

## 音声の高さ、大きさ、速さ感覚と物理関連量

小林 聡、北澤 茂良

静岡大学 電子科学研究科  
静岡県浜松市城北 3-5-1

{skoba,kitazawa}@cs.inf.shizuoka.ac.jp

あらまし 非言語的特徴を音声の認識や合成に利用するためには、どの程度の変化によって非言語的特徴が意味有るものとして認識されるのかを知る必要が有る。我々は声の高さ、大きさ、速さそして卓立についてその変化のラベル付けを行ない、物理的な量との関連を調べた。その結果、多くのラベルが実際の変化を記述しており、また大きな変化ほど、ラベルが書かれやすく、またそのラベルも実際の変化を記述している割合が高いことが分かった。しかし、大きな変化であっても、必ずしもラベルが書かれているわけではなかった。これにより、非言語的特徴は、単に物理的な量のみで判断されているわけではないことが分かった。

キーワード 書き起こし、非言語的特徴、卓立、音響的变化

## Correlates between Feeling of Hight, Loudness and Tempo of Speech and Physical Features

Satoshi KOBAYASHI, Shigeyoshi KITAZAWA

The Graduate School of Electronic Science and Technology, Shizuoka Univ.  
3-5-1, Johoku, Hamamatsu, Shizuoka

### Abstract

To use paralinguistic features on speech recognition and synthesis, we must know how changes of the features are recognized as meaningful. We transcribed paralinguistic features and changes of sounds in the aspects of hight, loudness, tempo and prominence. As the result, many transcriptions of paralinguistic features represented real changes. And the larger the changes, the more correctly labeled. Even a large change was not always transcribed. The paralinguistic features mean other aspects of utterances, not only physical features.

key words Transcription, Paralinguistic Features, Prominence, Acoustic Difference

## 1 はじめに

対話音声には、言語レベルでの書き起こしでは表現できない非言語情報が含まれる。非言語情報には、ジェスチャーなども含まれるが本研究ではあくまで音声における非言語情報を考える。音声に含まれる非言語情報としては、声質や声の高さ、大きさ、速さなどが有る。本研究においては、声の高さ、大きさ、速さ、および卓立の感覚を調査の対象とした。

声の高さ、大きさ、速さのうち、日本語においては声の高さはアクセントとして音声の言語的な内容の理解にも利用されている。しかし、そのような場合、問題となっているのはどの音節が高く発声されているかなどである。本研究においては、より大域的な変化の認識を対象としている。大域的な特徴として、単に話者の声が高いなどの場合も、1つの非言語的特徴と考えられる。しかし、単に話者の声が高いなどという特徴は、発話内容に関連して何かの情報を伝えるわけではない。そのため非言語的特徴として重要なのは、声が高くなったなどの非言語的特徴の相対的な変化であると考えられる。同様のことが、声の大きさや速さについても言える。

そのような非言語的特徴を、音声認識・音声合成などに利用するためには、さまざまな非言語的特徴のそれぞれがどの程度変化することにより、人間がそれを意味ある変化としてとらえるかを知る必要がある。これまでも、感情や強調（および卓立）の表現による非言語的特徴が調べられてきたが[2, 3]、それらは人工的な状況におけるものであった。しかし、非言語的特徴は、人間がそれと判断することによって成立する。そのため、自然な発話の聴取においてどの程度の音響的变化を人間が意味あるものとして捕えるのかという点が、非言語的特徴を利用する場合には重要である。そこで本実験においては、音声にどのような非言語的特徴が含まれると被験者が判断したかの結果を得るために、非言語的特徴のラベリングを行なった。特に、声が高くなったあるいは低くなったという判断、声が大きくなったあるいは小さくなったという判断、発話速度が速くなったあるいは遅くなったという判断、そしてある部分が卓立されているという判断のラベリングを行なった。

非言語的特徴の判断については、いくつかの疑問点がある。その1つは、ラベルが書かれるのは音響的な変化と無関係ではないと思われるが、そのラベルは信頼性の問題である。つまり、実際の音響的变化を表わすラベルが書かれるものかどうかという問題である。2つめは、実際にどの程度の変化によってラベルが書かれているのかという問題

である。3つめは、ここで我々は、ラベルが書かれるのは音響的な変化と無関係ではないと考えているが、ならば人間が行なうような非言語的特徴の識別を自動的にできるかどうかという問題である。そこで、本実験では、これらの声の高さ、大きさ、速さ、および卓立についてのラベルと音声データにおける音響的变化の比較を行なった。

以降、2節においては、実験条件を述べる。3節では、ラベル内容と音響的变化との一致率を最大にする評価方法について述べる。4節では、被験者間でのラベル内容の一致率および正解率について述べる。5節では、ラベルがつけられた位置などにおける音響的な変化の特徴について述べる。6節では、卓立について、音響的な変化の特徴について述べる。

## 2 実験条件

本実験に用いた音声資料は、ラジオ番組において読み挙げられたニュースおよびそれに続く自発的と思われる発話であり、話者は男性アナウンサー1名、サンプリング周波数は16kHz、長さはおよそ37秒である。なお、この音声資料はスタジオ録音されたものを使用している。また、音声資料には59文節が含まれ、その平均文節長は540 msec.である。

ラベリング作業は、Sparc Station上で音声波形を表示し、グラフィカル・ユーザー・インターフェースを使い、ヘッドフォンによる聴取を通して行なった。ラベルは、非言語的特徴の中でも特に声の高さ、大きさ、発話速度の変化と卓立に注目して行なった。

本実験の被験者は5名である。ラベルはTEI[1]に準拠した記述法を用いた。ただし、被験者中3名はこのような作業の経験がなかった。そこで、3回のトレーニングおよびミーティングを行なった上で、本実験で使用したデータの作成を行なった。

また、ラベルは、原則として文節の開始点に記述するよう指示した。これは、ラベルの評価を容易にするための指示だが、あくまで原則であり、文節の開始点以外にラベルを付けることも認めていた。ただし、すべての文節にラベルを付ける必要は無く、対象となる非言語的特徴が明らかに変化したと思われる部分のみ、かつ必要なラベルのみを付けるよう指示した。更に、卓立も非言語的特徴として重要な役割を担っていると考えられるため本実験の対象とし、卓立であると被験者が感じた部分（始点および終点）のラベル付けも行なった。なお、TEIにおいてはこのような複合的な変化を示すラベルは定義されていないため、独自にラベルを用意した。

### 3 評価方法

声の高さ、大きさ、速さの変化についての感覚は、ある特定の範囲内における最大値、平均値、最少値のいずれかの変化、もしくはそれらの変化の複合的な感覚として感じられているものと仮定した。

本実験では、評価方法として次の式のように dB を用いて評価を行なう。声の高さや発話速度に対して dB を使うのは一般的ではないが、3 者間の比較をしやすくするため、ここでは声の高さ、声の大きさ、発話速度のいずれも dB を用いて評価している。

$$20 * \log(\text{ラベル前の値} / \text{ラベル後の値}) \quad (1)$$

ここで、“ラベル前の値”と“ラベル後の値”としては、ラベルが書かれている位置を中心に、その前後に等しい幅の窓を取り、その窓内での平均値、窓内における最大値、窓内における最小値を考える。ただし、最大値と最小値に関しては、1 点のみの極端な値を避けるため、計測対象の窓内に 100 msec. の小さな窓を取り、その局所的な平均値を考える。この局所的な窓は大きな窓内をステップ幅 10 msec. で移動し、大きな窓内における局所的な平均値として最大値と最小値を計算する。

この結果が正の値であれば、声は高くなり、声は大きくなり、発話速度は速くなるというラベルと、実際の変化が一致することになる。負の値であれば、声は低くなり、声は小さくなり、発話速度は遅くなるというラベルと、実際の変化が一致することになる。なお、以下では、ラベル内容と実際の変化が一致する場合をラベルが正しく記述されていると考えている。

非言語的特徴は、人間がそれととらえることによって成立する。そこで、可能な限り人間の判断に近い結果をもたらす評価方法を求めることにより、つまり、ラベル内容の正解率が最大となる評価方法および窓幅を求めることにより、ラベリング作業の際に被験者が、どの程度の区間を基準にして非言語的特徴を判断しているのかを推定できると考えられる。

なお、この計測において、窓内に現れるポーズ区間、および声の高さと声の大きさについては促音の区間を除いて計算している。また、声の高さについては基本周波数の値を、声の大きさはサンプリング・データの自乗平均を使用した。発話速度 (モーラ数 / 秒) については、区間内に含まれるモーラの継続長の逆数の平均を用いた。なお、窓内にはモーラの断片が現れる場合もあるが、その場合は、窓内に含まれるモーラの割合をモーラ数として使用した。

結果は次の図 1 に声の高さの結果、図 2 に声の大きさの結果、図 3 に速さについての結果を挙げる。図中、“Avr”とは、窓内での平均値による評価、“Max”は窓内における最大値による評価、“Min”は窓内における最小値による評価にもとづいたラベルの正解率である。横軸は、msec. 単位での窓幅を示す。

これらの図において、平均値と最大値による計測の結果では、ほぼ同程度の正解率が得られている。しかし、最少値では他の 2 者よりも低い正解率となっている。この結果から、最小値は声の高さ、大きさ、速さの感覚のいずれに対しても少なくともそれのみではあまり影響を与えていないか、もしくは今回のような、ラベルが書かれた時点を中心としてその前後に等しい幅の窓を取るという方法では最小値に基づいて変化を捕えることはできないと考えられる。

また、全被験者の正解率の平均が最大となる窓幅は、表 1 のようになる。窓内における最少値にもとづく結果は正解率が低いために挙げていない。表 1 中、“平均値”とは窓内での平均値にもとづく正解率であり、“最大値”とは窓内における最大値にもとづく正解率を表わす。なお、表 1 において、下線が付けてあるものを以下では評価用の窓幅として使用している。これは、複数の窓幅が書いてある場合、高さなどのラベルごとに、すべてのラベルに対して dB の自乗平均を得た時に最大値を与える窓幅とした。これは、変化がより大きな値として得られる窓幅であると考えられる。

表 1 から、最大の正解率を挙げている窓幅は 500 msec. 前後もしくは 800 msec. 前後となっている。特に声の高さでは、平均値の場合も最大値の場合も 500 msec. 周辺、発話速度の場合は、平均値の場合も最大値の場合も 800 msec. 周辺において最も高い正解率となっている。

以下では、表 1 に示した、最大の正解率を得た窓幅と評価方法を用いる。つまり、声の高さについては窓幅 450 msec. 内における最大値、大きさについては窓幅 450 msec. 内の平均値、速さについては窓幅 750 msec. 内の最大値を用いて評価を行なう。

表 1: 正解率の最大値と窓幅

	平均値		最大値	
	窓幅 [msec.]	正解率 [%]	窓幅 [msec.]	正解率 [%]
高さ	550	91.8	450, 500, 550	92.5
大きさ	<u>450</u>	88.8	800, 850, 900	80.2
速さ	850, 900, 950	81.8	<u>750</u> , 800	83.7

#### 4 被験者間における非言語情報ラベルの一致

まず、それぞれの被験者ごとの正解率を表 2 に挙げる。

表 2 より、いずれの対象に対しても、全ての被験者を通して 80% 程度以上、多くは 90% 程度の正解率が得られている。これにより、書かれたラベルは被験者によらず声の高さなどの変化を的確にとらえていることが分かる。

次の表 3、4、5 に、被験者間で同じ位置に同じラベルを書いていた場合の正解率などを挙げる。ここで、“Conf.”とは、被験者間で矛盾するラベルが書かれていた場合である。

いずれの場合も、一致する被験者が多いほど正解率も高くなる傾向が見られる。特に、3 名以上の被験者のラベルが一致している割合は全ラベルの 50% 程度となっており、また、ラベル内容の誤りは見られない。

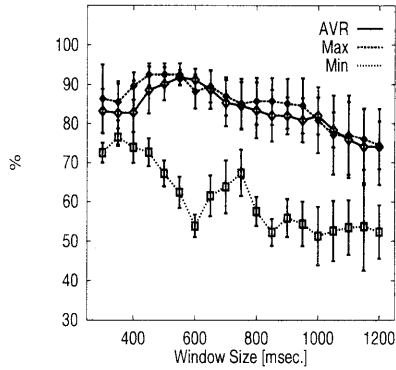


図 1: 窓幅による”声の高さ”ラベルの正解率

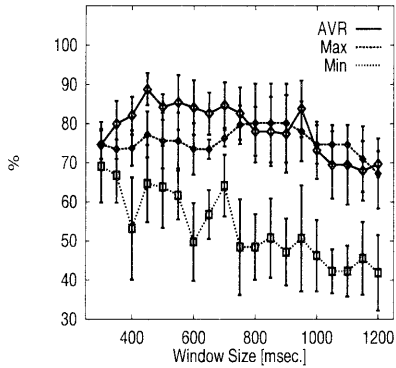


図 2: 窓幅による”声の大きさ”ラベルの正解率

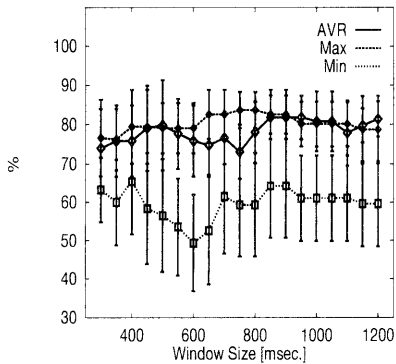


図 3: 窓幅による”速さ”ラベルの正解率

表 2: 被験者ごとの正解率 [% (Correct/Total)]

被験者	高さ	大きさ	速さ
A	94.1 (16/17)	91.7 (22/24)	85.7 (12/14)
B	92.3 (36/39)	88.5 (23/26)	88.5 (23/26)
C	96.4 (27/28)	86.7 (13/15)	78.9 (15/19)
D	89.5 (17/19)	83.3 (10/12)	78.6 (11/14)
E	90.3 (28/31)	93.8 (15/16)	86.7 (13/15)

表 3: 被験者間での”声の高さ”ラベルの一致

一致する被験者数	ラベル数	全ラベルに対する割合 [%]	正解率 [%]
1	19	14.2	68.4
2	30	22.4	93.3
3	18	13.4	100.0
4	56	41.8	100.0
5	5	3.7	100.0
Conf.	6	4.5	66.7

表 4: 被験者間での”声の大きさ”ラベルの一致

一致する被験者数	ラベル数	全ラベルに対する割合 [%]	正解率 [%]
1	25	26.9	84.0
2	18	19.4	88.9
3	27	29.0	100.0
4	4	4.3	100.0
5	10	10.8	100.0
Conf.	9	9.7	55.6

表 5: 被験者間での”速さ”ラベルの一致

一致する被験者数	ラベル数	全ラベルに対する割合 [%]	正解率 [%]
1	15	17.0	66.7
2	22	25.0	63.6
3	21	23.9	100.0
4	8	9.1	100.0
5	20	22.7	100.0
Conf.	2	2.3	50.0

ただし、一致する被験者数が少ないラベルについても、必ずしも誤っているというわけではない。また、被験者間において矛盾する内容のラベルを同じ位置に記述した例はごく少数であり、被験者間におけるラベルの違いの多くは、変化を表わすラベルを書くかどうかの判断の違いであることが分かる。

## 5 音響的变化

本節では、ラベルが書かれた位置などにおいて、どの程度の音響的な変化が起きているのかについて述べる。

ラベルが書かれた位置などにおける変化の平均値と標準偏差を表6に挙げる。ここで、“High”および“Low”はそれぞれ、声の高さならば高くなるおよび低くなるというラベル、声の大きさならば大きくなるおよび小さくなる、速さについては速くなるおよび遅くなるというラベルを表わす。また、“無し”とは、文節の開始点でラベルがつけられていないものを表わす。なお、文節の開始点以外にラベルがつけられないわけではないが、ラベリング位置の原則的な基準点を文節の開始点と被験者に指示したため、ここではラベルがつけられていない文節の開始点のみを、“無し”と考えた。“誤り (High)”とは、付けられたラベルの内容は“High”であったが、実際の変化は“Low”に対応するものを表わす。逆に“誤り (Low)”とは、付けられたラベルの内容は“Low”であったが、実際の変化は“High”に対応するものを表わす。

表 6: 音響的变化の統計的特徴

	変化	個数	平均値 (S.D.)
高さ	High	81	2.7(1.8)
	Low	53	-2.9(2.3)
	無し	167	0.1(2.4)
	誤り (High)	5	-1.0(0.8)
	誤り (Low)	5	1.6(0.4)
大きさ	High	60	5.7(7.3)
	Low	33	-4.5(5.1)
	無し	210	0.4(7.4)
	誤り (High)	6	-5.6(4.5)
	誤り (Low)	4	4.4(4.6)
速さ	High	50	2.5(2.5)
	Low	38	-2.7(3.2)
	無し	200	0.0(3.4)
	誤り (High)	8	-2.2(1.4)
	誤り (Low)	6	3.2(1.9)

以下の図 4から図 6に、声の高さ、大きさ、速さの変化のヒストグラムを挙げる。ここで“HIGH”とは、声の高さならば高く、声の大きさならば大きく、発話速度ならば速くなるというラベルを意味する。ここで“LOW”とは、逆に声の高さならば低く、声の大きさならば小さく、発話速度

ならば遅くなるというラベルを意味する。“Miss”は、誤ったラベルであり、“Not Labeled”は、ラベルが付けられなかった文節開始点における変化である。なお、文節開始点以外に付けられていたラベルの、全ラベルに対する割合は、声の高さの場合は12.7%(17個)、大きさの場合は14.0%(13個)、速さの場合は3.4%(3個)となっている。

声の高さ、大きさ、速さのいずれの場合でも、誤りも広く分布している。ただし、声の高さと速さについては、およそ -3 ~ 3dB の間に誤りの多くが含まれている。また、声の大きさについては、自乗平均にもとづいて dB が得られているが、声の高さや大きさにおけるものとはほぼ対応する、-5 ~ 5dB の間に誤りの多くが含まれている。ただし、それよりも大きな変化においてもラベルの誤りは存在し、また、大きな変化の場合であっても、必ずしもラベルが付けられているわけではないことが分かる。

図 7から 9に、文節の頭およびそれ以外でのラベルが付けられた位置における dB ごとの、ラベルがつけられた割合、およびそのラベルの正解率を挙げる。ここで、“Labeled / Total”とは、ラベルが付けられた割合を表わし、“Correct/Labeled”は、正しいラベルが付けられた割合を表わす。また、“(High)”とは、声が高く、あるいは大きく、速くなったというラベルの場合であり、“(Low)”とは、逆に声が引くく、あるいは小さく、遅くなったというラベルの場合を表わす。この図から、少なくとも、声の高さと、速さについては、大きな変化であればラベルが書かれやすく、かつそのラベルが正しい割合が高くなるという傾向が見られる。

しかし、変化が大きい場合であっても、必ずしもラベルが付けられるわけではない。現時点においては、どのような場合にラベルが付けられ、どのような場合にはラベルが付けられないかは明確ではない。今後、その基準について調査を行ないたい。これらの結果は、人間はここで挙げた非言語的特徴の識別において、単に音響的变化のみで判断しているわけではないことを示している。そのため、音響的な変化のみによって、人間が行なっているような非言語的特徴の認識を行なうのは困難であると考えられる。

## 6 卓立

次に、卓立であるとラベルづけされている場合のピッチなどの変化を示す。卓立とは、ピッチ、声の大きさ、音の長さ、強勢、聞こえのいずれかあるいは組み合わせによって、対象が際だって聞こえるという現象である。これまで、卓立(や強調)は主にピッチの上昇が報告されている[2]。また、音声合成規則として、卓立する部分においてピッチと

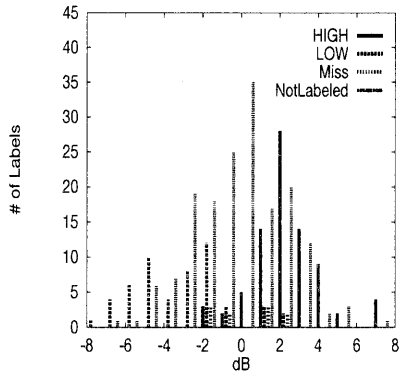


図 4: ピッチ変化のヒストグラム

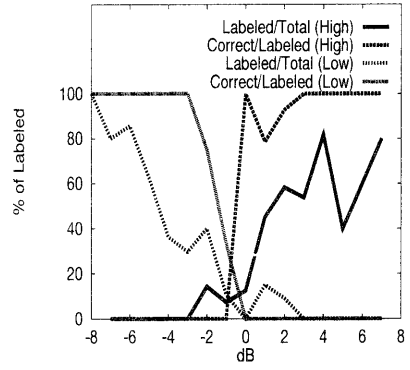


図 7: ピッチの変化とラベル付けされる割合

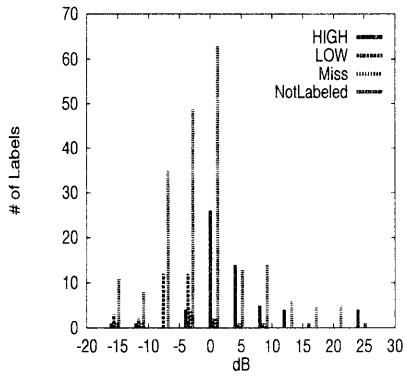


図 5: 大きさの変化のヒストグラム

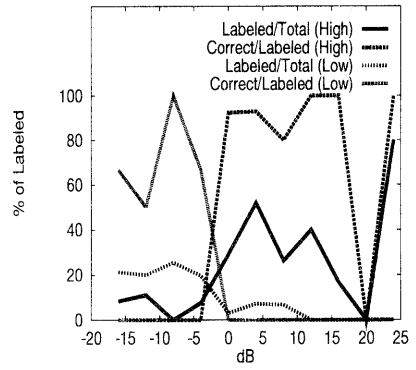


図 8: 大きさの変化とラベル付けされる割合

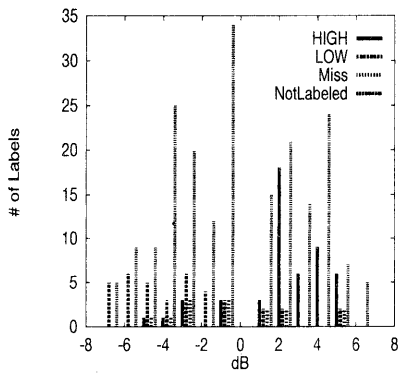


図 6: 速さの変化のヒストグラム

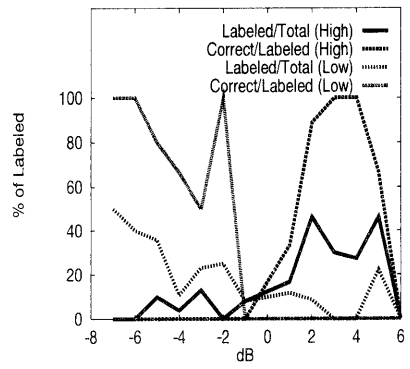


図 9: 速さの変化とラベル付けされる割合

声の大きさの両方を制御することで、卓立と聴取されるとの報告もあるが、その変化のルールは明確ではない[5]。英語においては卓立した母音では高周波が強くなっているという報告が有る[4]。また、発話速度に注目したものとしては、英語の読み上げにおいて卓立の際に発話速度が明確に遅くなるという報告が有る[6]。

しかし、それらは合成音声による1文程度の発話によるものであったり、あるいは英語の読み上げを対象としたものである。卓立は上記のようなピッチ、声の大きさ、発話速度のいずれにおいても変化が認められることや、本実験のラベリングは声の高さ、声の大きさ、発話速度の感覚に注目して行なったことなどから、本実験においては自然な発話において卓立と認識される部分における声の高さ、声の大きさ、発話速度の変化を調べた。次の図10から、12に、卓立の開始点および終了点とラベルづけされた位置における音響的な変化のヒストグラムを挙げる。また、表7に、それぞれの平均値および標準偏差を挙げる。

表 7: 卓立ラベルの変化の統計的特徴

	平均値 (S.D.)	
	開始点	終了点
高さ	3.0(1.7)	-2.4(2.3)
大きさ	2.0(6.8)	-4.0(7.4)
速さ	-1.0(3.7)	1.5(2.7)

卓立の開始点を示すラベルは、全被験者合わせて39個あった。また、終了点においては、音声区間の終了点に1人の被験者が1つ書いており、そこでは変化を計算できないため、計38個が対象となっている。

表7および図10~12から、卓立の開始点においては声は高く、大きく、また遅くなる傾向が読み取れる。また卓立の終了点では、開始点とは逆に声は低く、小さく、また遅くなる傾向が読み取れる。

そこで、卓立ラベルが付けられた点における変化に対し、検定を行なった。その結果を表8および9に示す。その結果、いずれの場合も仮説  $H_0$  は帰却され、表7から見られる変化の傾向は、統計的に明らかなものと言える。ただし、図10に示す声の高さの場合にはかなり明確に変化の傾向に沿っているが、図11に示す声の大きさや、図12に示す発話速度については、図からは、声の高さほどはっきりとした傾向は見られない。このように、卓立の開始点において声が高く、大きく、また発話速度は遅くなるという傾向や、あるいは卓立の終了点において、声が低く、小さく、また発話速度は速くなるという傾向が見られるものの、全ての卓立がこの傾向に沿って変化しているわけではない。

これらの変化は、平均して声の高さの場合は41%、大き

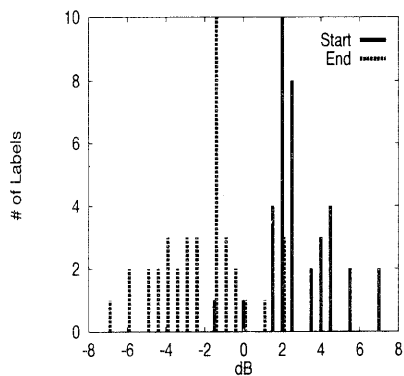


図 10: 卓立ラベルでの”声の高さ”の変化

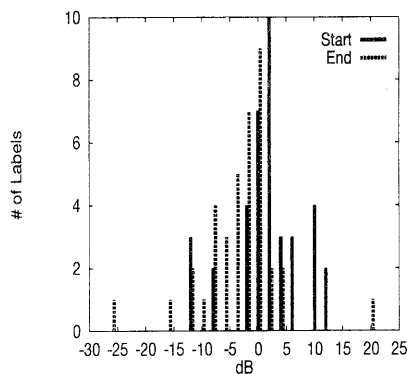


図 11: 卓立ラベルでの”声の大きさ”の変化

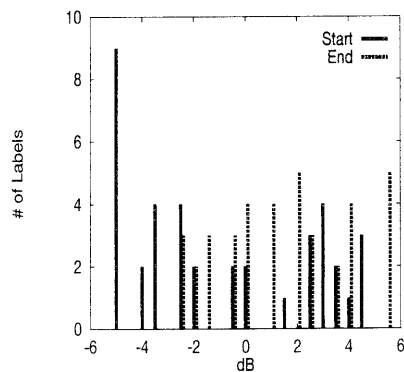


図 12: 卓立ラベルでの”速さ”の変化

表 8: 卓立開始点における変化の検定

開始点	仮説	結果
高さ:	$H_0: \mu = 0, H_1: \mu > 0$	有意水準 0.01 で帰却
大きさ:	$H_0: \mu = 0, H_1: \mu > 0$	有意水準 0.05 で帰却
速さ:	$H_0: \mu = 0, H_1: \mu < 0$	有意水準 0.05 で帰却

表 9: 卓立終了点における変化の検定

終了点	仮説	結果
高さ:	$H_0: \mu = 0, H_1: \mu < 0$	有意水準 0.01 で帰却
大きさ:	$H_0: \mu = 0, H_1: \mu < 0$	有意水準 0.01 で帰却
速さ:	$H_0: \mu = 0, H_1: \mu > 0$	有意水準 0.01 で帰却

さであれば 26%、速さについては 11% の変化で卓立の開始とみなされていることになる。ただし、声の高さと速さについては区間内の最大値の比である。合成音を用いたピッチの変化についてはおよそ 40% 程度の上昇によって強調と聴取されたと報告されているが [2]、本実験での結果も、同程度の変化で卓立ととらえられている。

## 7 まとめ

本実験では被験者に、聴取にもとづいて、声の高さ、大きさ、速さについてその変化をラベリングし、あるいは卓立においてその開始点と終了点をラベリングし、そのラベルと実際の音響的な変化の比較を行なった。

その結果、非言語的特徴ラベルの前後それぞれ 500 msec. もしくは 800 msec. 程度の窓によって、最も高い正解率が得られた。これにより、人間が非言語的特徴として声の高さ、大きさ、速さを認識する場合にも、この程度の窓によって判断していると推定される。

複数の作業者が同じ位置に同じ内容のラベルを記述している場合には、そのラベルの信頼性は高かった。特に 3 名以上が同一の位置に同一の内容を書いている場合、今回用いた評価方法においては全てのラベルが正しいものであった。そうでない場合でも一概にラベルの信頼性が低いとは言えない。また、被験者間におけるラベル内容の違いの多くは、変化についてのラベルを書くかどうかの判断の違いであった。

ラベルづけされた部分における実際の音響的な変化については、大きな変化ほど、ラベルが書かれやすく、またその内容も正しい割合が高くなっていった。しかし、仮に大きな変化であっても、必ずしもラベルが付けられているわけではなかった。

また、卓立については開始点では声は高く、大きくなり、発話速度は遅くなる傾向が見られた。終了点については、

声は低く、小さく、発話速度は速くなる傾向が見られた。

非言語的特徴は、人間がそれをどのように判断するかが現状における唯一の基準である。声の高さ、大きさ、速さなどは数量的にとらえることが可能である。実際、今回の評価方法によって高いラベルの正解率が得られたことから、音響的な変化も判断に関係していることは明らかである。しかし、変化が大きければ必ずラベルがつけられているというわけではなかった。これにより、人間は音響的特徴の変化のみによって非言語的特徴をとらえているわけではないことが分かる。そのため、人間が行なうような非言語的特徴の識別は、音響的な変化のみによって行なうことは困難であると思われる。

## 参考文献

- [1] C.M.Sperberg-McQueen, Lou Burnard edited: "Base Tag Set for Transcription of Spoken Texts", TEI P3 chapter 11, Text Encoding Initiative, Chicago, Oxford, 1994.
- [2] 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", 信学論 J70-A, pp. 816-821 (1987).
- [3] 小林 豊, 駒沢 健一郎, 村家 志奈子, 新美 康永: "音声に含まれている感情表現について", 日本音響学会平成 7 年度秋期研究発表会講演論文集, 1-1-2, pp.201-202(1995).
- [4] Nick Campbell, Mary Beckman: "Stress, Loudness, and Spectral Tilt", 日本音響学会平成 7 年度春期研究発表会講演論文集, 3-4-3, pp.279-280(1995).
- [5] 鈴木 司, 生方 正夫, 有泉 均: "規則音声合成における強調規則について", 日本音響学会平成 8 年度春期研究発表会講演論文集, 1-4-12, pp.241-242(1996).
- [6] Claire Gerard, Delphine Dahan: "Durational Variations in Speech and Didactic Accent During Reading", Speech Communication, Vol. 16, pp.293-311, 1995.