音声言語情報処理 13－1
（1996．10．25）

発話末の基本周波数とパワーのパターン分類とその分析

Classfication for Pattern of F0 and Power on the Last Part of Utterance Fragment and its Analysis

佐々木 聡 Hajime Sasaki
神田 祐和 Hirokazu Kanda
堀内 靖雄 Yasuo Horiuchi
市川 熹 Akira Ichikawa

千葉大学
Chiba University

Abstract

The prosody seems to play important roles for the understanding of spontaneous spoken dialogue.It was reported that the pattern of prosody on the end of utterance fragment is available for dialogue control. But so far, the pattern was infered subjectively. In this paper, we represent the pattern of prosody as codes sequence to infer the pattern objectively. First we transform the pattern of prosody into approximate lines. Then we code them by Vector quantization of inclination of approximate line. Using this codes sequence of prosody, we can infer the dialogue control type with about 70% accuracy. And we reached the conclusion that it is sufficient to use code book size 9 or 11 in Vector quantization. In the latter part of this paper, we examine about automation of all process that are auto-labeling, approximation and classification.

1 はじめに

音声対話システムとユーザがスムーズに自然な対話を進行させるためには、システムとユーザの間で適切な話者交代が行なわれるようなモデルが必要となる。現在の音声対話システムでは、音韻情報のみを利用した対話モデルが用いられている。しかし、自然な対話の中には、文法的に不的確な表現や述語の省略、同音にもかかわらず発話が終了する場合と継続する場合があるなど、音韻情報だけでは処理できないものも存在している。このような場合、音韻情報のみを用いた対話モデルでは、適切な話者交替を行なうことができず、その結果、ユーザの発話に制限を加えることになり、自然な対話を行なうことが困難になる。

このように音韻情報だけでは処理が困難な場合、発話者は他の何らかの情報を用いて相手に自分の意志を伝えていると考えられる。その１つとして音声に含まれている抑揚情報が考えられる。音韻情報だけで処理が困難な部分では、文字化する場合に欠落してしまう抑揚情報が、何等かの働きをして処理の手助けをしていると考えられる。実際に、発話されたものの抑揚情報を見てみると、発話された状況や発話者の意志によってその波形は変化している。自然な対話の中で、述語が省略されたり、同音にもかかわらず発話が終了する場合と継続する場合があっ

ても対話がスムーズに進行できるのは、発話の終端部において抑揚情報が重要な働きをしていると考えられる。

その観点から、発話断片末における音韻情報と抑揚情報である基本周波数パターンとパワー波形パターンを用いて発話の終了、継続を予測しようという研究が行なわれている [1]。その結果、この 3 つの情報を組み合わせることで、かなり高い精度で発話終了、継続の判定が可能であるということが報告されている。しかし、この研究では基本周波数パターンとパワー波形パターンの判断は研究者が主観的に行なっているため、システムに応用するにはこれらのパターンの客観的な判断による発話の分類、さらに、処理を自動化し、大量のデータを用いて、定量的な検討を行なう必要がある [2]。

そこで本研究では、抑揚情報として基本周波数、パワーに着目し、それらのパターンと発話制御との関係を分析した。なお、[1] と異なり、大量のデータを客観的に分析することを目標としているため発話継続、発話終了のラベリング、抑揚パターンのコード化などすべて機械的に行なう。そこで本研究では以下の手順で分析を行なった。

1. 発話の基本周波数パターンとパワー波形パターンのベクトル量子化によるコード化

2. 手動でラベルづけしたデータに対して、コード列を発話制御の関係を分析し、ベクトル量子化のコード数を決める

3. 大量データに対し、発話継続、発話終了の自動ラベリング、および上記の方法で決定したベクトル量子化により各発話末の抑揚パターンをコード化し、コード列を発話制御との関係を分析

2 抑揚情報のコード化

基本周波数パターン、パワー波形パターンを客観的に判断するために、そのパターンをコード列として表現する。そのために、以下の処理を行なった。

1. 基本周波数パターン、パワー波形パターンの直線近似

2. 近似直線のコード化

音声データから基本周波数およびパワーを抽出し、その波形の直線近似を行なう。その後、直線の勾配を利用したコード化を行ない、抑揚情報のパターンをコード列として表現する。

今回コード化数を決定するのに利用した音声対話データは、千葉大学地図課題コーパス [3] の 8 対話 (2312 発話、男性対話 6、女性対話 2) である。また、本研究では、前後を 400[msec] 以上の無音区間で区切られた単位を 1 つの発話として扱っている。

2.1 抑揚情報パターンの直線近似およびコード化

抑揚情報に対する直線近似は最小二乗法を用いて行なった。その時、たて軸として対数スケールを用いて直線近似を行なった。

これによって得られた近似直線の勾配に対し、量子化を行なうことにより対応するコードを決定する。また、発話の時間長をコードの個数として表している。この処理で抑揚情報の各パターンのコード化を行なった。基本周波数パターンの直線近似とコード化の例を図 1に示す。基本的なコードとして、上昇が [u] 、平坦が [f] 、下降が [d] で表現されている。さらに、上昇と下降に関しては近似直線の勾配の値に応じて、各コードの後ろに数字が付加される。

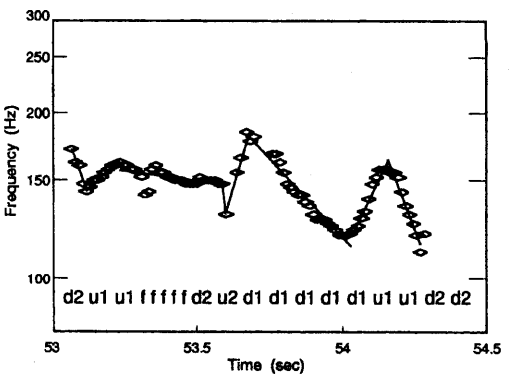

図 1: コード化の例 (F0)

なお、今回使用した音声対話データでは、発話の終端部の近似直線の平均時間長が約 75[msec] であることから、本研究ではコードの単位時間長を 75[msec] としてコード化を行なっている。また、基本周波数

コード数	幅	
	F0	Power
3	3.0	720.0
5	2.0	240.0
7	1.2	144.0
9	0.8	102.9
11	0.66	80.0
13	0.54	66.5
15	0.46	55.4

表 1: 量子化の種類

コード	勾配
u3	$3.0 < x$
u2	$1.8 < x \leq 3.0$
u1	$0.6 < x \leq 1.8$
f	$-0.6 \leq x \leq 0.6$
d1	$-1.8 \leq x < -0.6$
d2	$-3.0 \leq x < -1.8$
d3	$x < -3.0$

表 2: コード対応例 (F0)

に関しては、対数周波数に対する直線の勾配をとることにより、男女の音声を同一に扱うことが可能なことも報告されている [2]。

近似直線のコード化を行なう場合、直線の勾配の量子化をどの程度まで細かく行なうかが問題となる。そこで今回は、表 1に示す 7 種類の量子化を行ない、各量子化によるコードパターンで発話の分類を行なった場合の発話行為の判定の正解率を調べ、その結果から量子化のパラメータを決定する。

なお、発話の終端部の近似直線の勾配は、基本周波数は-3.0 から 3.0、パワー波形では-360.0 から 360.0 の範囲に約 90%の分布があったので、この範囲を表 1 の幅の値で量子化を行なった。また、0.0 が平坦を表すコード [f] の中心になるように量子化を行なっている。基本周波数パターンをコード数 7 でコード化する場合の近似直線の勾配とコードの対応例を表 2 に示す。

2.2 コード列による発話の分類

発話の客観的な分類を行なうために、前述の方法によりコード化された抑揚情報を用いて、各発話の終端部のコード列による発話の分類を行うことを考える。抑揚情報のパターンをコード列から客観的に判断を行ない、同じコード列を持つ発話を 1 つのグループに分類を行なっている。このように分類を行ない、分類された発話を分析することで、各パターンが発話制御において果たしている役割を調べることが可能となる。

発話の分類の方法は、コード列として表現された各発話の終端部の 3 コードのマッチングを行うことで発話の分類を行なった。コード列のマッチングは、発話の最後尾のコードから順に行なっていく。つまり、最後尾のコードが一致するものを 1 つのグループに分類し、さらにそのグループの中で次のコード (後ろから 2 番目のコード) が一致するものをさらにグループにまとめる。この処理を繰り返すことにより、さまざまなコード長 1 によるコード列と発話制御の関係を調べることができるが、本研究では発話末 1 モーラ程度の情報が発話制御に関与しているという報告 [1] をもとに、最後尾から 3 コードのコードパターンで分類を行なった (図 2)。この分類方法を用いることにより、最後尾のコードから階層的に分類が可能である。

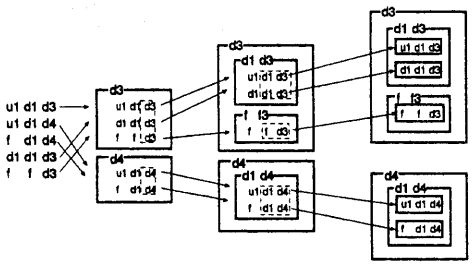

図 2: コードによる分類の例

3 勾配の量子化の検討

先に述べたように、近似直線の勾配を量子化する場合、どの程度まで細かく量子化を行なうべきか、また、どの程度細かく量子化を行なえば十分なのか

が問題になる。そこで本研究では、表1に示す7種類の各量子化を実際に行ない、その量子化によって得られるコード列で発話の分類を行なった場合、どの程度発話終了と発話継続が判定できるか調べ、量子化レベルの検討を行なった。

今回の検討の方法は、音声対話データの発話すべてに対して、発話継続、発話終了、あいづちのラベル付けを上述の8対話に対して手動で行ない、各量子化レベルを用いて抑揚情報のコード化および発話の分類を行った。そして、各コードパターンに分類された発話について出現頻度が最も高いラベル（継続、終了、あいづち）を正解とする。そして、全発話のうち正解として扱われる発話数を調べ、正解率を求めた。各量子化での正解率の比較を行ない、量子化の検討を行なった。

基本周波数に関する量子化の結果を図3に示す。図3は、各コード数で量子化を行なった場合に扱わなければならないコードパターンの最大数とそのコード数での正解率を示している。

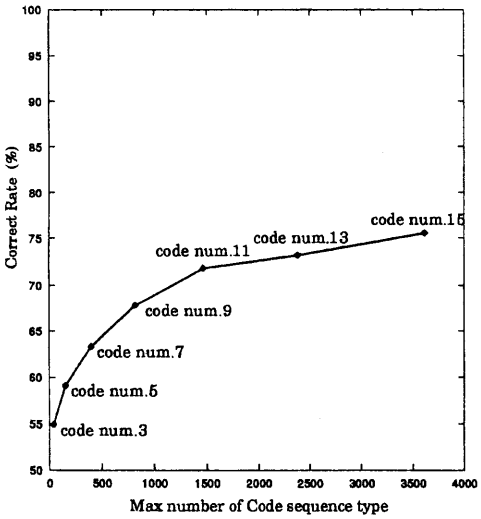

図3: 最大コードパターン数と正解率（基本周波数）

正解率は、コード数を増やすにつれて向上している。しかし、勾配を細かく量子化する場合、出現頻度が低いパターンの数が増えることによって、正解率が上昇すると考えられる（例えば出現頻度が1のものは必ず正解となるため）。しかし、各コード数で量子化を行なった場合に扱わなければならないコードパターンの最大数と正解率の関係を見てみると、コード数9以上では最大パターン数に比べて正解率は飽和している。このことから、近似直線の勾配の量子化を行なう場合には、コード数を9または11程度で行なえば十分であると考えられる。また、コード数が9以上では、基本周波数のパターンのみを用いても、正解率が約70%程度得られることがわかった。

さらに、上記の検討で不正解として扱われる発話について検討を行なった。今回、最も発話数が多く分類されたコードパターン [fff] について検討を行なった。また、コード数は先の検討から9個のコードを用いた場合を選択している。コードパターン [fff] では、発話継続の出現頻度が頻度が高いため、発話終了とあいづちは不正解として扱われる。そこで、発話終了となる発話を分析してみると、66.7%が音韻情報から発話終了と判断できるものとなっていた(図4)。したがって、音韻情報と組み合わせていくことで正解率の向上が図れると考えられる。

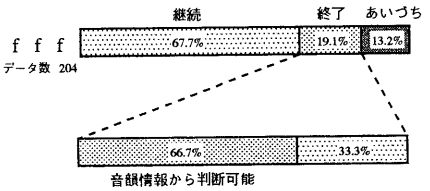

図4: コードパターン [fff] の検討

パワーに関する結果を図5に示す。パワーの場合も基本周波数の場合と同様にコード数を増やすことで正解率が向上するが、コード数9以上では最大コードパターン数に対して、正解率が飽和していると考えられる。また、パワー情報のみではコード数9以上で約65%程度の正解率が得られた。

4 定量化

本研究では、定量的な分析を行なうために処理の自動化に関して検討を行なった。発話に対する継続、終了のラベルづけの自動化を行ない、基本周波数とパワーのコード列による分類の結果に関して検討を行なった。

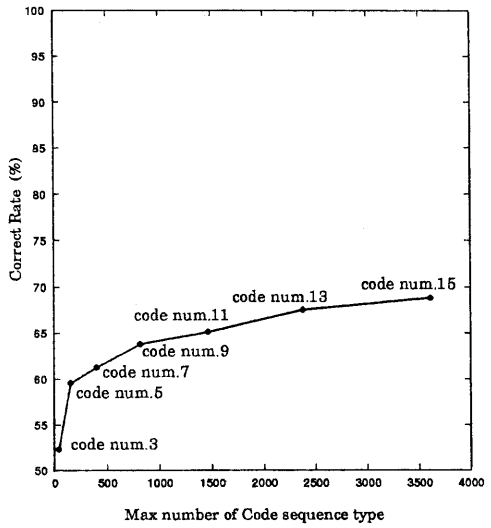

図 5: 最大コードパターン数と正解率（パワー）

4.1 検討方法

1. 基本周波数、パワーの計算
2. 最小二乗法による直線近似
3. 近似直線に対する量子化
4. 発話に対する継続、終了のラベルづけ
5. 発話終端部の 3 つのコード列とその発話のラベルとの比較

なお、今回の量子化は先に述べた理由によりコード数 9 とした。また発話のラベルづけを自動的に行なうため、一つの発話に対して次の発話の話者が移っているものには発話終了、移っていないものには発話継続というラベルをつけた。ただし、次の発話が 100msec 以内の重複発話（ラッチング発話）になっているものに対しては終了とした。

4.2 使用データ

今回は千葉大学地図課題コーパスで収録された対話の中の 24 対話（発話数 10430、男性対話 16、女性対話 8）を使用した。また、一つの発話は 400msec の無音区間で挟まれているものとする。

コード列	ラベル			
	手動		自動	
	継続	終了	継続	終了
d1 d1 d1	51	35	136	208
d1 d4 d1	1	9	15	24
f f d4	27	2	56	31
d1 d4 f	1	12	28	35
d1 f f	9	2	19	27
f f f	138	66	541	441
f f u2	13	7	33	67
d1 d3 u4	2	12	1	17
d2 d4 u4	2	9	18	11
f f u4	12	4	36	63

表 3: 同コード列でのラベルの違い（基本周波数）

4.3 検討結果

この実験によって基本周波数パターンからはコード列のパターン数が 757、正解率は 64.38%になり、パワー波形パターンからはパターン数 440、正解率 62.14%という結果が得られた。これらは、2 節で示したコード数 9 の正解率に比べ、データ量が増えているにも関わらずいくらか低くなっている。

その原因として、2 節の実験で用いたラベルづけは手動で行なったため、あいづちを含めていることがあげられる。すなわち、あいづちの前発話は手動ラベリングでは継続ラベルとすることができるが、機械的にラベリングをする場合には、あいづちの認識が不可能であるため、あいづちの前発話は発話終了ラベルとなってしまう。しかし、[1] によれば、あいづちの前の発話の韻率的特徴は話者継続の傾向を示すと考えられる。そのため、自動ラベリングによる正解率が低くなってしまう。そこで人手によってラベルづけしたものと自動的にラベルづけを行なったものをコード列別にラベルの割合を比較したものの幾つかをを表 3、 4に示した。

これらの表から、手動でつけたラベルと自動でつけたラベルに大きな違いがあるのが分かる。例えば、表 3の一番上のコード列 (d1 d1 d1) においては、手動によるラベルづけでは約 6 割が発話継続であるのに対し、自動によるラベルづけでは反対に約 6 割近くが発話終了になってしまっている。

コード列	ラベル			
	手動		自動	
	継続	終了	継続	終了
f u2 d1	3	16	21	51
f f d2	127	41	334	378
u4 u2 d2	2	17	6	10
f f d3	43	8	119	96
u3 f d3	6	25	21	35
u4 f d3	14	49	41	43
u4 u1 d3	3	23	11	10
f f d4	11	5	29	30
u4 f d4	5	15	24	17

表 4: 同コード列でのラベルの違い（パワー）

5 まとめ

本稿では、発話末の抑揚情報を利用することで発話の終了、継続が高い精度で判定できるという報告[1]から、それを客観化、定量化するために、基本周波数パターンおよびパワー波形の直線近似を行ない、近似直線の勾配を利用したコード化を行なった。また、それから得られるコード列を利用した発話の分類に関して検討を行なった。また、自動的処理を行ない大量データを扱った際の問題点を指摘した。

その結果、今回のコード化を用いた場合、基本周波数パターンのみで70%程度の正解率で発話終了、発話継続、あいづちが判定できる見通しが得られた。また、抑揚情報パターンをコード化する場合、近似直線の勾配の量子化はコード数9程度が妥当であることが分かった。

同様にパワー波形パターンのみを用いた場合も、60%強の正解率が得られた。また、量子化のコード数は9程度が妥当であろうと思われる。

また、コードパターンによる分類で、不正解として扱われる発話を分析した結果、それらの発話の多くが音韻情報により判断できることが分かった。したがって、音韻情報と今回のコード化された抑揚情報をうまく組み合わせることで、発話終了、発話継続の判定が高い精度で行なえると考えられる。

しかし、大量データの分析を行なうためにはすべてを機械的に行えることが望ましい。そこですべての処理を自動化し、データ量を増やしてみたが、あまり良い結果を得ることが出来なかった。その理由として、発話のラベリングのためにはあいづちなどの認識が必要であり、単純に発話の遷移だけからは決定できないということがあげられる。

現在、国内外で対話のタグ付けの検討が行なわれているが、コーパスにタグが付与されていれば、その情報を用いたラベリングにより、大量のデータから、音韻情報・抑揚情報と発話制御との関係の分析が可能になると考えられる。

参考文献

[1] 小磯 花絵 他. 言語的・韻律的情報を利用した発話の終了／継続の予測. 人工知能学会全国大会論文集.pp.407–410. 1996.

[2] 市川 熹. 抑揚情報を利用した対話音声理解モデル構築のための基礎的研究. 重点領域研究報告書.pp.371–378. 1996.

[3] 青野 元子 他. 地図課題コーパス (中間報告). 人工知能学会研究会資料.SIG–SLUD–9402.pp.25–30. 1994.

[4] 市川 熹 他. 対話音声の抑揚の記述. 音響学会講演論文集. 1–P–4.1996.

[5] 佐々木 聡 他. 発話制御におけるコード化された抑揚情報の利用. 音響学会講演論文集.1–3–5.1996.

[6] 小磯 花絵 他. 先行発話断片の終端部分に存在する次発話者に関する言語的・韻律的特徴について. 電子情報通信学会技術研究報告 NLC.pp.25–30. 1996.

[7] 堀内 靖雄 他. 自発的音声対話における話者交替の制御に関わる発話末の統計的・韻律的特徴. 情報処理学会研究報告 SLP.pp.45–50. 1996.