概要

音声合成の手法の一つとして波形接続型音声合成がある．この手法は録音音声から音節を単位とし，波形素片を取り出し，信号処理をせずに接続することで，自然性の高い音声を合成できる．また，前後音素環境，モーラ情報，アクセント型などの付加的な情報を用いることで，品質が向上することが知られている．

しかし，任意の一般名詞を作成しようとするには大量の録音単語が必要である，音声データベースとして ATR 単語発話データベース Aset (5240 単語) を使用した場合 5240 単語中の 470 単語しか作成できない．

そこで任意の音声を合成可能にするために，収録されている DB に対して木に基づくクラスタリングを行い，音響パラメータが似た音節素片をグループ化する．グループ化された情報を利用して波形接続型音声合成することで音声を合成し，作成した音声の音質の評価を行う．

聴覚実験ではオピニオン評価実験と対比較実験を行った．

その結果，木に基づくクラスタリング (以後クラスタリング合成) を用いた合成音声では 3.7，自然音声では 4.9，オリジナルの波形接続型音声接続 (以後オリジナル合成) で 4.3 というオピニオンスコアが得られた．クラスタリング合成は自然音声には大きく及ばなかったが品質の高い音声が作成できたことがわかる．

対比較実験結果では自然性とクラスタリング合成の結果は 91 %が自然音声の方が良いと判定されたが 9 %が自然音声よりも良い音だと判断されたことから高い品質の合成音声が作成可能であることが分かった．

目 次

図 目 次

表 目 次

1　はじめに

現在，カーナビゲーションシステムや電車の社内アナウンスなどのように，音声ガイダンスを利用したシステムやサービスが様々な場面において利用されている．このようなシステムでは，録音編集方式が広く使われている．録音編集方式では，まず，システムやサービスに必要となる音声を，システム利用者の入力やサービスの利用される場所・時間などに依存するような比較的短い単語音声 (以下,『可変部』) と，それ以外の比較的長い文節・文音声 (以下,『固定部』に区別する．そして，可変部と固定部を別々に録音しておき，必要に応じて組み合わせることで出力音声を構築する．

例えばカーナビゲーションシステムにおいて,「目的地は○○でよろしいですか.」というガイダンス音声を出力したい場合，○○の部分には，駅名や建物名などの単語音声が挿入される．ユーザーが目的地に「東京駅」を指定した場合，ガイダンス文は「目的地は“ 東京駅 ”でよろしいですか.」となる．例の場合,『東京駅』などの駅名や建物名などの単語音声が可変部,「目的地は　」という部分が固定部となる．

録音編集方式を用いた音声合成においては，可変部と固定部を接続した場合の違和感を軽減するために，一般に同一話者の音声が必要となる．可変部と固定部を分離して録音することにより，必要となるすべての音声を録音する場合に比べて話者に対する負担は若干軽減されるが，可変部に挿入する単語が増大した場合，同一話者から全ての音声を録音することは困難となる．さらに，録音環境の違いにより発話速度 F_0 周波数にばらつきが出るため，安定した品質の音声を得ることは非常に困難となる．

そこで，固定部と可変部に必要になる音声をすべて音声合成によって作成する方法が考えられる．例えば，音素や音節，CV，VCV を単位とした規則音声合成がある．規則音声合成は，古くから TTS 音声合成において用いられてきた方法であり，基本的には，音声の特徴をパラメータとして抽出し，変形することによって合成音声を作成する．PSOLA 方式による音声合成については，現在も多くの研究がなされている．また，最近では HMM を用いて直接音声を合成する研究も行われている．しかし，いずれの場合においても，直接人の声を録音した音声のように，高い品質を安定して得ることが難しい点が問題である．

一方，録音した音声波形の一部 (以下「音声素片」) を用いて別の音声を合成する方法があり，一般に，波形接続型音声合成と呼ばれる．波形接続型音声合成は，音声素片を取り出し，接続することによって合成音声を作成する．接続単位については，音素，CV，VCV，CVC など，様々な単位が提案されている．

しかし，波形接続型音声合成においては，音声波形に信号処理を加えないため，韻律の扱いが問題となる．最も，波形接続型音声合成に限らず，一般に音声合成において，韻律制御は重大な課題であるが，音声合成の対象として小さな単位である単語を合成する場合においては，地名などの固有名詞では F_0 周波数のばらつきが比較的小さく，アクセント型がほぼ一意に決まるため，F_0 周波数とモーラ情報の依存関係を効率的に利用することが可能である．そして，固有名詞を対象とした実験では，実用的な品質が得られたと報告されている．また，普通名詞に適用した場合も，明瞭性の高い合成音声が作成でき，さらにアクセント情報としてアクセント型を考慮することで，より自然性に近い合成音声に近い合成音声の作成が可能であることが示されている．

しかし，波形接続音声では音節素片選択時の条件が厳しく作成できる音声の数が少ないという問題がある，そこで本研究では木に基づく状態共有を用いて素片選択の条件を緩和することで作成できる音声を増やし，条件を緩和したことによる作成した音声の劣化が考えられるので音質の評価を行った．

以降，2 章，3 章で波形接続型音声合成と木に基づく状態共有の説明する，そして 4 章で評価実験に関する説明を行い，5 章で実験結果を報告する．実験により表れた問題点を 6 章で述べ，7 章で 6 章で述べた問題の考察をする．

2　波形接続型音声合成

2.1　モーラ情報とアクセント情報

波形接続型音声合成を含め，一般的に音声合成においては，韻律の扱いが問題となる．韻律を扱う場合，録音音声および出力音声の F_0 周波数が必要となる．しかし，正確な F_0 周波数を直接推定することは困難である．

一方，音声合成の対象として小さな単位である単語を合成する場合においては，地名などの固有名詞では F_0 周波数のばらつきが比較的小さく，アクセント型がほぼ一意に決まるため， F_0 周波数とモーラ情報が依存関係を効果的に利用することが可能である．そして，このモーラ情報は音素ラベリングや音声認識など分野において効果があることが報告されている．

しかし，より一般的な普通名詞では例えば「雨」と「飴」のように同音異義語が多数表れるため，モーラ情報を考慮しただけでは不適切な音声素片が選択される場合がある．そして，普通名詞で素片選択においてモーラ情報に加えてアクセント型を考慮した研究が行われており，アクセント型が合成音声の自然性を向上するために有効であることが示されている [1].

2.2 波形接続型合成音声の問題点

過去の研究 [1] より波形接続型音声合成の有用性が示されたが波形接続型音声合成では音節素片選択の条件が厳しいために作成できる音声が少ないという問題がある，音声データベースとして，ATR 単語発話データベース Aset(5240 単語) を使用した場合，5240 件中の 470 単語しか作成できない [1].

そこで任意の音声を合成可能にするために，収録されている DB に対して木に基づくクラスタリング [3] を行い，音響パラメータが似た音節素片をグループ化する．波形接続型音声合成でこれを利用して音声を合成し音質の評価を行う.

2.3 波形接続型音声合成の概説

本研究で用いる波形接続型音声合成では、まず、以下の情報が一致する素片を選択する．単語のアクセントについては，NHK 日本語発音アクセント辞典 [4] を参考にラベルデータに対してアクセントを付加する.

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

そして，素片の開始時間と終了時間を元に波形データを切りだし，接続して合成音声を作成する.

2.4 波形接続に関する捕捉

波形接続型音声合成では，接続部の違和感の発生が自然性に大きく影響する．本研究では，波形の接続位置を音素境界とする．さらに，接続部における 2 素片間の波形の位相を考慮し，接続部の振幅の差がゼロに近付くように調整を行う．具体的には，あらかじめラベル付けされた素片終了時間をもとに，振幅が負から正に変わる部分を，波形が短くなる方向 (開始時間は進む方向，終了時間は戻る方向) に探し，抽出する位置を修正する．

2.5 波形接続型音声合成の例

本研究で作成した合成音声の例を以下に示す．なお「__|¯」は音の強弱（アクセント）を表し，太字の部分が音節素片である．

一帯 (/i_/q̄/t̄a/i/) = 一掃 (/**i**_/**q̄**/t̄a/i/)
+ 実態 (/j**i**_/**q̄**/**t̄a**/i/)
+ 絶対 (/z**e**_/**q̄**/**t̄a**/**i**/)
+ 組合 (/k**u**_/**m̄i**/**a**/**i**/)

感化 (/i_/q̄/t̄a/i/) = 看護 (/**ka**_/**N̄**/ḡo/)
+ 頑固 (/g**a**_/**N̄**/**k̄o**/)
+ 文化 (/b**u**_/**N̄**/**k̄a**/)

最善 (/s**a**_/**ī**/**z̄e**/N/) = 災害 (/s**a**_/**ī**/ḡa/i/)
+ 外人 (/g**a**_/**ī**/**z̄h**ī/N̄/)
+ 改善 (/k**a**_/**i**/**z̄e**/**N̄**/)
+ 回転 (/k**a**_/**ī**/**t̄e**/**N̄**/)

3　木に基づく状態共有について

木に基づくクラスタリング [3] は，音声認識で学習データを効率良く使うためによく使われている．音響的特徴が類似した triphone HMMの状態集合に対して音声の決定木に基づいてクラスタリングを行う．

本研究において，前後素環境に加えて，モーラ長，モーラ位置，アクセント型を考慮した質問を用いて木に基づく状態共有を行い，共有されたHMMに基づいて音声合成を行う．質問の例を以下に示す．

- 前音素環境は鼻音であるか?
- モーラ数は 3 または 4 で，モーラ位置は１であるか?
- アクセント型は１，２，３のどれであるか?

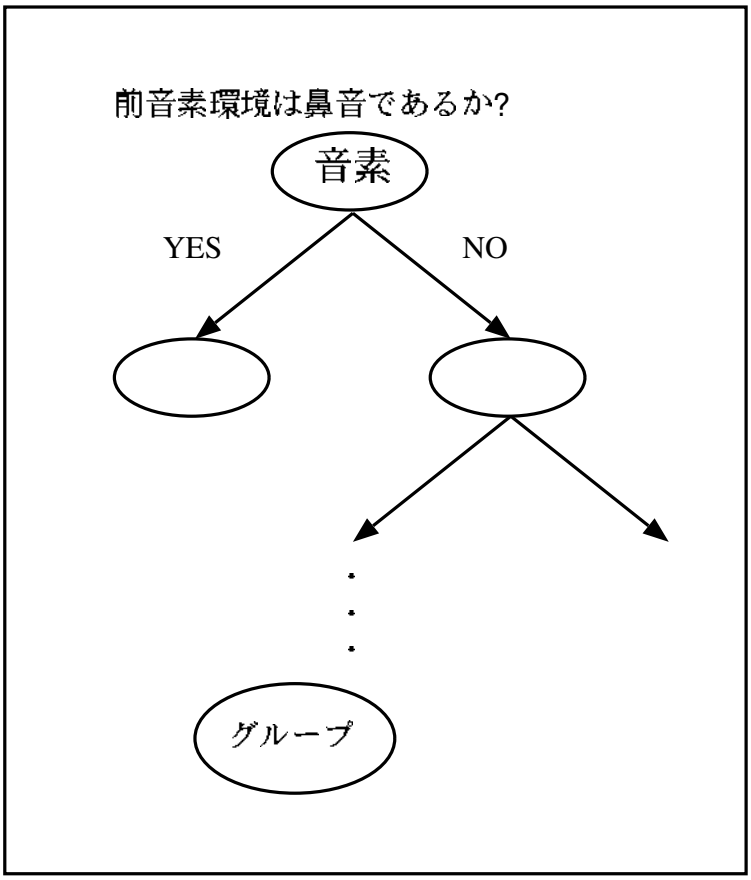

図 1: 木に基づくクラスタリングの例

状態共有された音素の例を表１に示す．例では，前音素やアクセント型の異なるHMMが状態共有されている．なお，HMMの学習データにはAsetの話者fynの奇数番２６２０単語を用いる．

状態を共有された音素は同一とみなすことで波形接続型音声合成の条件の緩和を行った．

表 1: 状態共有された音節素片の一部 (te)

音節	前音素	後音素	モーラ数	モーラ位置	アクセント型	アクセントの高低
	a	n	3	2	0	高
	a	g	3	2	2	高
	i	N	3	2	0	高
te	i	k	3	2	0	高
	i	ts	3	2	0	高
	i	N	4	3	0	高
	i	k	4	3	0	高

表 2: 状態共有された音節素片の一部 (N)

音節	前音素	後音素	モーラ数	モーラ位置	アクセント型	アクセントの高低
	a	k	3	2	1	低
N	a	k	4	2	0	低
	a	k	5	2	1	低

表 3: 状態共有された音節素片の一部 (te)

音節	前音素	後音素	モーラ数	モーラ位置	アクセント型	アクセントの高低
	q	pau	3	3	0	高
	q	N	4	3	0	高
ki	q	Ns	4	3	3	高
	q	r	4	3	3	高

3.1 木に基づくクラスタリングを利用した波形接続型合成音声の例

本研究で作成した合成音声の例を以下に示す．なお「_|¯」は音の強弱（アクセント）を表し，太字の部分が音節素片である．

一帯 (/i_/q_/ta_/i_/)= 威張る (/**i**_/ba_/ru_/)
+ 切手 (/ki_/q_/te_)
+ めでたい (/me_/de_/**ta**_/i_/)
+ 洪水 (/ko_/o_/zu_/**i**_/)

感化 (/ka_/N_/ka_/_/) = 観点 (/**ka**_/N_/te_/N_/)
+ 観光 (/ka_/**N**_/ko_/o)
+ 悪化 (/a_/q_/**ka**_/)

最善 (/sa_/i_/ze_/N_/) = 細工 (/**sa**_/i_/ku_/)¯
+ 追い出す (/o_/**i**_/da_/su/)
+ 安全 (/a_/N_/**ze**_/N_/)
+ 回転 (/ka_/i_/te_/**N**_/)

4　評価実験

4.1　実験環境

作成した合成音声の評価の為に聴覚実験を行う.

本実験では音声データベースとして，ATR 単語発話データベース Aset(5240 件) を使用する．そして，Aset に含まれる 3,4 モーラ語の 100 音声 (各 50 音声) を利用する．そして以下の条件の音声を準備する.

- 自然音声
- オリジナルの波形接続型合成音声 (以後オリジナル合成音声)
- 木に基づく状態共有を用いた波形接続型合成音声 (以後クラスタリング合成音声)

4.2 評価方法

合成音声の評価のために，音声研究に関わったことのない 4 名を対象に，自然音声と合成音声をランダムにヘッドフォンから被験者に聴かせ，オピニオン評価，対比較実験の 3 つの実験を行う．

(1) オピニオン評価

音声の自然性を調べるためにオピニオン評価を行う．オピニオン評価では，自然に聞こえた度合を 5 段階 (5 が最も自然，1 が最も不自然) で評価するように指示する．

(2) 対比較実験

作成した音声の評価のために対比較実験を行う．対比較実験では自然音声と合成音声の同じ内容の文節で 2 種類の音声を続けて聞かせ，どちらの音声が自然に聞こえたかを判定してもらう．

5　実験結果

5.1　オピニオン評価結果

オピニオン評価の全被験者の平均を下に示す．クラスタリング合成の結果は 3.7 とな

表 4: オピニオン評価の結果

	オピニオンスコア 評価音節数：100
自然音声	4.9
オリジナル合成	4.3
クラスタリング合成	3.7

り高い品質の合成音声が作成できたことわかったが，やはり自然音声と比べてみると自然性の面では差があることがわかった．

5.2　対比較実験結果

以下の三通りの対比較実験を行った，その結果を表 3 に示す．

1) 自然音声とオリジナル合成

2) 自然音声とクラスタリング合成

3) オリジナル合成とクラスタリング合成

表 5: 対比較実験の結果

	比較対象 1	比較対象 2
(1)	自然音声：79%	オリジナル合成：21%
(2)	自然音声：91%	クラスタリング合成：9%
(3)	オリジナル合成：76%	クラスタリング合成：24%

対比較実験ではクラスタリング合成が良い音声だと判定された文節が 9.25 %あった．この結果から合成音声は品質が高い音声が作成されたと言えるが，自然性の面では自然音声との間にはまだ差がある事がわかる．

6　問題点

6.1　接続部の違和感

波形接続型音声接続では，人手でラベリングされた情報を利用して音節素片を切り出す．しかし，ラベルに誤りがある場合がある. このとき余分な前後の音声が入ることがある．

木に基づくクラスタリングを利用しない波形接続型音声合成では前後の音素を考慮している為に多少前後の音があっても違和感無く音声が合成できる．しかし，木に基づくクラスタリングを利用した合成音声では，グループ化された音素からランダムに音を選択している為に前後環境は考慮されていない．その結果，合成音声に違和感が生じ品質の低下に繋がったと考えている．

6.2　アクセントの高低

木に基づくクラスタリングでグループ化された音響パラメータが似た音節素片の中には，アクセントの高低が違うものがグループ化されている事があった．しかしこれをグループ化すると品質が劣化することは明白であるのでアクセントの高低の違いに関しては本実験ではグループ化から除外した．

6.3　アクセントの位置

アクセントの高低は同じであるがアクセント位置が違う単語がある．例をあげると “宛名” の “TE” と “建具” の “TE” である．この二つの音節素片は木に基づくクラスタリングで音響パラメータが似ていると判断された，しかし，宛名の “TE” は次の音素でアクセントが変化しないが “建具” の “TE” は次の音素でアクセントが下がるためにこれらを同じと “TE” とみなすことで音質の劣化が見られる場合がある．

7　考察

7.1　接続部の違和感の問題

人手でラベリングされた情報を用いるために前後の音素が混入してしまうのは回避するのは難しい，しかしこのように前後の音素が混入するケースはそれほどに多くはない，そのために，素片選択時にまず前後の音素を考慮しその条件の音が無い場合に前後の音素を無視するなどの手法をとることによりある程度は軽減できると考える．

7.2　アクセント問題

本実験ではアクセントの高低に関しては，考慮して実験を行った，しかしアクセントの位置については考慮に入れてなかった，アクセントの位置を考慮に入れることで多少オピニオンスコアが上がると考えているが作成できる音声の数が減少するとも懸念される．

そこで今後はどの程度のオピニオンスコアでどの程度の数の音声をカバーできるかのボーダーを考えて行くことが重要であると考えている．

8　おわりに

本実験では，波形接続型音声合成で任意の合成音声を作成するために音声が収録されている DB に木に基づくクラスタリングを行い，DB の音素を音響的に似ているとみなせるグループにグループ化した．

波形接続型音声合成の素片選択の時に，クラスタリングされた情報を用いることで素片選択時の条件が緩和され，作成できる音声が増加したが音質の劣化が懸念されるためにオピニオン評価実験と対比較実験の二つの聴覚実験を行った．

聴覚実験の結果，オピニオンスコアで 3.7 という値が得られた．クラスタリングにより条件を緩和したことで波形接続型合成音声よりも多少オピニオンスコアが下がったものの自然性の高い音声が作成でき，任意の音が作成可能であることがわかった．

今後は，木に基づくクラスタリングにアクセント位置を考慮に入れるなどの検討を行い，より自然性の高い合成音声の作成を目指したい．

謝辞

本研究・本論文作成に際して，多大なる検討と種々の御助言をしていただきました鳥取大学工学部知能情報工学科計算機Ｃ工学講座池原研究室の池原悟教授，村上仁一助教授に心からお礼を申し上げます．また, 論文を執筆にあたり, 助言を頂いた徳久助手にお礼を申し上げます．さらに木に基づくクラスタリングに関して計算機工学講座博士前期課程 2 年の堀田波星夫さんに，聴覚実験の被験者には計算機工学講座学部 4 年生の方に協力して頂きました．心より御礼申し上げます．

参考文献

[1] 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”，音響全体，2-Q-18, pp.1-409,410(2003-3).

[2] 堀田 波星夫，村上 仁一，池原 悟. “不特定話者における同音異義語音声認識”，NO. 2-1-11(2006).

[3] 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).

[4] “NHK 日本語アクセント辞典 新版”，NHK 出版，ISBN4-14-011112-7(1998).

付録1：作成したクラスタリング合成音声一覧

表 6: 作成した音声の一部 (30)

音声	音節 1	音節 2	音節 3	音節 4
温泉	恩 (/**o**/N/)	金星 (/ki/**N**/se/i/)	光線 (/ko/o/**se**/N/)	運転 (/u/N/te/**N**/)
改革	海外 (/**ka**/i/ka/ku/)	大会 (/ta/**i/ka/i/)**	快活 (/ka/i/**ka**/tsu/)	なるべく (/na/ru/be/**ku**/)
開催	開始 (/**ka**/i/sa/i/)	愛想 (/a/ɯ/so/)	体裁 (/te/i/**sa**/i/)	温かい (/a/ta/taa/**ka**/i/)
改善	会議 (/**ka**/i/gi/)	会話 (/ka/**i**/wa/)	安全 (/a/N/**ze**/N/)	点検 (/te/N/ke/**N**/)
威厳	異義 (/**i**/gi/)	無期限 (/mu/**ki**/ge/N/)	保険 (/ho/ke/**N**/)	
一帯	威張る (/**i** / ba/ ru/)	切手 (/ki/**q**/ te/)	めでたい (/me /de /**ta**/i/)	洪水 (/ko/o/zu/**i**/)
移転	委託 (/**i**/ta/ku/)	回転 (/ka /i /**te**/N/)	縁 (/ e/**N**/)	
会費	海外 (/**ka** /i / ga/ i/)	相手 (/a /**i** /te /)	朝日 (/a /sa /**hi** / /)	
買い物	開始 (/**ka** /i /shi /)	着物 (/ki/**mo** /no /)	飲み物 (/no/mi/**mo**/no /)	もの (/mo/**no** /)
会話	改革 (/**ka**/i/ka /ku/)	最大 (/sa /**i** /da/i/)	しわ (/shi/**wa** /)	
拡大	覚悟 (/**ka**/ku/da/i/)	肉体 (/ni/**ku** /ta/i/)	問題 (/mo/n/**da**/i/)	明るい (/a/ka/ru/**i**/)
角度	書く (/**ka**/ku/)	昨日 (/ sa/**ku** /zi /tsu /)	制度 (/se /i /**do**/)	
形見	形 (/**ka**/ta/chi /)	あらためて (/a/ra/**ta** /me/te/)	波 (/na/**mi** /)	
瓦	乾かす (/**ka**/wa/ka/su/)	慌てる (/a/**wa** /te/ru/)	皿 (/sa/**ra** /)	
代わり	帰り (/**ka**/e /ri /)	慌てる (/a /**wa**/te/ru/)	限り (/ka/gi/**ri**/)	
外人	外部 (/**ga**/i/bu/)	いじる (/**i** /zhi/ru/)	叔父 (/o/**zhi** /)	税金 (/ze/i/ki/**N**/)
学期	柄 (/**ga**/ra/)	トラック (/to/ra/**q** /ku/)	鉄筋 (/te/q/**ki**/N/)	
頑固	癌 (/**ga**/N/)	三角 (/sa/**N** /ka/ku/)	過去 (/ka/**ko** /)	
基板	記録 (/**ki**/ro/ku/)	番地 (/**ba**/N/chi/)	酸 (/sa/**N** /)	
検査	憲法 (/**ke**/N/po/o/)	現代 (/ge/**N** /da/i/)	審査 (/ke/N/**sa** /)	
現役	言語 (/**ge**/N /go/)	原料 (/ge/**N** /ryo/u/)	絵本 (/**e**/ho/N/)	演劇 (/e/N/ge/**ki**/)
盛り	悟る (/**sa**/to /ru/)	疲れる (/tu/**ka** /re /ru /)	限り (/ka/gi/**ri**/)	
盛ん	里 (/**sa**/to/)	半ば (/na/**ka** /ba/)	案 (/a/**N**/)	
酸素	桟橋 (/**sa**/N/ba/shi/)	男子 (/da/**N** /shi/)	味噌 (/mi/**so**/)	
雑談	残業 (/**za** /tsu /da/N/)	実話 (/zhi/**tsu** /wa/)	診断 (/shi/N/**da**/N/)	海岸 (/ka/i/ga/**N**/)
資源	無期限 (/**mu**/ki/ge/N/)	期限 (/ki/**ge** /N/)	多分 (/ta/bu/**N**/)	
次第	知らせ (/**shi**/ra/se/)	左 (/hi /**da** /ri/)	白い (/shi /ro/**i**/)	
尺度	車庫 (/**sya**/ko/)	握手 (/a/**ku** /syu/)	温度 (/o /N /**do** /)	
写真	しゃがむ (/**sya**/ga/mu/)	純真 (/zyu/N/**shi** /N/)	イン (/i/**N** /)	
進化	審議 (/**shi**/N/ka/)	人口 (/zhi/**N**/ko /o/)	国家 (/ko/q/**ka** /)	
辞典	時間 (/**zhi** /ka/N/)	宛名 (/a/**te**/na/)	点 (/te/**N**/)	
地盤	時代 (/**zhi**/da/i/)	看板 (/ka/N/**ba**/N/)	癌 (/ga/**N**/)	
純真	熟練 (/**zhu**/ku/re/N/)	印刷 (/i/**N** /sa/tsu/)	最新 (/sa/i/**shi**/N/)	更新 (/ko/o/shi/**N**/)
人権	人種 (/**zhi** /N/shju/ /)	陰気 (/i/**N** /ki/)	団結 (/da/N/**ke**/tsu/)	点検 (/te/N/ke/**N**/)
説得	責任 (/**se**/ki/ni/N/)	エチケット (/e/chi/ke/**q**/to/)	もっとも (/mo/q/**to**/mo/)	継続 (/ke/i/zo/**ku**/)
設立	石炭 (/**se**/ki/ta/N/)	脱落 (/da/**tsu**/ra/ku/)	効率 (/ko/o/**ri**/tsu/)	鉛筆 (/e/N/pi/**tsu**/)
選挙	銭湯 (/**se**/N/to/o/)	限度 (/ge/**N**/do/)	検挙 (/ke/N/**kjo**/)	

付録2：クラスタリングされた情報の一部