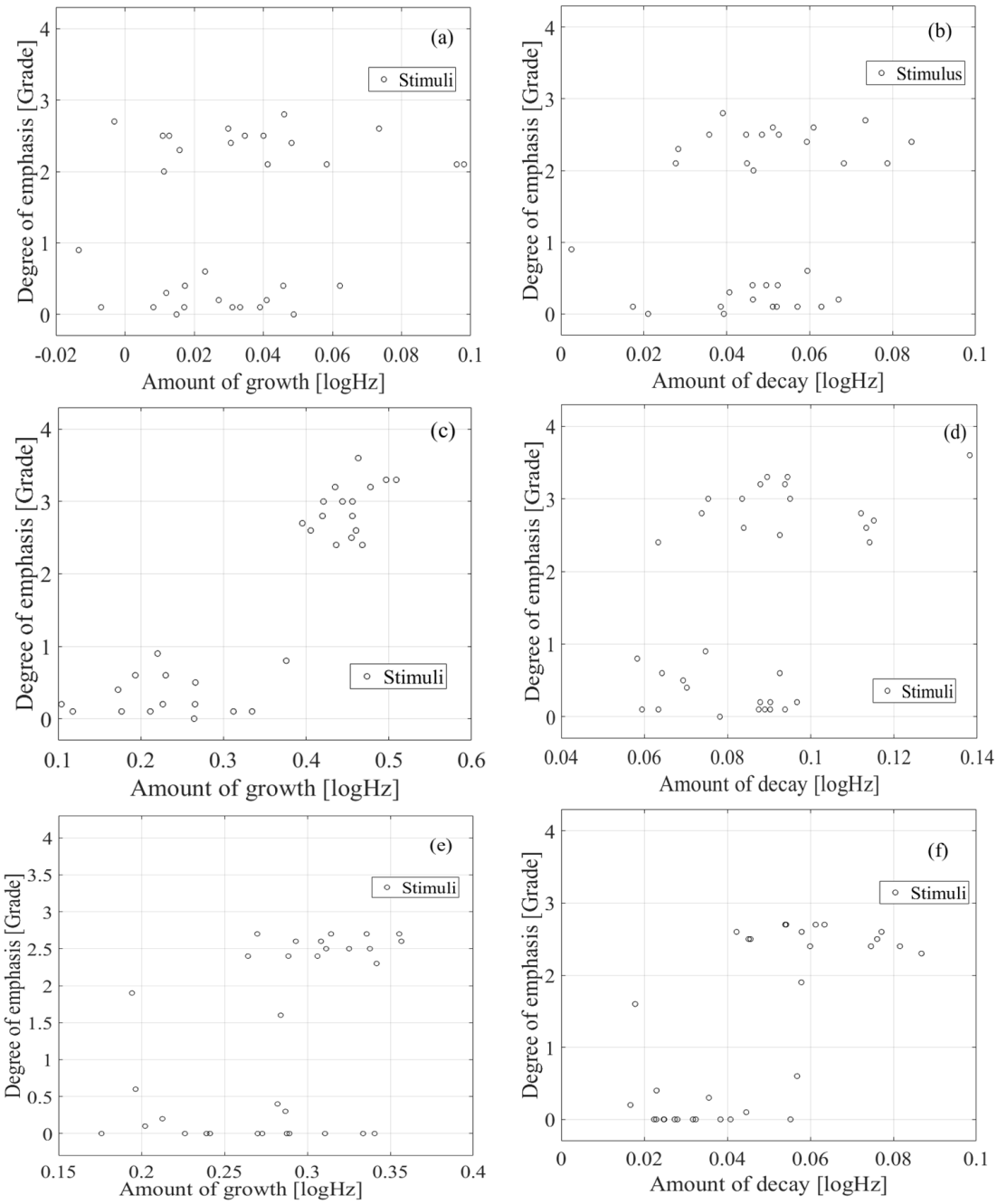

図 4.1: 強調単語のアクセント型が 0 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係，(a)1 番目の単語が強調された場合のアクセント核までの F0 の増加量，(b)1 番目の単語が強調された場合のアクセント核からの F0 の減少量，(c)2 番目の単語が強調された場合のアクセント核までの F0 の増加量，(d)2 番目の単語が強調された場合のアクセント核からの F0 の減少量，(e)3 番目の単語が強調された場合のアクセント核までの F0 の増加量，(f)3 番目の単語が強調された場合のアクセント核からの F0 の減少量

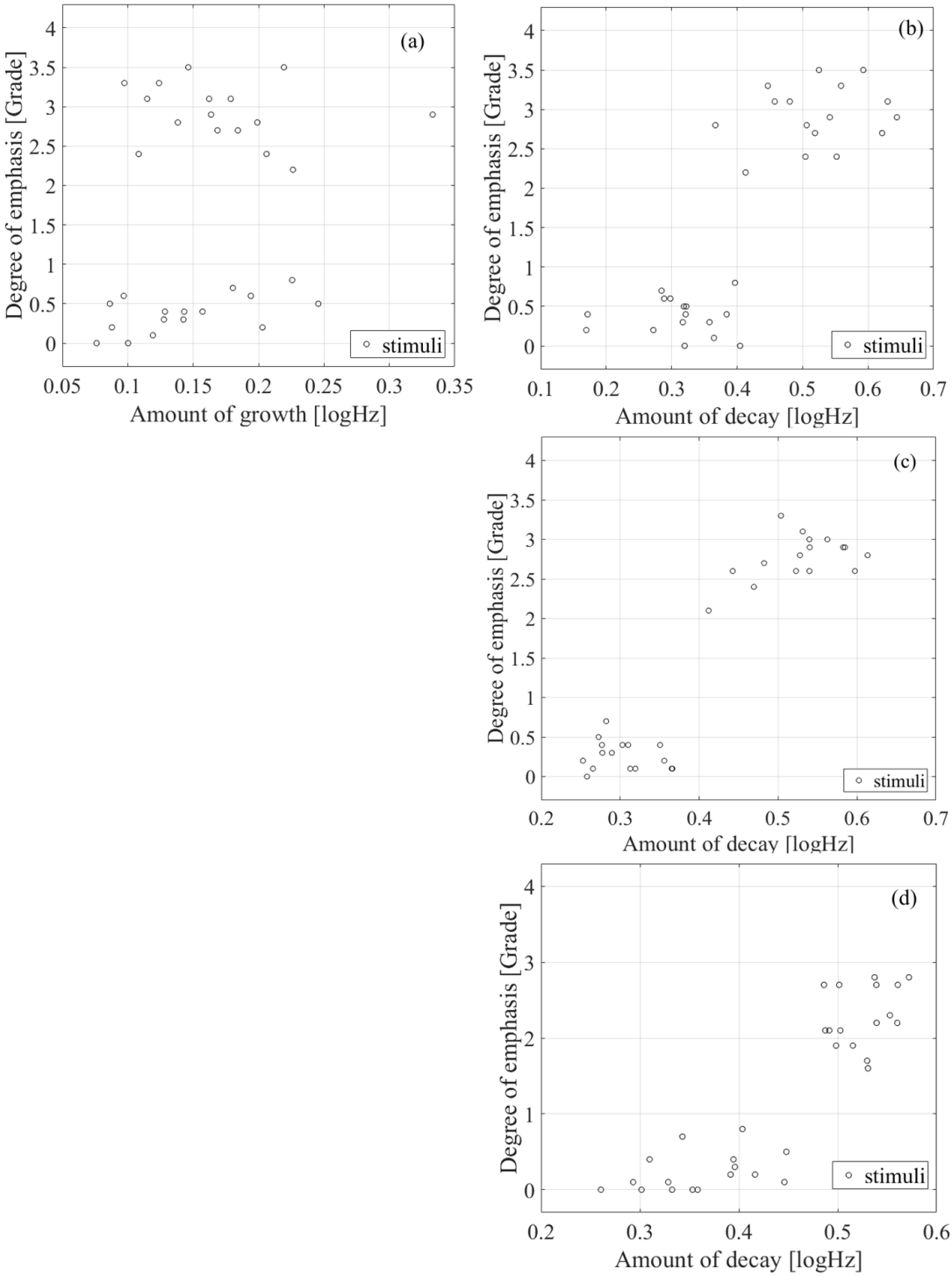

図 4.2: 強調単語のアクセント型が 1 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係，(a)1 番目の単語が強調された場合のアクセント核までの F0 の増加量，(b)1 番目の単語が強調された場合のアクセント核からの F0 の減少量，(c)2 番目の単語が強調された場合のアクセント核からの F0 の減少量，(d)3 番目の単語が強調された場合のアクセント核からの F0 の減少量

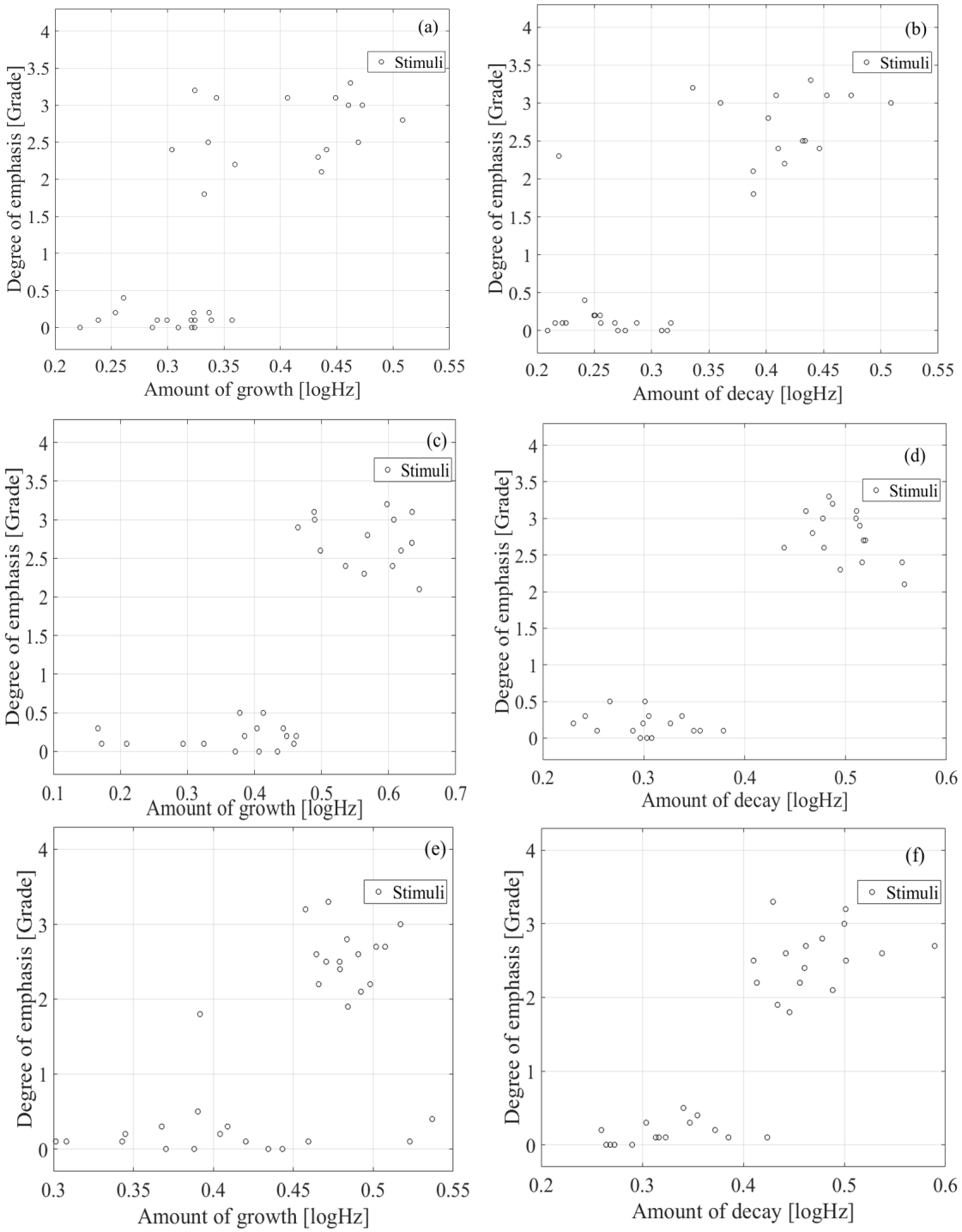

図 4.3: 強調単語のアクセント型が 2 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係，(a)1 番目の単語が強調された場合のアクセント核までの F0 の増加量，(b)1 番目の単語が強調された場合のアクセント核からの F0 の減少量，(c)2 番目の単語が強調された場合のアクセント核までの F0 の増加量，(d)2 番目の単語が強調された場合のアクセント核からの F0 の減少量，(e)3 番目の単語が強調された場合のアクセント核までの F0 の増加量，(f)3 番目の単語が強調された場合のアクセント核からの F0 の減少量

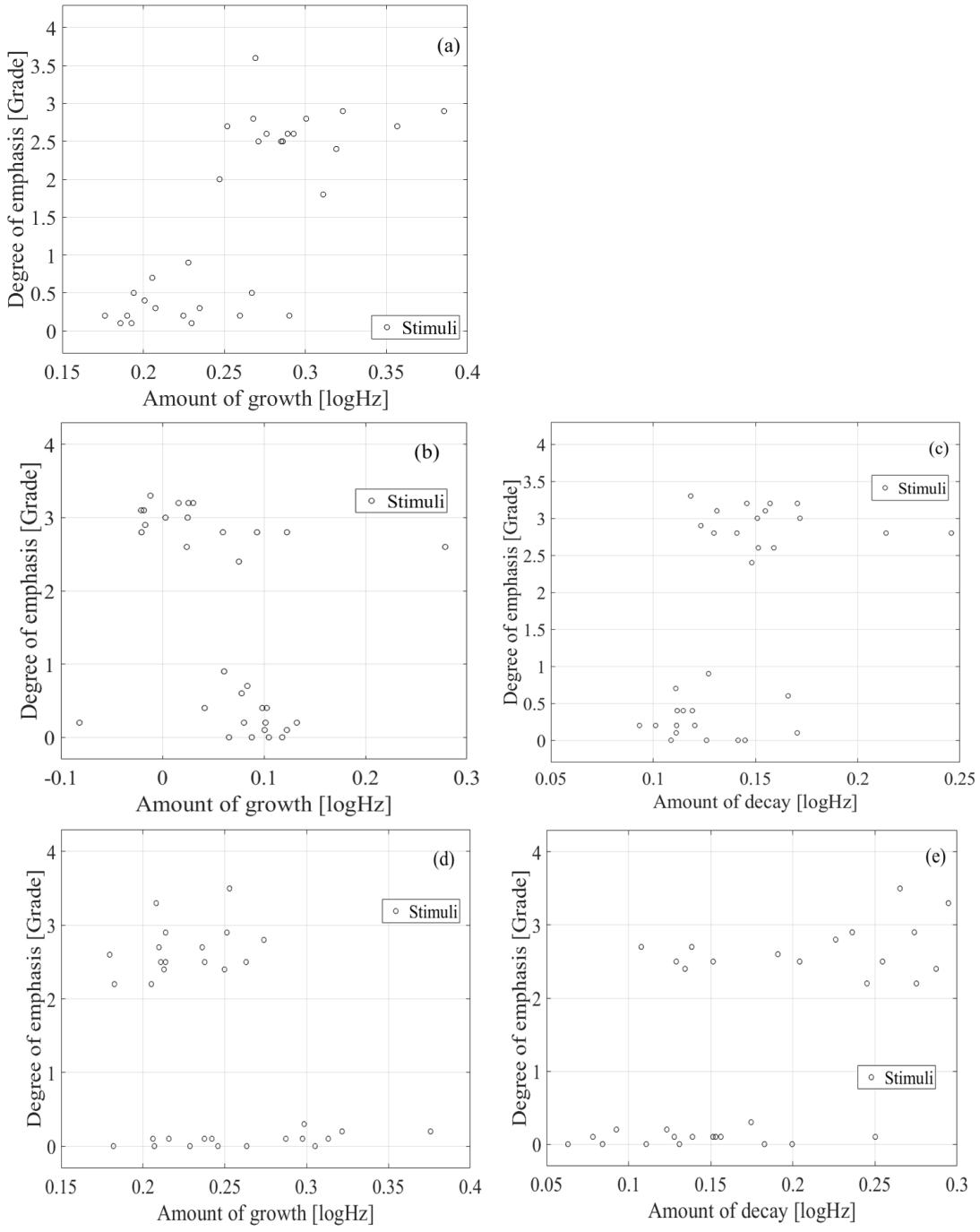

図 4.4: 強調単語のアクセント型が 3 型の場合の強調の程度とアクセント核前後の F0 の変化量の関係，(a)1 番目の単語が強調された場合のアクセント核までの F0 の増加量，(b)2 番目の単語が強調された場合のアクセント核までの F0 の増加量，(c)2 番目の単語が強調された場合のアクセント核からの F0 の減少量，(d)3 番目の単語が強調された場合のアクセント核までの F0 の増加量，(e)3 番目の単語が強調された場合のアクセント核からの F0 の減少量

4.3　文章中の各単語間における点ピッチの比較

各音声の強調の程度と各アクセント核の差の関係を比較する．各音声の強調の程度は実験 1 の結果から得られた値を示す．文章中の各単語の F0 の違いは各アクセント核の点ピッチの差から計算する．

$$A_d(a) = \ln a_1 - \ln a_2 \tag{4.3}$$

$$A_d(b) = \ln a_2 - \ln a_3 \tag{4.4}$$

$$A_d(c) = \ln a_1 - \ln a_3 \tag{4.5}$$

a_1 a_2 a_3 はそれぞれ 1 番目 2 番目 3 番目の単語のアクセント核の点ピッチの値を示す．

図 4.5 から図 4.8 に強調単語がそれぞれ 0 から 3 型のアクセント型の場合の各音声の強調の程度とアクセント核の前後における F0 の増加量・減少量の関係を比較した図を示す．それぞれの図の (a) は 1 番目と二番目の単語のアクセント核の差と強調の程度との関係を示している．(b)(c) は同様に 2 番目と 3 番目，1 番目と 3 番目の単語のアクセント核の差と強調の程度との関係を示している．黒星印は強調なしの場合の音声の関係を示し，赤丸，青四角，緑三角，はそれぞれ 1 番目，2 番目，3 番目の単語が強調された場合の音声の関係を示している．

図 4.5，図 4.6，図 4.7，図 4.8 の (a) より，2 番目の単語が強調された場合，$A_d(a)$ は 0 未満であることから，2 番目の単語のアクセント核の F0 の値が 1 番目の単語のアクセント核より大きいことが示された．図 4.6，図 4.7，図 4.8 の (b) より，2 番目の単語が強調された場合，$A_d(b)$ の値は 0.5 以上と他の単語が強調された場合と比較して大きくなっていることから，3 番目の単語のアクセント核との差がより増加していることが示された．また，3 番目の単語が強調された場合，$A_d(b)$ は 0 未満であることから，3 番目の単語のアクセント核の F0 の値が 2 番目の単語のアクセント核より大きいことが示された．図 4.6，図 4.7，図 4.8 の（c）より、3 番目の単語が強調された場合，$A_d(c)$ の値が 0 であることから，1 番目と 3 番目の単語のアクセント核の差がなくなっていることが示された．一方で，図 4.5(b) より，3 番目の単語が強調された場合の $A_d(b)$ が 0 以下という結果はアクセント核のある単語と同様の結果であったが，2 番目の単語が強調された場合の結果は異なり，前の単語が強調された場合と同じ値をとっていた．また，図 4.5(c) より 3 番目の単語が強調された場合の $A_d(c)$ が 0 以下であることから，アクセント核のある単語とは異なり，1 番目と 3 番目の単語のアクセント核の差が存在することが示された．

以上の結果から、強調単語がアクセント核を持つ場合，強調された単語と次の単語との間のアクセント核の点ピッチの差はより大きくなり，強調された単語と前の単語とのアクセント核の点ピッチの差はより小さくなることが示された．しかし、強調の程度が変わってもアクセント核の点ピッチの差の減少量は変化がなかった．強調単語がアクセント核を持たない場合，2 番目の単語が強調された場合はアクセント核を持つ単語と同様の結果が示されたが，3 番目の単語が強調された場合はダウンステップが観測されたが F0 ピークの減少量は小さくなっていた．そのため、各アクセント核の F0 の差は強調の有無と関連があると考えられる．

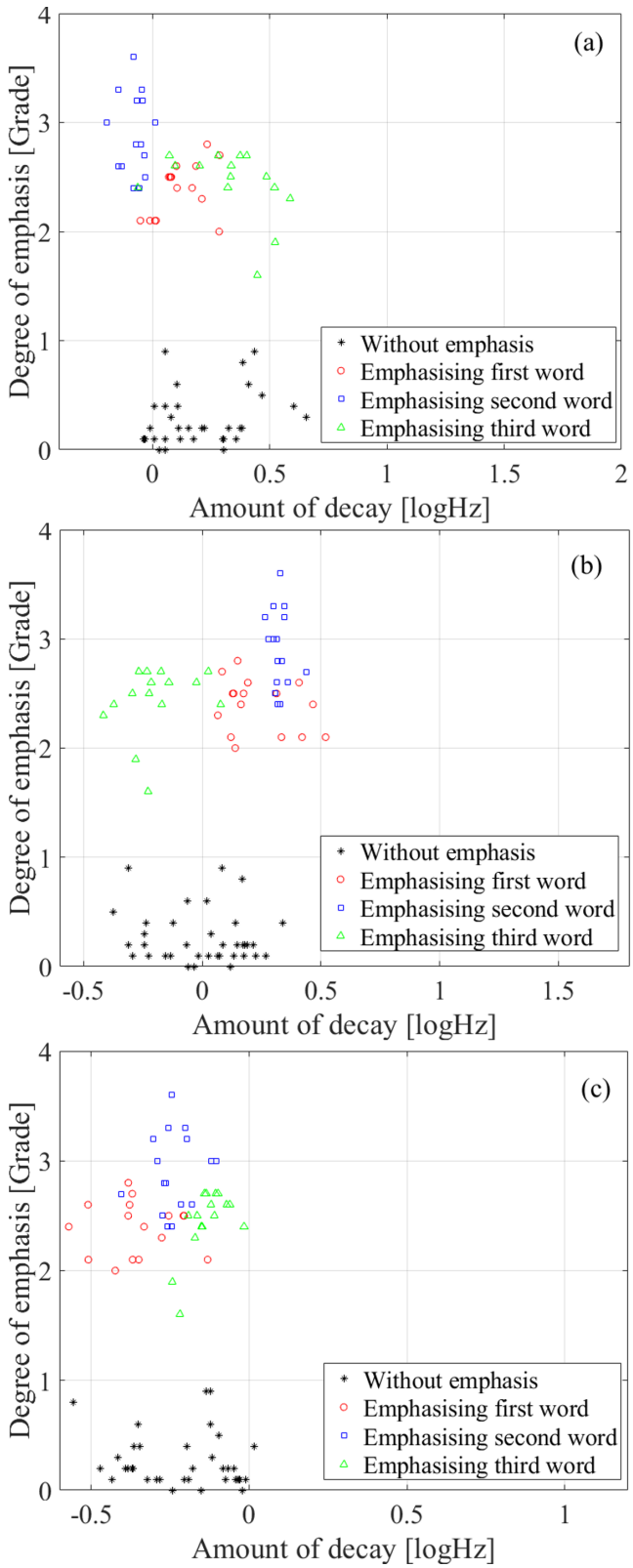

図 4.5: 強調単語のアクセント型が 0 型の場合の強調の程度と各アクセント核の F0 の差の関係，(a)1 番目の単語と 2 番目の単語のアクセント核の F0 の差，(b)2 番目の単語と 3 番目の単語のアクセント核の F0 の差，(c)1 番目の単語と 3 番目の単語のアクセント核の F0 の差

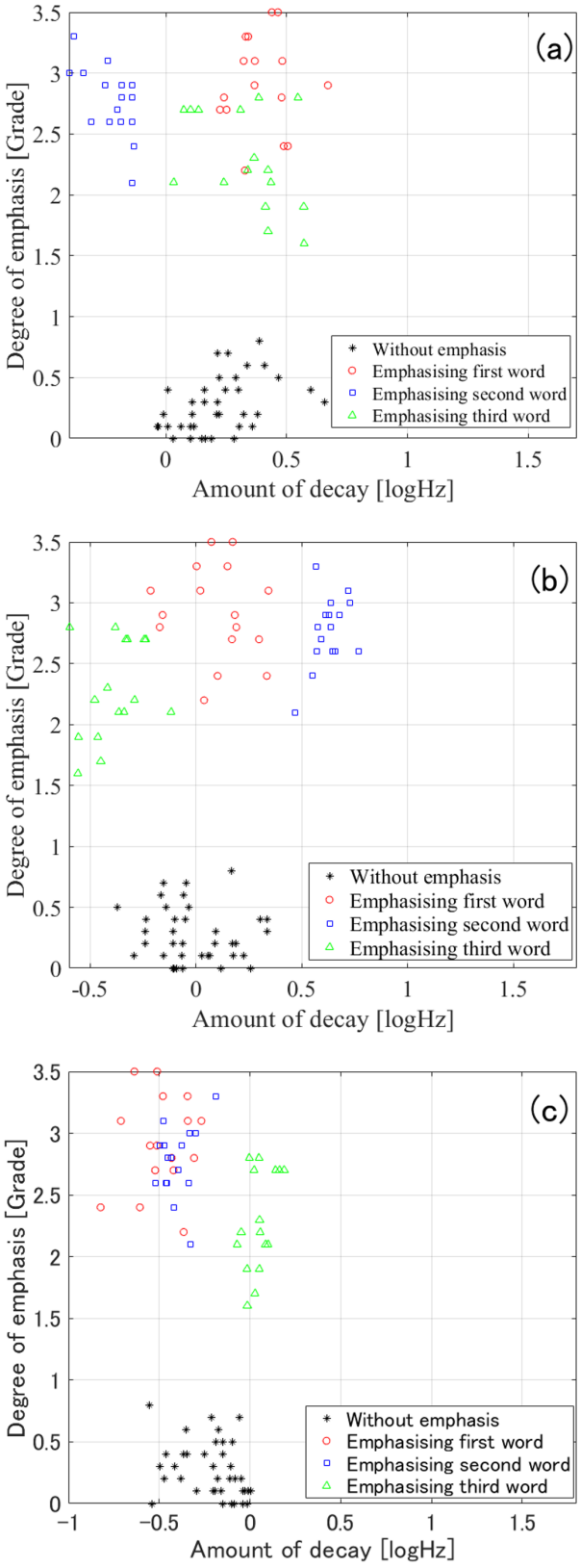

図 4.6: 強調単語のアクセント型が 1 型の場合の強調の程度と各アクセント核の F0 の差の関係，(a)1 番目の単語と 2 番目の単語のアクセント核の F0 の差，(b)2 番目の単語と 3 番目の単語のアクセント核の F0 の差，(c)1 番目の単語と 3 番目の単語のアクセント核の F0 の差

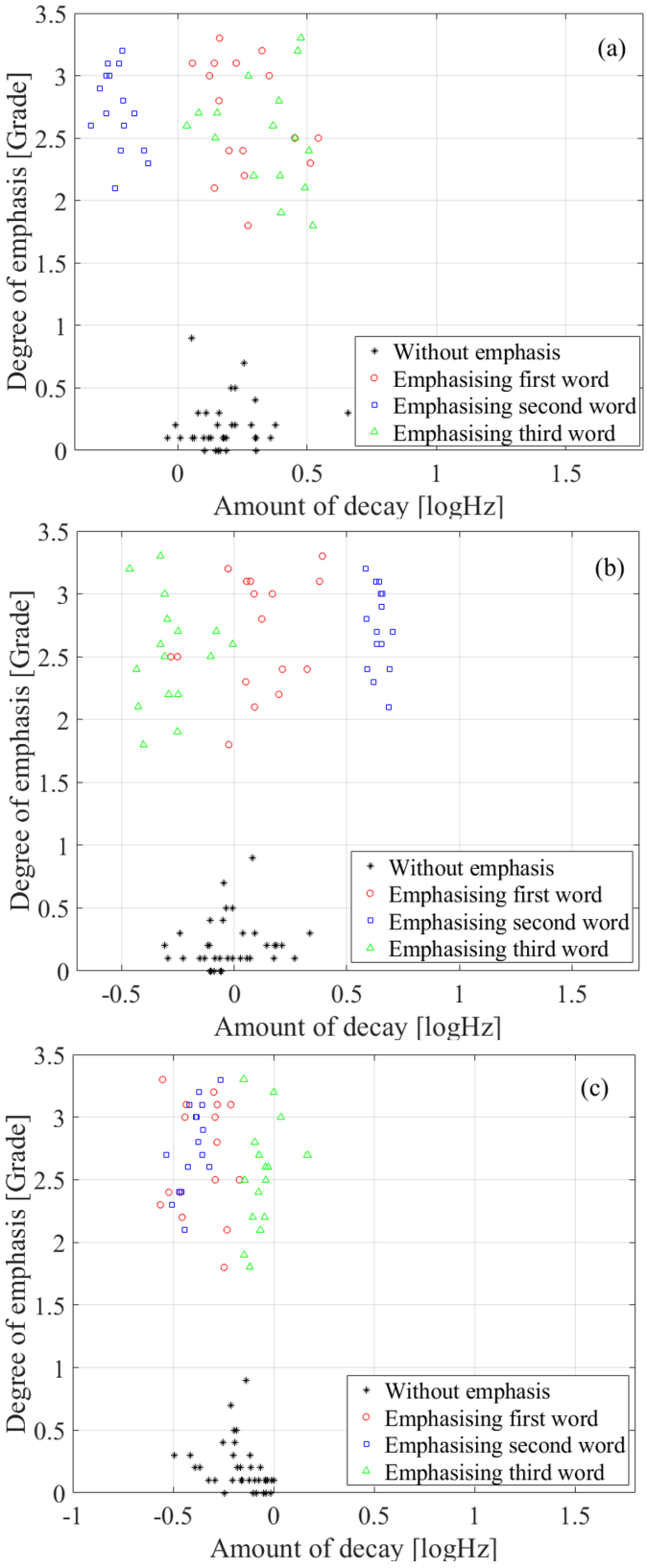

図 4.7: 強調単語のアクセント型が 2 型の場合の強調の程度と各アクセント核の F0 の差の関係，(a)1 番目の単語と 2 番目の単語のアクセント核の F0 の差，(b)2 番目の単語と 3 番目の単語のアクセント核の F0 の差，(c)1 番目の単語と 3 番目の単語のアクセント核の F0 の差

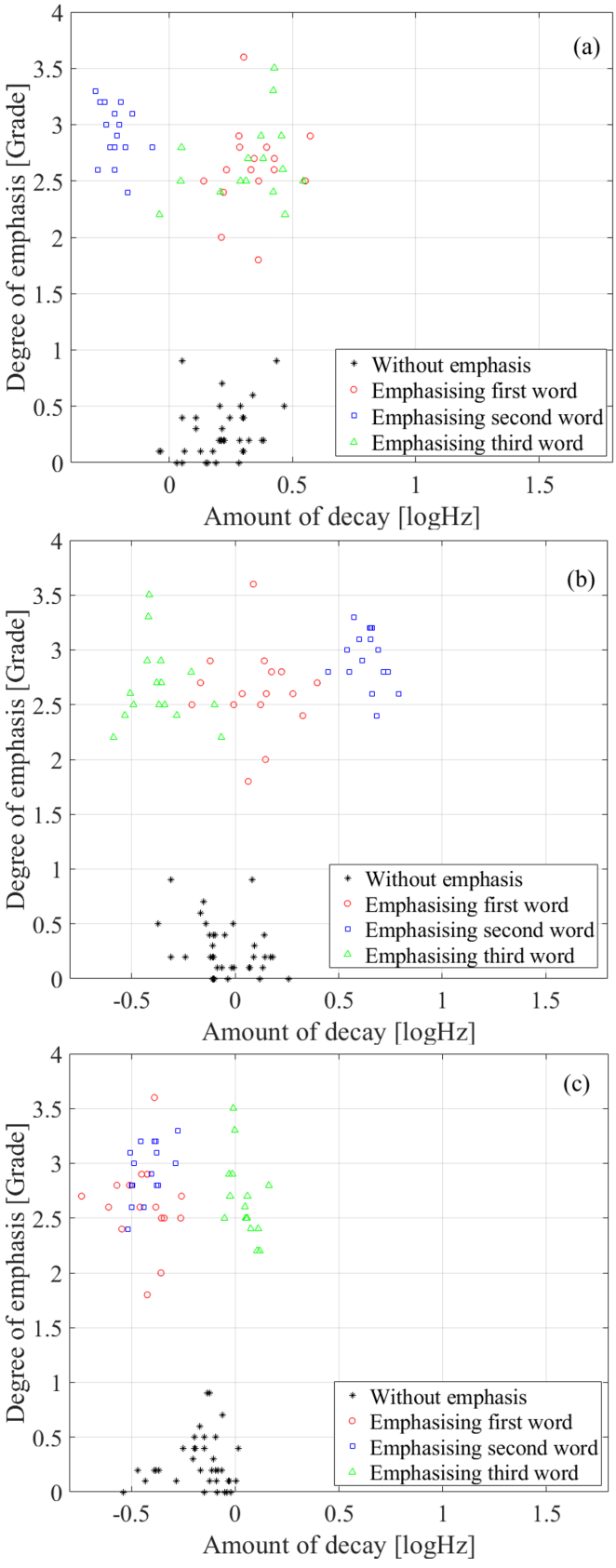

図 4.8: 強調単語のアクセント型が 3 型の場合の強調の程度と各アクセント核の F0 の差の関係，(a)1 番目の単語と 2 番目の単語のアクセント核の F0 の差，(b)2 番目の単語と 3 番目の単語のアクセント核の F0 の差，(c)1 番目の単語と 3 番目の単語のアクセント核の F0 の差

4.4　実験 2 分析結果の評価

上述の結果に従って F0 軌跡を修正し，合成した音声から人が強調を知覚できるか評価するために実験 2 を行った．

4.4.1　実験刺激

実験には，同様の発話内容で三種類の音声 (実験 1 で強調なしと評価された音声を再合成した音声，実験 1 で強調なしと評価された音声の F0 を点ピッチから推定される F0 に置き換えて再合成した音声，分析結果に従って操作を行った点ピッチから推定される F0 に置き換えて再合成した音声) を用いた．音声の合成には STRAIGHT を用いた．合成時に使用する点ピッチから推定される F0 は，有声区間の始終点を仮点ピッチとし，仮点ピッチと点ピッチの間を直線で内挿することで求められた．また，無声区間の F0 の値は 0 とした．点ピッチから推定される F0 を用いた音声合成の研究は多く行われており，自然な合成音が得られることが報告されている [21] [27] [28] [29]．合成音声の分類は，操作を行ったアクセント句が文頭・文中・文末の 3 種類で，アクセント句に含まれる名詞単語のアクセント型が 0 型〜3 型の 4 種類で 12 種類の組み合わせであった．

図 4.9 に合成に使用した F0 軌跡の図を示す．青線は音声の F0 軌跡を，赤線は点ピッチから推定される F0 軌跡を，黒線は修正した点ピッチから推定される F0 軌跡を示す．緑星印は点ピッチを赤丸は操作した点ピッチを示す．音声の発話内容は「兄さんから仲人に弁当を渡してほしい」であった．

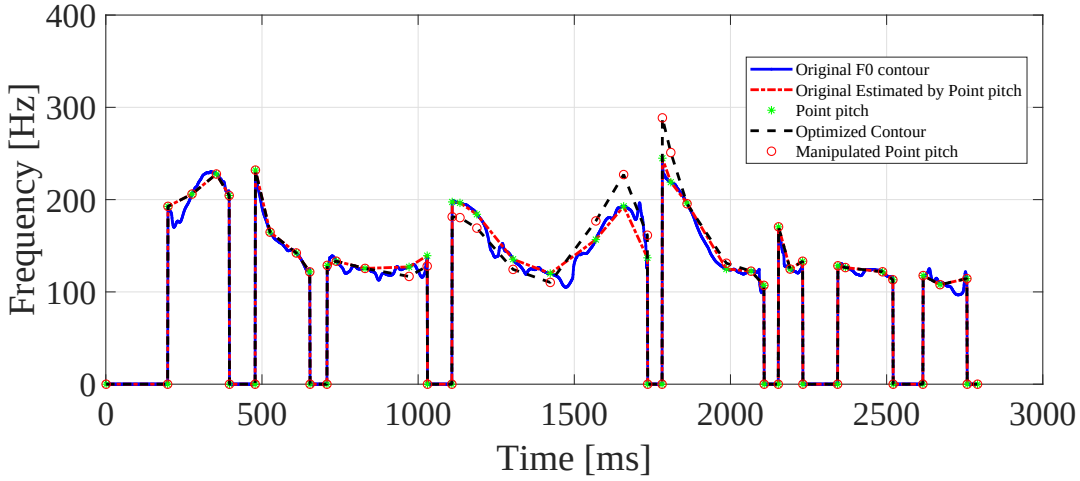

図 4.9: 合成に使用した F0 軌跡

4.4.2　実験参加者

北陸先端科学技術大学院大学の 20 代の成人男女 10 名（男性 6 名女性 4 名）が実験に参加した．すべての被験者は正常聴力を有していた．

4.4.3　実験方法と実験装置

上述の実験刺激 36 音声を被験者にランダムで提示した．これらの音声刺激の提示制御は PC(Windows 8，MATLAB 2015a) で行い，オーディオインターフェース (Fireface UCX) で DA 変換したのちアンプ (STAX SRM-1/MK-2pp) で信号増幅し，ヘッドホン (STAX SR-404) で提示した．各刺激は 1 度だけ被験者に提示した．実験は防音室で行った．実験は実験 1 と同様の GUI を用いて，各刺激に対して被験者に強調の有無を回答するよう求めた．加えて，強調があると回答した場合には強調の程度を 4 段階 (1〜4) で回答するよう求めた．実験は 30 分程度で行った．

4.4.4　実験結果

表 4.1 から表 4.4 に実験結果を示す．表 4.1 から表 4.4 はそれぞれ 0 型から 3 型のアクセント型が強調された場合を表している．表は，発話内容が同じで，STRAIGHT 分析再合成された音声，点ピッチから推定された F0 情報を用いて再合成された音声，分析結果に従って操作された点ピッチから推定された F0 情報を用いて再合成された音声の強調の有無の回答率を示す．

表 4.2 から表 4.4 より，アクセント核が存在する単語に対して，アクセント核からの F0 の減少量やアクセント核ごとの F0 の差の結果に従って点ピッチを操作した音声から強調があるとの回答率が高く，操作を行った音声から強調が知覚されることが示された．しかし，表 4.1 より，アクセント核を持たない単語について，強調ありの回答率は 40%から 60%であった．このことから，アクセント核を持つ単語が対比強調された場合についてアクセント核からの F0 の減少量は重要であることが考えられる．

表 4.1: 0 型アクセント単語が強調された場合の音声の強調の有無の回答

	強調操作を行った箇所		
	第 1 単語	第 2 単語	第 3 単語
STRAIGHT を用いて分析再合成された音声	30%	0%	10%
点ピッチから推定された F0 情報を用いて再合成された音声	20%	0%	0%
結果に基づき操作された点ピッチから推定された F0 情報を用いて再合成された音声	60%	50%	40%

表 4.2: 1 型アクセント単語が強調された場合の音声の強調の有無の回答

	強調操作を行った箇所		
	第 1 単語	第 2 単語	第 3 単語
STRAIGHT を用いて分析再合成された音声	20%	10%	60%
点ピッチから推定された F0 情報を用いて再合成された音声	30%	0%	40%
結果に基づき操作された点ピッチから推定された F0 情報を用いて再合成された音声	90%	70%	100%

表 4.3: 2 型アクセント単語が強調された場合の音声の強調の有無の回答

	強調操作を行った箇所		
	第 1 単語	第 2 単語	第 3 単語
STRAIGHT を用いて分析再合成された音声	10%	10%	50%
点ピッチから推定された F0 情報を用いて再合成された音声	0%	0%	30%
結果に基づき操作された点ピッチから推定された F0 情報を用いて再合成された音声	90%	90%	90%

表 4.4: 3 型アクセント単語が強調された場合の音声の強調の有無の回答

	強調操作を行った箇所		
	第 1 単語	第 2 単語	第 3 単語
STRAIGHT を用いて分析再合成された音声	20%	0%	20%
点ピッチから推定された F0 情報を用いて再合成された音声	40%	0%	10%
結果に基づき操作された点ピッチから推定された F0 情報を用いて再合成された音声	90%	90%	90%

4.5　まとめ

強調の程度と本研究の着眼点（アクセント核の前後の F0 の増加量・減少量，ダウンステップの影響による F0 のピーク値の低下量）との関係を明らかにするために，各音声の着眼点の F0 の変動量と強調の程度の関係を求めて比較した．

結果，対比強調音声の強調単語のアクセント核からの F0 の減少量は平静音声に比べて増加していることが示された．また，F0 の減少量が異なると強調の程度が異なる傾向が見られた．アクセント核までの F0 の増加量の傾向はアクセント型によって異なっており，対比強調音声のアクセント核までの F0 の増加量は平静音声に比べ，2 型の場合では増加し，文中に位置する 3 型の場合は減少していることが示された．

強調単語と強調単語に後続する単語とのアクセント核の差は，平静音声に比べ増加し，強調単語と強調単語の前の単語との差は減少していることが示された．しかし，強調の程度が変化しても各アクセント核の差に変化は見られなかった．

比較結果に従って点ピッチを操作し，推定される F0 軌跡を用いた再合成音声を用いた聴取実験の結果，再合成音声から対比強調が知覚された．従って，本研究の着眼点（アクセント核の前後の F0 の増加量・減少量，ダウンステップの影響による F0 のピーク値の低下量）は対比強調の有無に関連する特徴であることがわかった．しかし，強調の程度が異なってもダウンステップの影響による F0 のピーク値の低下量に違いは見られなかった．

第5章　結論

5.1　本研究で明らかになったこと

本研究では，強調の程度に関連している音響特徴やその変化を明らかにするために，対比強調の音声学的実現に重要であるピッチ（声の高さ）に関わる日本語の特徴の F0 の変動量の違いが，任意の強調の程度の実現に有効であると仮説を立てた．本研究の目的は，仮説を評価するために，音声ごとの強調の程度と F0 の変動量の違いの関係を比較・評価することである．アクセント核からの F0 の下降量は，対比強調された単語がアクセント核を持つ単語（有核語）の場合，対比強調音声は平静音声に比べ F0 の下降量は増加することが示された．また，F0 の下降量が異なると，強調の程度に違いが見られた．強調単語と後続する単語の間のアクセント核の F0 の差は平静音声や他の箇所が強調されてる場合に比べ大きくなり，強調単語と前の単語とのアクセント核の F0 の差はより小さくなることが示された．分析結果に従って F0 軌跡を修正し，対比強調のない音声に対比強調の付加した音声の再合成を行った．作成した合成音声を用いた主観評価実験によって分析結果を評価した結果，アクセント核のある単語（有核語）について F0 軌跡の修正を行った音声から対比強調が知覚された．よって有核語を対比強調する場合，本研究で着目した F0 の変動量や F0 の差は強調の有無に関連がある特徴であることが分かった．

本研究では分析結果の評価のために，点ピッチを用いて再合成する音声の F0 軌跡の推定を行った．実験結果から操作を行った箇所から対比強調が知覚されたため，合成音声の任意の箇所に対比強調を付加する音声再合成への活用が期待される．また，再合成には点ピッチを用いて再合成する音声の F0 軌跡の推定値を用いるため，少ない計算量で再合成が行える．

5.2　今後の課題

5.2.1　強調の程度との関連

アクセント核からの下降量が異なった場合，強調の程度に違いが見られたため，強調の程度とも関連があると推察する．着目した F0 軌跡の特徴と強調の程度と関連を明らかにするには，分析した結果の増減量や低下量の違いをモデル化し，モデルに従って再合成した音声を用いてモデルを評価する必要がある．強調の強弱のモデル化の際には，アクセン

ト型や強調の有無だけでなく強調の程度についても統一のとれたデータを用いてさらに分析を進める必要がある [6].

5.2.2 使用データの個人性

本研究では，使用した音声の発話者が一人であったため，分析結果の使用データの個人性による影響を考えなければならない．そのため，分析した結果の増減量や低下量の違いのモデル化には，音声データを増やし，さらに分析を進める必要がある．

5.2.3 アクセント核が存在しない単語

アクセント核を持たない単語（無核語）の強調に関しては強調あるとの回答率が低かった．今回 F0 の操作をアクセント核のある単語と同様にするとアクセント型が崩れてしまう恐れがあった．日本語の場合，音の高低が単語の意味が決まり，同じ発話内容でもアクセント型が変わると無意味語となる報告がある [30]．そのため，アクセント型を考慮しつつ，無核語について分析を進める必要がある．また，ダウンステップの阻止のために無核語の F0 を増加させると，音声全体でのダウンステップがなくなる恐れがある．音声全体のダウンステップがなくなると，すべてのアクセント句が対比強調されているように聞こえ，きわめて不自然に聞こえる [20]．そのため，無核語の対比強調音声の特徴を抽出するためには，F0 以外の強調に関連した他の物理量の検討が効果的であると考えられる．

5.2.4 強調に関連するその他の物理量の検討

実験 1，2 の内観報告では声の高さ以外に声の大きさや声の高さの変化の急峻さを手がかりに強調を評価しているという報告があった．先行研究でも，4 つの手法（声の大きさ，話速，間，声の高さ）に関連した物理量の操作が行われている [10] [12] [13] [14]．そのため，ほかの物理量についても強調の程度との関連を検討する必要がある．

また，本研究では点ピッチを用いて着目した F0 の変化を分析した．点ピッチを用いた分析は計算量が少なく，モーラのピッチと基本周波数との対応がとりやすいため，着眼点の F0 の変化量について少ない計算量で分析することができた．しかし，時間情報が壊れてしまうため，話速を考慮した場合のアクセント核に着目した F0 の分析が行えない．従って，音高変化の急峻さなどのように話速を考慮した場合のアクセント核に着目した F0 の分析の際には別の分析方法の検討が必要となる．

参考文献

[1] 藤崎 博也, “韻律研究の諸側面とその課題,,” 日本音響学会講演論文集, pp. 287–288, 1994.

[2] 能勢 隆, “統計モデルに基づく多様な音声の合成技術,” 電子情報通信学会論文誌, Vol. J100–D, No. 4, pp. 556–569, 2017.

[3] 中嶋 秀治, 水野 秀之, 阪内 澄宇, “テキストからの表現豊かな音声合成を目指した宣伝文からの強調アクセント句の予測,” 情報処理学会論文誌, Vol. 56, No. 12, pp. 2384–2394, 2015.

[4] 田中 晴美ほか (編) “現代言語学辞典,” 成美堂, 東京, 1988.

[5] Kumar, V., Sridhar, R., Nenkova, A., Narayanan, S. and Jurafsky, D. “Detecting prominence in conversational speech: pitch accent, givenness and focus,” Proc. Speech Prosody, pp.453–456, 2008.

[6] 前川 喜久雄, 北川 , “音声はパラ言語をいかにつたえるか,” Cognitive studies, Vol. 9, No. 1, pp. 46–66, 2002

[7] 福岡 昌子, “パラ言語的情報の強調順位に関する日本語学習者の知覚,” 音声研究, Vol. 21, No. 3, pp. 1–14, 2017.

[8] Ladd, R., K. Silverman, F. Tolkmitt, G. Bergmann and R. Scherer “Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect.,”Journal of the Acoustical Society of America Vol. 78, No. 2, pp. 435–444, 1985.

[9] 郡 史郎, “フォーカス実現における音声の持続時間, 強さ, F0 の役割,” 音声言語 III, pp. 29–38, 1989.

[10] 武田 昌一, 市川 熹, “日本語文音声におけるプロミネンスの韻律的特徴の解析,” 日本音響学会論文誌, Vol. 47, No. 6, pp. 386–396, 1991.

[11] 白井 克彦, 岩田 和彦, “音声合成のための単語の強調表現の規則化,” 電子情報通信学会論文誌, Vol. J70–A, pp. 816–821, 1987.

[12] 武田 昌一, 市川 熹, “日本語文音声のプロミネンス生成規則の作成と評価,” 日本音響学会論文誌, Vol. 47, No. 6, pp. 397–404, 1991.

[13] 浜田 洋, 千葉 仁一, “音声合成における音声強調インターフェースの設計法,” 情報処理学会 HCI 研究報告, Vol. 41, No. 10, pp. 69–76, 1992.

[14] 浜田 洋, 千葉 仁一, “GUI による音声コントローラと音声強調インタフェースへの応用,” 情報処理学会論文誌, Vol. 34, No. 12, pp. 2569–2577, 1993.

[15] 益子 貴史, 徳田 恵一, 小林 隆夫, 今井 聖, “動的特徴を用いた HMM に基づく音声合成,” 電子情報通信学会論文誌, Vol. J79–D, No. 12, pp. 2184–2190, 1996.

[16] 吉村 貴克, 徳田 恵一, 益子 貴史, 小林 隆夫, 北村 正, “HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化,” 電子情報通信学会論文誌, Vol. J83–D–2, No. 11, pp. 2099–2107, 2000.

[17] 森實 久美子, 中村 圭吾, 戸田 智基, 猿渡 洋, 鹿野 清宏, “HMM に基づく音声合成における強調音声の生成,” 情報処理学会 SLP 研究報告, Vol. 75, No. 6, pp.27–32, 2009.

[18] 日本音響学会（編）, “音響用語辞典,” コロナ社, 東京, 2003.

[19] 前川 喜久雄, “イントネーション研究の現状と課題,” 日本語学, Vol. 33 No. 7, pp. 4–14, 2014.

[20] 田窪 行則, 前川 喜久雄, 窪園 晴夫, 本田 清志, 白井 克彦, 中川 聖一, “言語の科学 音声,” 岩波書店, 東京, 2002.

[21] 橋本 新一朗, “日本語アクセントの諸性質, ” 電子情報通信学会論文誌, Vol. 56–D, No. 11, pp.654–661, 1973.

[22] 板橋 秀一（編著）, “音声工学,” 森北出版, pp. 160–161, 東京, 2005.

[23] Kawahara, H., Masuda-Katsuse, I., De Cheveigne, A. “Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” Journal of Speech Communication, Vol. 27, No. 3, pp. 187–207, 1999.

[24] 佐藤 大和, “共通語アクセントの成因分析,” 日本音響学会論文誌, Vol. 49, No.11, pp.775–784, 1993.

[25] 武田 一哉, 匂坂 芳典, 片桐 滋, 桑原 尚夫, “音声データベース構築のための視察に基づく音韻ラベリング,” ATR 日本語音声データベース付属資料, 1988.

[26] 佐藤 大和, “PARCOR-VCV 連鎖を用いた音声合成方式,” 電子情報通信学会論文誌, Vol. J61–D, No. 11, pp. 858–865, 1978.

[27] 阿部 匡伸, 佐藤 大和,“音節区分化モデルに基づく基本周波数の 2 階層制御方式,” 日本音響学会論文誌, Vol. 49, No. 10, pp. 682–690, 1993.

[28] 山田 真裕, 岩野 公司, 古井 貞熙, “数量化 I 類による F0 パターン生成の制御要因に関する検討,” 情報処理学会 SLP 研究報告, Vol. 38, No. 3, pp.15–20, 2001.

[29] 箱田 和雄, 佐藤 大和, “文音声における音調規則,” 電子情報通信学会論文誌, Vol. J63–D, No. 9, pp. 715–722, 1980.

[30] 小野 明日香, “日本語ピッチアクセント音声データベースの構築,” 日本音響学会聴覚研究会資料, Vol. 48, No. 8, pp. 817–822, 2018.

研究業績一覧

国際会議

1. OHTANI, Y., AKAGI, M., “Study on Relationship between Degree of Emphasis and Acoustic Feature for Synthesizing Emphasized Speech,” 2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’19), Hawaii, USA, March, 2019.

謝辞

本研究を進めるにあたり，多大なるご指導ならびにご鞭撻を賜りました赤木 正人 教授に心から御礼申し上げます．

研究室会議やミーティングなどの場において，数多くの助言を頂いた鵜木 祐史 教授に心より感謝致します．

日頃から御助言ならびに御助力をいただきました博士後期課程３年 鳥谷 輝樹 氏, 博士後期課程１年 高橋 響子 氏，磯山 拓都 氏に感謝致します．

共に切磋琢磨した，音情報処理分野　赤木・鵜木研究室博士前期課程２年の皆様に感謝致します．

研究で行き詰まったときに相談に乗って頂いた，音情報処理分野　赤木・鵜木研究室の皆様に感謝致します．

最後に本学での研究生活を支え，温かく見守ってくれた両親に心から感謝致します．