

言語情報のみを用いた音節クラスタリングの音節波形接続型音声合成法への適用

植村 和久[†] 村上 仁一[†] 池原 悟[†]

[†] 〒 680-8552 鳥取県鳥取市湖山町南 4-101 鳥取大学工学部知能情報工学科
E-mail: †{s032007,murakami,ikehara}@ike.tottori-u.ac.jp

あらまし 音声合成法の手法の一つとして、音節波形接続型音声合成法が提案されている。この手法の問題点の一つとして、任意の一般名詞を作成する際に大量の録音単語が必要となる。そこで従来の研究では、理論上全ての音声を作成するために、音声が収録されている DB に対して音響情報と言語情報の両方を用いてクラスタリングを行った。しかし、音声品質が非常に悪い音声もあった。この原因として、本論文では音響的特徴の類似性を用いて合成音声を作成することが、音声品質に悪い影響を与えると仮定した。そこで、言語情報のみを用いてクラスタリングを行い、音声品質の改善を行った。音声品質の評価には、オピニオン評価実験および対比較実験を用いた。オピニオン評価実験の結果、“言語クラスタ”の音声が 3.22、“音響クラスタ”の音声が 2.48 という値が得られた。また、対比較実験の結果、“言語クラスタ”の音声が 77.8%、“音響クラスタ”の音声が 22.2%となった。以上より、言語情報のみを用いたクラスタリング手法は、言語情報と音響情報の両方を用いたクラスタリング手法と比べ、良い音声品質の合成音声を作成できることが分かった。

キーワード クラスタリング, 音節波形接続型音声合成, 音声合成, 言語情報

Syllable Clustering using Only Linguistic Information for “Word Synthesis by Concatenating Syllabic Components”

Kazuhisa UEMURA[†], Jin'ichi MURAKAMI[†], and Satoru IKEHARA[†]

[†] Faculty of Engineering, Tottori University, Minami 4-101, Koyama-cho, Tottori-shi, Tottori-ken, 680-8552
Japan

E-mail: †{s032007,murakami,ikehara}@ike.tottori-u.ac.jp

Abstract “Word synthesis by concatenating syllabic components” is proposed as a speech synthesis method. As a problem of this technique, large amount of recording words is needed when we make arbitrary common noun. In the past research, syllable clustering using acoustic information and linguistic information can make arbitrary common noun. However, very low quality synthesized speech is sometimes generated. As this reason, I assumed that making the synthesized speech by using the similarity of an acoustic feature might have a bad influence for the speech quality. Then in this paper, our aim is to improvement of the speech quality by doing syllable clustering using only linguistic information. The mean opinion score (MOS) and the ABX test are used for the evaluation of the speech quality. As the result of the MOS, the “Linguistic Cluster” was obtained 3.22 and the “Acoustic Cluster” was obtained 2.48. Moreover, as the result of the ABX test, the “Linguistic Cluster” was obtained 77.8% and the “Acoustic Cluster” was obtained 22.2%. Therefore, syllable clustering using only linguistic information can make good speech quality compared with syllable clustering using acoustic information and linguistic information.

Key words tree-based clustering, concatenating syllabic components, word synthesis, linguistic information

1. はじめに

音声合成法の手法の一つとして、音節波形接続型音声合成

法 [1] が提案されている。この手法は、録音した音声波形の一部（以下、音節素片）を取り出し、接続することによって合成音声を作成する。音声波形に信号処理を加えないため、自然性

の高い音声を作成出来るが、音節素片選択時に、全ての言語情報を一致させなければならない。そのため、任意の一般名詞を作成する際に大量の録音単語が必要となる。

その問題を解決するために、従来研究において Tree-based Clustering (以下、クラスタリング) 手法が提案されている [2], [3]。文献 [3] では、MFCC の距離尺度やモーラ数、前後音素環境等の情報を用いている。この手法は、音節素片選択時の全ての言語情報を完全に一致させるのではなく、一部の言語情報をクラスタリングを用いて緩和する。緩和された言語情報を用いることで、理論上全ての合成音声を作成可能となる。

しかし、音声品質の非常に悪い音声ができる場合がある。この原因として、音声合成において、MFCC の距離尺度は、音声品質との相関が低いためであると考えている。そこで本研究では、MFCC の距離尺度を全く使用せず言語情報のみを用いてクラスタリングを行うことで、良い音声品質が得られると仮定した。

以上より本研究では、言語情報のみを用いてクラスタリングを行い、作成した合成音声の音声品質の調査を行う。

2. 音節波形接続型音声合成法

2.1 音節波形接続型音声合成の概要

音節波形接続型音声合成では、以下の手順に従い合成音声を作成する。

(1) データベース中の各音節素片に対して、単語のモーラ数、モーラ位置、アクセント型、各モーラ位置のアクセントの高低、前後の音素環境のラベルを付与する。付与された音節素片の例を表 1 に示す。

表 1 音節素片の詳細 (例: o-ri0402001+m)

前音素環境	o
中心音節	ri
モーラ数	04
モーラ位置	02
アクセント型	00
アクセントの高低	1 (高)
後音素環境	m

(2) 合成する単語内に含まれる音節素片と、言語情報の一致する音節素片をデータベース中から選択する。音節波形接続型音声合成で用いる言語情報を表 2 に示す。

表 2 音節波形接続型音声合成における言語情報

中心の音節	単語のモーラ位置
直前の音素 (前音素環境)	単語のアクセント型
直後の音素 (後音素環境)	単語のアクセントの高低
単語のモーラ数	

(3) 手順 2 で選択した音節素片を接続し、合成音声を作成する。「乗り物 (no/ri/mo/no)」を作成する際に用いる音節素片の例を図 1 に示す。

図 1 において「 」はアクセントの高低を表す。また、太文字で示している部分は、合成に使用する音節素片である。

乗り物 (<u>no</u> /ri/mo/ <u>no</u>) = 乗換 (<u>no</u> /ri/ka/e/)
+ 織物 (<u>o</u> /ri/mo/ <u>no</u>)
+ 履き物 (<u>ha</u> /ki/mo/ <u>no</u>)
+ 入れ物 (<u>i</u> /re/mo/ <u>no</u>)

図 1 音節波形接続型音声合成法の例

2.2 合成可能な音声数の問題

従来研究 [5] より、音節波形接続型音声合成法の有用性が示された。しかしこの手法では、音節素片選択時に言語情報が全て一致する音節素片が必要であるため、作成できる音声が少ない。例えば、ATR 単語発話データベース Aset (5,240 単語) を使用して同一発話内容の音声を作成する場合、5,240 単語中の 470 単語しか作成できない [5]。

3. 言語情報のみを用いたクラスタリング

3.1 概要

言語情報のみを用いたクラスタリングは、予備実験による言語的な知識を用いて、データベースに存在しない音節素片を、最も言語的特徴の類似したデータベース中の音節素片と同じクラスターへ分類する。

予備実験では、アクセント、継続時間、調音の 3 種類の情報に対して、合成音声の音声品質に与える影響の調査を行った。予備実験の結果、アクセント、調音、継続時間の順に音声品質に与える影響が大きいと分かった。また音節波形接続型音声合成法では、アクセント情報として、単語のアクセント型とアクセントの高低、継続時間情報として、単語のモーラ数とモーラ位置、調音情報として、前後音素環境を用いる。

そのため、本研究では、モーラ数、モーラ位置、後音素環境、前音素環境、アクセント型、アクセントの高低の順に言語情報の緩和を行う。言語情報の緩和は、優先順位を付与したクラスタリングのルールを作成することでを行い、上位のルールから順に用いる。本研究では、全部で 17 個のルールを作成した。作成したルールを付録の表 A.1 に示す。

ルールを用いてクラスタリングを行うと、複数のクラスタが作成できる。各クラスタには言語的特徴の類似した音節素片が含まれる。本研究では、クラスタ内の音節素片を同一とすることで、音節波形接続型音声合成法の条件の緩和とみなす。表 3 にクラスタの例を示す。

表 3 クラスタの例

クラスタ名	N_5	pe_32
クラスタ内の音節素片	a-N0202001+pau	N-pe0403001+k
	a-N0303001+pau	i-pe0302011+N
	o-N0303001+pau	o-pe0302000+e
	u-N0202001+pau	q-pe0403001+i
	u-N0303001+pau	

3.2 クラスタリングを利用した音節波形接続型音声合成法
本研究では、データベースに存在しない音節素片を、データベース中の音節素片を用いて、任意の一般名詞を作成する。本

研究における音声合成の流れを図 2 に示す。また、本研究で作成した合成音声の例を図 3 に示す。

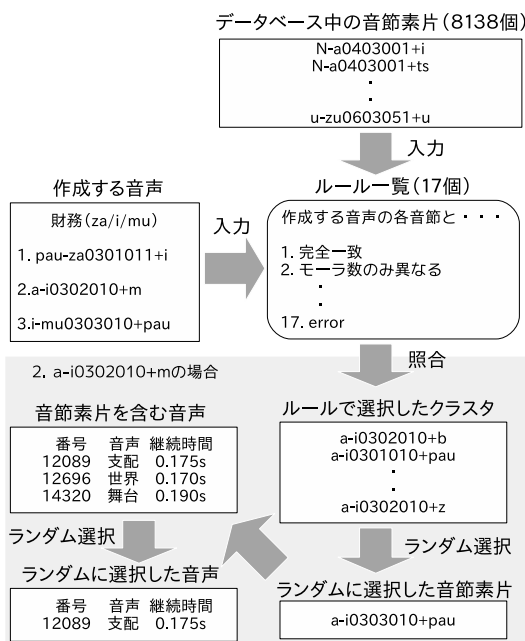


図 2 クラスタリングを利用した波形接続型音声合成の流れ図

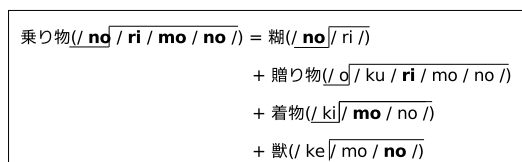


図 3 クラスタリングを利用した音節波形接続型音声合成法の例

本研究における、クラスタリングを利用した波形接続型音声合成の手順を以下に示す。

- (1) データベース中の音節素片 8,138 個と、作成する音声の各音節素片を入力データとして準備する (図 2 では「2.a-i0302010+m」が入力された場合を想定する)。
- (2) 手順 1 の 2 種類の音節素片を付録の表 A.1 に示す 17 個のルールを用いて順に照合する。
- (3) 上位のルールから順に照合し、適合した音節素片が 1 つ以上あればそのルールで打ち切る。
- (4) 再下位のルールである 17 までに適合する音節素片が見付からない場合、エラーを出力する。
- (5) ルールで選択したクラスタ中の音節素片を、ランダムに 1 つ選択する。
- (6) 手順 5 で選択した音節素片を含む音声を、ランダムに 1 つ選択する。

図 2 では、「財務 (/za/i/mu/)」の「a-i0302010+m」を作成する場合に、「支配 (/shi/ha/i/)」の「a-i0303010+pau」を使用する。

図 3 は「乗り物 (no/ri/mo/no)」を作成する際に用いる音節素片の例である。2.1 節の図 1 と比較すると、図 3 で合成に使

用した音節素片は、全てモーラ数およびモーラ位置の異なる音節素片であることが分かる。

4. 言語情報のみを用いたクラスタリングの改良実験

4.1 実験条件

本章の実験では、女性話者 1 名に関して、言語情報のみを用いたクラスタリングで作成した音声を 20 単語作成する。音声の評価者は、音声研究に関わった経験のある 1 名とし、評価実験として、オピニオン評価実験を行う。

4.2 初期実装

本節の実験では、言語情報のみを用いたクラスタリングの初期実装 (3.2 節図 2) における音声品質を調査する。

評価実験の結果、オピニオンスコアとして 3.1 を得た。しかし作成した音声を聞くと、前後音素環境にポーズが含まれている場合、音声品質が劣化することがあると分かった。

4.3 ポーズの考慮

本節の実験では、ポーズの位置を考慮してクラスタリングを行った場合の音声品質を調査する。以下に実験手順を示す。

- (1) データベース中の音節素片 8,138 個をポーズ (“pau”) の位置に基づいて、3 つのグループに分類する。各グループは前音素環境に “pau” を含む (立ち上がり)、後音素環境に “pau” を含む (終わり)、それ以外 (中間) の 3 つのグループとする。
- (2) 手順 1 の 3 つのグループと、作成する音声の各音節素片を入力データとして準備する。

(3) 付録の表 A.1 に示す 17 個のルールを用いて順に照合する (作成する音声の音節素片によって、照合するグループを変更)。

(4) 上位のルールから順に照合し、適合した音節素片が 1 つ以上あればそのルールで打ち切る。

(5) 再下位のルールである 17 までに適合する音節素片が見付からない場合、エラーを出力する。

(6) ルールで選択したクラスタ中の音節素片を、ランダムに 1 つ選択する。

(7) 手順 6 で選択した音節素片を含む音声を、ランダムに 1 つ選択する。

上記手順 3 において、作成する音声の入力音節素片が前音素環境に “pau” を含む音節素片であり、かつ前音素環境以外の条件が一致するルールの場合は、立ち上がりのグループに対してのみ照合を行う。例えば、作成する音声の入力音節素片が「pau-za0301011+i」で、表 A.1 において、ルール 9 や 10 の場合、立ち上がりのグループに対してのみ照合を行う。同様に作成する音声の入力音節素片が後音素環境に “pau” を含む音節素片であり、かつ後音素環境以外の条件が一致するルールの場合は、終わりのグループに対してのみ照合を行う。例えば、作成する音声の入力音節素片が「i-mu0303010+pau」で、表 A.1 において、ルール 5 や 6 の場合、終わりのグループに対してのみ照合を行う。その他のルールや入力音節素片の場合は、全てのグループに対して照合を行う。

オピニオン評価実験の結果、ポーズを考慮したことで、初期

実装に比べオピニオンスコアが 0.3 向上した。

また本節の実験では、手順 7 において、複数の音節素片からランダムに 1 つ選択した。しかし各音声には継続時間のばらつきがあり、継続時間は音声品質に多大な影響を与える可能性がある。

4.4 継続時間の考慮

本節の実験では、4.3 節の手順 7 において、各音節の継続時間に着目し、最も継続時間の長い音節素片を持つ音声を選択した場合の音声品質を調査する。

オピニオン評価実験の結果、4.3 節のポーズを考慮した実験と同じオピニオンスコアとなった。しかし作成した音声を聞くと、促音 q を含む音声は、明らかに音声品質が改善されていることが分かった。また連続母音を含む音声は、音声品質が改善されている場合と劣化している場合の両方があった。

4.5 連続母音の考慮

4.4 節で継続時間を考慮した結果、連続母音を含む音声の音声品質が劣化することがあった。そこで本節の実験では、連続母音に対して、以下の対処を行う。

- 連続母音の数は 2 つまでとする。
- 連続母音の最初の音節を“音節 1”、2 番めの音節を“音節 2”とする。
- “音節 1”と“音節 2”は同じ音声番号の音声から選択する。
- “音節 1”の後音素環境が“音節 2”の中心音節と異なる場合、連続母音として考慮しない。
- 連続母音を考慮した結果、ポーズの位置が考慮されない場合、連続母音として考慮しない。

オピニオン評価実験の結果、連続母音を考慮した実験は、継続時間を考慮した実験と比べ、オピニオンスコアが 0.2 向上した。しかし、付録の表 A.1 に示すルール 11 と 14 で選択されたクラスタ内の音節素片は、他のルールで選択されたクラスタ内の音節素片と比べて数が多く、音声品質に悪影響を及ぼす可能性が高いと分かった。

4.6 全体の実験結果

本章で行った各実験のオピニオン評価実験結果を表 4 に示す。

表 4 オピニオン評価実験結果（総評価単語数：20）

音声の種類	オピニオンスコア
初期実装	3.1
ポーズの考慮	3.4
継続時間の考慮	3.4
連続母音の考慮	3.6

表 4 より、初期実装から連続母音の考慮の実験において、オピニオンスコアが確実に向上しているのが分かる。しかし各実験の評価者は、音声研究に関わった経験のある 1 名のみであるため、信頼性が低い。

5. 言語情報のみを用いたクラスタリングの音声品質調査実験

5.1 実験の目的

本章の実験では、言語情報のみを用いたクラスタリングの最

終的な音声品質を調査する。実験の流れ図を図 4 に示す。また実験の手順を以下に示す。

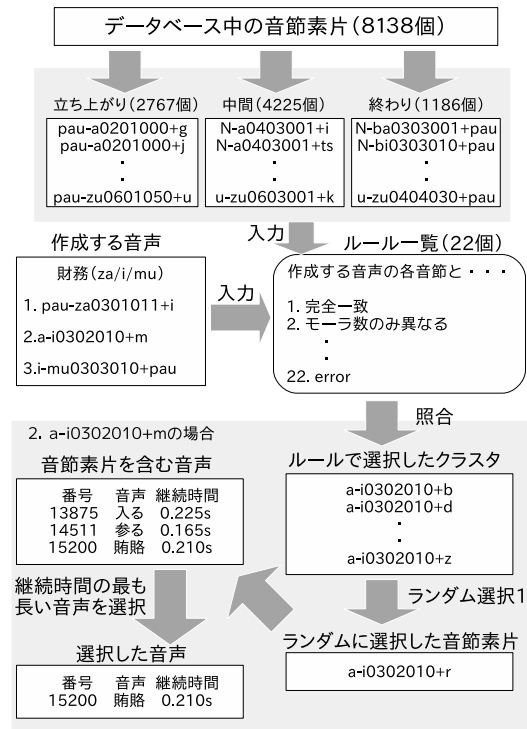


図 4 言語情報のみを用いたクラスタリング手法の流れ図

(1) データベース中の音節素片 8,138 個をポーズ (“pau”) の位置に基づいて、3 つのグループに分類する。

(2) 手順 1 の 3 つのグループと、作成する音声の各音節素片を入力データとして準備する (“a-i0302010+m” が入力された場合を想定)。

(3) 付録の表 A.2 に示す 22 個のルールを用いて順に照合する (作成する音声の音節素片によって、照合するグループを変更)。

(4) 上位のルールから順に照合し、適合した音節素片が 1 つ以上あればそのルールで打ち切る。

(5) 再下位のルールである 22 までに適合する音節素片が見付からない場合、エラーを出力する。

(6) ルールで選択したクラスタ中の音節素片を、ランダムに 1 つ選択する。

(7) 手順 6 で選択した音節素片を含む音声の中で、最も継続時間の長い音声を選択する。ただし連続母音を含む音声の場合は、連続母音の音節素片が同じ音声番号となるように選択する。

5.2 実験条件

本章の実験では、クラスタリングのルールとして、付録の表 A.2 を使用する。また実験に使用する音声は、女性話者 1 名に関して、3 または 4 モーラの 100 単語とする。作成する音声を表 5 に示す。

ただし“言語クラスタ”と“音響クラスタ”は、データベースに存在しない音声を作成する。一方、“自然音声”と“オリジナル合成”は、データベース中の音声を作成する。

表 5 作成する音声

名称	説明
自然音声	ATR 単語発話データベース ASET に含まれる音声
オリジナル合成	クラスタリングを利用しない 音節波形接続型音声合成法で作成した音声
言語クラスタ	言語情報のみを用いた クラスタリングで作成した音声
音響クラスタ (従来手法 [3])	音響情報と言語情報の両方を用いた クラスタリングで作成した音声

また音声品質評価の信頼性を高めるために、合成音声の評価者は、音声研究に関わった経験のない5名とし、評価実験として、オピニオン評価実験と対比較実験を行う。

5.3 実験結果

5.3.1 オピニオン評価実験

オピニオン評価実験の結果を表6に示す。またオピニオン評価実験において、“言語クラスタ”の合成音声の評価例を表7に、合成音声の作成方法を付録の表A.3に示す。

表 6 オピニオン評価実験結果 (総評価単語数:500)

音声の種類	オピニオンスコア
自然音声	4.79
オリジナル合成	4.31
言語クラスタ	3.22
音響クラスタ	2.48

表6より、“言語クラスタ”は、“自然音声”や“オリジナル合成”より低いオピニオンスコアとなった。しかし“音響クラスタ”よりは高いオピニオンスコアであった。

表 7 オピニオン評価実験における“言語クラスタ”の評価例

合成音声	評価者 A	評価者 B	評価者 C	評価者 D	評価者 E	平均
筆箱	5	5	5	4	5	4.8
品数	3	3	3	3	4	3.2
太刀魚	2	1	1	2	1	1.4

表7より、“言語クラスタ”には、非常に良い音声品質の音声もあるが、一方で非常に悪い音声品質の音声もできると分かった。しかしいずれの音声も、評価者によるばらつきはほぼ見られなかった。

5.3.2 追比較実験

対比較実験の結果を表8に示す。

表 8 対比較実験結果 (総評価単語数:1,000)

	言語クラスタ	音響クラスタ
対比較実験	77.8%	22.2%

表8より、“言語クラスタ”は77.8%と高い値を得た。

以上より言語情報のみを用いたクラスタリング手法を音節波形接続型音声合成法へ適用することで、音響情報と言語情報の両方を用いたクラスタリング手法よりも音声品質が改善されることを確認した。

5.4 考察

5.4.1 オピニオン評価実験結果

表6より、“言語クラスタ”は、“自然音声”や“オリジナル合成”より低いオピニオンスコアとなった。この原因として、“言語クラスタ”は、一部の言語情報が緩和された音節素片を用いたのが原因だと考えられる。一方、“自然音声”や“オリジナル合成”は2.1節の表2に示す言語情報が完全に一致した音節素片を用いた。

以上より、一部の言語情報を緩和することで、音声品質が劣化することを確認した。しかし言語情報のみを用いたクラスタリング手法の改良により、さらに“自然音声”や“オリジナル合成”の音声品質に近づけると考えている。

5.4.2 継続時間の影響

4.4節の実験では、継続時間の最も長い音声を選択した結果、オピニオンスコアに変動はなかった。しかし本章の実験で作成した音声は、継続時間の影響により音声品質に悪影響が出た。この原因として4.4節の実験では、評価単語が20単語と少ないため、偶然オピニオンスコアが変わらなかったと考えられる。しかし促音qは継続時間を考慮したことで、良い音声品質を得た。

以上より、今後は促音qのみ継続時間の最も長い音声を選択し、その他の音声は継続時間の平均長である音声を選択した方が良いと考えている。

5.4.3 言語クラスタ

本章の実験結果より、“言語クラスタ”は“音響クラスタ”よりも良い音声品質を得た。その一方で、“自然音声”や“オリジナル合成”と比較すると、まだ十分な音声品質ではない。

しかしながら、“言語クラスタ”は、5.2節の手順6のランダム選択や、付録の表A.2のルール見直し等の改良点がある。例えば、5.2節の手順6のランダム選択は、前後音素環境に着目し、音声合成における音素間の相関関係を調査することで、一意に音節素片を選択できる。

以上より、“言語クラスタ”の音声品質は、今後まだ改善される見込みがある。

5.4.4 音響クラスタ

本研究では、音声合成において、MFCCの距離尺度と音声品質間の相関が低いと仮定した。本実験の結果、“音響クラスタ”は“言語クラスタ”よりも悪い音声品質であり、仮定が正しいと分かった。しかし、本実験で使用したデータベース(5,240単語)では数が少なく、偶然MFCCの距離尺度の類似した音節素片が少なかった可能性がある。また本実験で作成した音声は、女性話者1名に関する音声であり、他の話者でも同様の結果が得られるとは限らない。

以上より、今後も“音響クラスタ”および“言語クラスタ”の両方の可能性を調査していきたい。

6. おわりに

本研究では、言語情報のみを用いたクラスタリング手法の音声品質の調査を行った。

データベースに存在しない単語を作成し、各モデルに対して

聴覚実験を行った。オピニオン評価実験の結果，“言語クラスタ”の音声は3.22，“音響クラスタ”の音声は2.48という値が得られた。また，対比較実験の結果，“言語クラスタ”の音声は77.8%，“音響クラスタ”の音声は22.2%となった。以上より言語情報のみを用いたクラスタリング手法は，音響情報と言語情報の両方を用いたクラスタリング手法に比べて，良い音声品質であると分かった。

今後は言語情報のみを用いたクラスタリング手法を更に改良し，自然音声やオリジナル合成の音声品質を目指す予定である。

謝辞 本論文を執筆するにあたり，参考にさせて頂いた論文，聴覚実験に協力して下さった鳥取大学工学部知能情報工学科池原研究室の友大 謙一氏，大森 裕貴氏，奥村 秀人氏，滝川 晃司氏，鏡味 良太氏に深く感謝いたします。

文 献

- [1] 村上仁一，水澤紀子，東田正信. “音節波形接続方式による単語音声合成”，電子情報通信学会論文誌，pp.1157-1165(2002)
- [2] S. J. Young, J. J. Odell, and P. C. Woodland. “Tree-based state trying for high accuracy acoustic modelling. Proc”, ICASSP, pp.307-312(1994)
- [3] 植村 和久，村上 仁一，池原 悟. “Tree Based Clustering を利用した音節波形接続型音声合成法に関する検討”，SP, pp.81-86(2008).
- [4] NHK 放送文化研究所，“NHK 日本語発音アクセント辞典新版”，NHK 出版 (1998)
- [5] 石田 隆浩，村上 仁一，池原 悟. “モーラ情報とアクセント情報を用いた波形接続型音声合成の普通名詞句への応用”，音響学会，pp.1-409,410(2003)

付 録

表 A.1 クラスタリングの初期ルール

優先順位	内容
1	全ての条件が一致
2	モーラ数以外の全ての条件が一致
3	モーラ位置以外の全ての条件が一致
4	モーラ数，モーラ位置以外の全ての条件が一致
5	後音素環境（子音）以外の全ての条件が一致
6	モーラ数，モーラ位置，後音素環境（子音）以外の全ての条件が一致
7	後音素環境以外の全ての条件が一致
8	モーラ数，モーラ位置，後音素環境以外の全ての条件が一致
9	前音素環境以外の全ての条件が一致
10	モーラ数，モーラ位置，前音素環境以外の全ての条件が一致
11	モーラ数，モーラ位置，前後音素環境以外の全ての条件が一致
12	アクセント型以外の全ての条件が一致
13	モーラ数，モーラ位置，アクセント型以外の全ての条件が一致
14	モーラ数，モーラ位置，前後音素環境，アクセント型以外の全ての条件が一致
15	アクセントの高低以外の全ての条件が一致
16	中心音節のみ一致
17	error

表 A.2 クラスタリングの初期ルール改良後

優先順位	内容
1	全ての条件が一致
2	モーラ数以外の全ての条件が一致
3	モーラ位置以外の全ての条件が一致
4	モーラ数，モーラ位置以外の全ての条件が一致
5	後音素環境（子音）以外の全ての条件が一致
6	モーラ数，モーラ位置，後音素環境（子音）以外の全ての条件が一致
7	後音素環境以外の全ての条件が一致
8	モーラ数，モーラ位置，後音素環境以外の全ての条件が一致
9	前音素環境以外の全ての条件が一致
10	モーラ数，モーラ位置，前音素環境以外の全ての条件が一致
11	モーラ数，モーラ位置，前後音素環境（子音）以外の全ての条件が一致
12	モーラ数，モーラ位置，前後音素環境以外の全ての条件が一致
13	アクセント型以外の全ての条件が一致
14	モーラ数，モーラ位置，アクセント型以外の全ての条件が一致
15	モーラ数，モーラ位置，後音素環境（子音），アクセント型以外の全ての条件が一致
16	モーラ数，モーラ位置，後音素環境，アクセント型以外の全ての条件が一致
17	モーラ数，モーラ位置，前音素環境，アクセント型以外の全ての条件が一致
18	モーラ数，モーラ位置，前後音素環境（子音），アクセント型以外の全ての条件が一致
19	モーラ数，モーラ位置，前後音素環境，アクセント型以外の全ての条件が一致
20	アクセントの高低以外の全ての条件が一致
21	中心音節のみ一致
22	error

表 A.3 5.3.1 節の表 7 における合成音声の作成方法

モーラ位置	筆箱 (fu/de/ba/ko)	品数 (shi/na/ka/zu)	太刀魚 (ta/chi/u/o)
	pau-fu0401000+d u-de0402001+b e-ba0403001+k a-ko0404001+pau	pau-shi0401030+n i-na0402031+k a-ka0403031+z a-zu0404030+pau	pau-ta0401020+ch a-chi0402021+u i-u0403020+o u-o0404020+pau
1 番目	不断 pau-fu0301000+d	支払う pau-shi0401030+h	たとえば pau-ta0401020+tt
2 番目	腕前 u-de0402001+m	災難 i-na0403031+N	落ちる o-chi0302021+r
3 番目	束縛 u-ba0403001+k	ごまかす a-ka0403031+s	不合理 o-u0403020+r
4 番目	煙草 a-ko0303001+pau	紙屑 u-zu0404030+pau	青 a-o0202010+pau

表 A.3 において，1 行目は，音節波形接続型音声合成法で合成音声を作成する際に用いる音節素片である。また 2 行目以降は，言語情報のみを利用したクラスタリング手法で合成音声を作成する際に用いた音節素片である。