音声言語情報処理 22－9
（1998. 7. 25）

韻律的特徴と周辺言語的特徴

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

我々は、韻律的特徴と周辺言語的な特徴を比較し、それぞれが表現される背景とともに表現の方法もまた異なると仮定した。周辺言語的特徴を音声の認識や合成に利用するためには、音響的な特徴などがどの程度変化することによって周辺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。我々は自由な発話に対して、声の高さ、大きさ、速さ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。

書き起こし、周辺言語的特徴、音響的変化、韻律的特徴

Measuring Prosodic Features or Paralinguistic Features?

Satoshi KOBAYASHI, Shigeyoshi KITAZAWA

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

We compared prosodic features and paralinguistic features assuming that prosodic features and paralinguistic features are represented in different way on different backgrounds. To use paralinguistic features on speech recognition and synthesis, we must know how changes of the features are recognized as meaningful. We attached labels of paralinguistic features in the aspects of height, loudness and tempo on spontaneous speech. Then we investigated the relation of labels and physical quantity. As the result, many labels had described physical changes. And when it is a bigger change, the rate to which a label was written was high. Moreover labels for bigger changes also had the higher rate of carrying out description corresponding to actual changes.

Transcription, Paralinguistic Features, Acoustic Change, Prosodic Features

1 はじめに

これまで、音声言語に付随する音響的な特徴の分析は、主に文などの短い発話を対象としていた。そのため、音響的特徴としては意味的な対立を表わす韻律が重要であった。しかし、自然な対話などにおける発話や、より長い発話に対しては、韻律のみではなく、周辺言語的情報も重要であると考えられる。

本報告においては、まず韻律的特徴について考察を行ない、それにより周辺言語的情報を表現する音響的な特徴について仮定を設ける。また音声に対してラベル付けを行ない、設けた仮定にもとづいて、周辺言語的情報を表現する音響的な特徴についての聴き手が行なう判断の傾向について述べる。

2 周辺言語としての声の高さ、大きさ、発話速度

2.1 周辺言語と韻律

韻律は元々ヨーロッパでの詩のリズム研究を指していた。音韻論の中で韻律的特徴・韻律素性として使用されるようになり、その後、韻律という言葉が非常に広く解釈される場合には韻律が全ての音声的特徴を包含するとする定義を与えることがあるので、韻律研究が全ての音声現象の研究を包含しているかのような解釈がなされることがる。パラ言語あるいは非言語伝達の研究と韻律研究との対比で位置付けが必要である。

言語学研究は言語そのものに関わる研究と社会・心理など他領域と関わる研究とに大きく 2つに分ける事ができる。前者は音声学、音韻論、形態論、統語論、意味論、語用論で、後者は社会言語学、心理言語学、神経言語学、応用言語学、コンピュータ言語学、文体論、言語哲学、認知言語学、人類言語学などである。アメリカ構造言語学の立場から、G. L. Trager(1906 ~ 1992) は言語学全ての領域を指す用語として大言語学 (macrolinguistics) を用いて、大言語学を前段言語学 (prelinguistics)・小言語学 (microlinguistics)・後段言語学 (metalinguistics) の三つの分野に分けた。前段言語学は音声の物理的・生理的研究を含み、調音音声学、音響音声学などの分野を含む。小言語学は言語学の中心部分で、音韻論、形態論、統語論などの分野が含まれる。後段言語学には、文より大きい単位の言語研究が含まれ、文体論、意味論、パラ言語学 (paralinguistics) などの分野で扱われる。後段言語学の中でもパラ言語学、社会言語学などは 1960 年以降急速に発達してきた。

人間が情報を伝達しようとする時、その言語行動に伴って起こる非言語的行動をパラ言語 (paralangugage) と呼び、それを研究する分野をパラ言語学・周辺言語学という。一般に、韻律素性 (prosodic feature) に含まれない声の質 (かすれ声、キーキー声など)、高さ (頭声・胸声など)、音量 (大声・小声など)、話し方 (流れるような、途切れがちの) などの声の調子が扱われる。

パラ言語は、言語行動とともに、話し手の性別・年齢・性格・健康状態・感情、話題、聞き手に対する態度など、さまざまな情報を伝える。身ぶりもパラ言語に加えられるがここでは除外して考える。

ここで、いわゆる韻律的特徴・韻律素性 (prosodic feature) は音韻論における概念である。音韻論は各言語における意味の区別にかかわる音声特徴を解明することであり、音声学において細かく記録・分析された言語音が、ある言語でどのような、意味を区別する音の体系を作っているかを明らかにする。これには音素や弁別素性の対立を明らかにする。一方、音を分節素 (segment) と超分節素 (suprasegment) に分けて、分節素音韻論と超分節素音韻論として論じられる事もある。

韻律的特徴・韻律素性とは、発話における分節音以外の音声的特徴すべてを指す場合と、超分節的な現象のうち、音の高さ (pitch)・強勢 (stress)・速さ (tempo)・リズム (rhythm) などを指していう場合とがある。英国の言語学者 J. R. Firth が提唱した韻律素分析では分節音以外の音声的特徴が文法的関係も含めて、必要に応じてできるかぎり取り扱われる。アメリカ音韻論では超分節素として強勢、語調そして長さを韻律素としている。Bloomfield は音素と、韻律素的単位を一次音素と二次音素と呼び、後継者達は分節音素と超分節音素と呼んだ。超分節的音素は強勢 (stress)、声調 (tone)、音調 (intonation)、そして長さ (length) を含む。これらは個々の分節 (音節) 内での変動として現れる場合と音節間で強弱・高低・長短の対立として分節を越えて現れる事もある。いずれも音韻論の枠内での意味的対立を表しているので、Trager が 1958 年に提唱した voice qualifiers, vocal characteristics, vocal qualifiers, vocal segregates としてのパラ言語とは異なるものである。

これらの、意味的対立を表現する韻律的特徴に対して、周辺言語的情報としては、話者の意図などを表現するために、話者によって意識的に変更される音響的な特徴であると考えられる。だが、音韻論的に意味的対立を表現する場合に重要なのはおもに語や句の中のどこで変化するか、あるいはどのような変化をするのかということであると考えられる。この点において、韻律においては比較的短い時間的な範囲においての変化が重要であると考えられる。それに対して我々は、周辺言語的情報を表現する場合においては、語や句内での変化が重要な韻律に比較して、よりも広い時間的な範囲において変化が持続することが重要であると考えた。

これまで、対象として文程度の長さの発話を使っていたことなどにより、音声言語における言語的情報、つまり狭義の言語符号によって表現可能なもの以外の情報としてはおもに韻律的特徴の研究がなされてきた。しかし、より自然かつ長い発話や対話を対象とする場合には、話

者の意図などを表現すると考えられる周辺言語的情報の分析が必要と考えられる。

2.2 周辺言語の判定

2.2.1 判定方法

周辺言語的情報は話者がそのような特徴や情報を音響的に表現して発話すると同時に、聞き手がそれらの特徴をとらえ、周辺言語的情報が含まれていると判断することによってはじめて成立すると考えられる。

ここで、音響的な変化によって周辺言語的情報が表現され、また聴き手がそれと判断すると考えると、聴き手はその位置や範囲において明らかな変化が有ると判断しているものと考えられる。

そこで、聴き手には、音声資料に対して明確に変化していると考えられる位置にラベル付けなどにより示してもらうことにより、周辺言語的情報を含むような部位の特定を行なうことが可能であると考えられる。

2.2.2 音声資料

本実験では、自由対話を収録し、その一部を音声資料として用いた。各対話は、それぞれ3名の話者によるもので、計3回の収録を行なった。それぞれの対話は約60分間行なっている。参加した話者は、計5名 (全員 男性) である。なお、この対話は、特にテーマなどは与えずに、自由に行なってもらった。

収録は無響室で行なった。各話者に指向性マイクを割当て、また全体の音声を収録するための無指向性マイクを使用し、計4チャンネルでDATを2台を用いて行なった。

このような対話から、話者3名に対して、20数秒程度の発話をそれぞれ5発話づつ、16kHz でサンプリングしたものを、本実験のための音声資料として使用した。これらの発話を抜き出す基準は、他の話者のオーバーラップや、体を動かすことなどによって発生するノイズ、咳払いなどの非言語音などが少なく、発話がはっきりと聞きとれる程度の大きさの声でなされており、かつ1発話が長過ぎないものとした。

なお、各作業者が対象とした音声は、それぞれ13発話、計約5分40秒程度となっている。

2.2.3 ラベル付け作業

ラベル付け作業は、Sparc Station 上で音声波形を表示し、GUIによる環境で、ヘッドフォンによる聴取を通して行なった。

韻律に対するラベル付けの方法としては、ToBIがある[1]。しかし、ToBIは、語や句内でのピッチ変動を記述するためのものである。また、そのようなピッチ変化の特徴やパターンに基づいて記述形式が規定されているため、周辺言語的情報を担うような音響的変化を記述するためには適していないと考えられる。

そこで、ラベルはTEI[2]に準拠したものを用い、ラベル付けは、周辺言語的情報の中でも特に声の高さ、大きさ、発話速度の変化 に注目して行なった。

本実験の作業者は3名である。ただし、作業者中2名はこのような作業の経験が無かった。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。

ラベル付けの対象としては、文節や句、節などに渡って声の高さ、大きさ、速さが明確に変化していると思われる部分のみに、かつ必要なラベルのみを付けるよう指示を行なった。また、1音節程度が極端に高いピッチもしくは大きな声で発話されていることを示す別のラベルを容易し、局所的な変化とより広域的な変化とは別のものとするよう指示した。

なお、ラベルは、原則として文節の開始点に付与するよう指示した。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。

2.3 被験者間 (内) でのラベルの一致

表1から3に、作業者間で同じ位置に同じラベルを書いていた場合の割合などを挙げる。ここで、"Conf."とは作業者間で矛盾するラベルが書かれていた場合である。

声の高さ、大きさ、発話速度のいずれの場合も、2名以上の作業者のラベルが一致している割合は全ラベルの40% 程度となっている。それに対し、作業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多くとも 10% 程度と少数である。

これらのことから、作業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。

表 1: 作業者間での"高さ"ラベル一致

	一致する作業者数	ラベル数	全ラベルに対する割合 [%] (3人 + 2人 [%])	正解率 [%]
3人	1	159	45.7(27.9)	65.6
	2	130	37.4(22.8)	76.6
	3	12	3.4(2.1)	75.0
	Cnf.	47	13.5(8.3)	51.1
2人	1	125	56.6(22.0)	69.9
	2	78	35.3(13.7)	84.6
	Cnf.	18	8.1(3.2)	50.0

表 2: 作業者間での"大きさ"ラベル一致

	一致する作業者数	ラベル数	全ラベルに対する割合 [%] (3人 + 2人 [%])	正解率 [%]
3人	1	122	41.9(27.5)	74.6
	2	90	30.9(20.3)	86.4
	3	63	21.7(14.2)	100.0
	Cnf.	16	5.5(3.6)	50.0
2人	1	88	57.9(19.9)	73.6
	2	60	39.5(13.6)	100.0
	Cnf.	4	2.7(0.9)	50.0

表 3: 作業者間での"速さ"ラベル一致

	一致する作業者数	ラベル数	全ラベルに対する割合 [%] (3人 + 2人 [%])	正解率 [%]
3人	1	118	39.6(25.8)	69.2
	2	86	28.9(18.8)	83.3
	3	54	18.1(11.8)	94.4
	Cnf.	40	13.4(8.7)	52.5
2人	1	100	62.5(21.8)	68.7
	2	48	30.0(10.5)	87.0
	Cnf.	12	7.5(2.7)	50.0

3 音響的変化の特徴

3.1 変化の計測のしかた

聴き手が、音響的変化に意味が有るととらると仮定した場合、どのような方法でその音響的変化の大きさを計測するのかが問題になる。F0 についてであれば、藤崎モデルにおけるフレーズ指令の大きさなどでの比較も可能と思われる。しかし、周辺言語的情報は、ある程度の長さの区間において表現されると仮定すると、局所的なパラメーターによる比較は適当ではない。

そこで、今回の実験においては声の高さ、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により作業者が感じているものと仮定した。この変化を、本実験では、次の式のように dB を用いて評価を行なう。

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

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 (大きな窓) を取り、その窓内での最大値、平均値、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少値を計算する。

ここで、得られた dB の値が正の値であれば、声は高く、大きくなり、また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とみなした。逆に dB が負の値であれば、声は低く、小さくなり、また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とみなした。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。またラベルの内容と、変化が食い違う場合を、ラベルの誤りと考えた。ただし、ラベルが書かれていない場合については、正解等は考えない。

表 4: 採用した窓幅と、方法

	窓幅 [msec.]	方法	正解率 [%]
高さ	350 msec.	平均	68.9 %
大きさ	500 msec.	平均	82.1 %
速さ	400 msec.	平均	74.3 %

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を、声の大きさは rms を使用した。発話速度 (モーラ / 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。この発話速度は、音節ラベルを作成し、それに基づいて計測している。なお、発話速度を計測するための窓内にはモーラの断片が現れる場合もあるが、その場合は該当モーラの全長に対して窓内に現われる部分の割合を、モーラ数として使用した。

これにより、図 1 から 3、および表 4に示すような窓幅などで最大の正解率が得られた。以後、この窓幅などを用いて評価を行なう。

3.2 変化の特徴

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 5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、発話速度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、"NotLabeled"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付けられないわけではないが、ラベル付けする位置の原則的な基準点を文節の開始点と作業者に指示したため、ここではラベルが付けられていない文節の開始点のみを、"NotLabeled"と考えた。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、声の高さで 7.9%、大きさで 7.4%、発話速度は 7.6% であった。このように、文節開始点以外に付けられたラベルは少数であった。"Miss"とは、付けられたラベルと実際の変化が異なるもの、例えばラベル内容は "HIGH" であったが、実際の変化は "LOW" に対応するものであった場合などを現わす。

図 4 から図 6 に、声の高さ、大きさ、速さの変化の

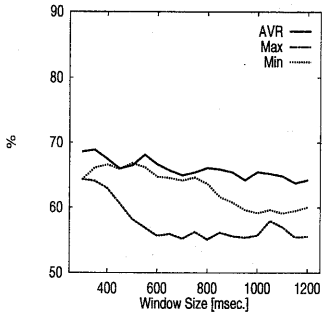

図 1: 窓幅によるラベルの正解率 (声の高さ)

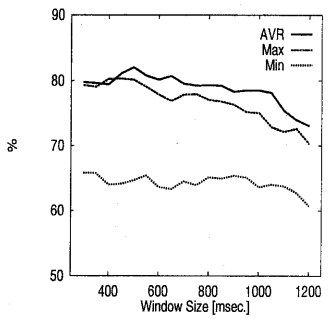

図 2: 窓幅によるラベルの正解率 (声の大きさ)

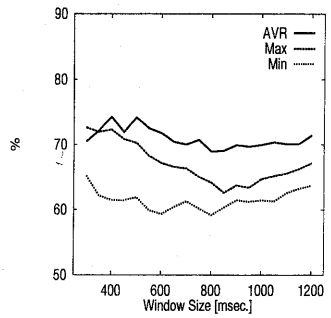

図 3: 窓幅によるラベルの正解率 (発話速度)

ヒストグラムを挙げる。声の高さ、大きさ、速さのいずれの場合でも、誤りおよびラベルが付けられなかったものが 0dB 付近で多くなっている。

図 7 から 図 9 に、文節の開始位置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 を横軸にとり、その変化の大きさで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、 "Lbld / Total" とは、ラベルが付けられた割合を表わし、 "Crrct / Lbld" は、正しいラベルが付けられた割合を表わす。これらの図から、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のかは明確ではない。これらの結果は、人間はここで挙げた周辺言語的情報の識別において、単に音響的変化のみで判断しているわけではないことを示していると考えられる。

表 5: ラベル前後での変化の大きさの平均 (S.D.)

	高さ	大きさ	速さ
HIGH	1.57(1.99)	4.55(4.77)	1.94(2.76)
LOW	-0.52(1.74)	-3.50(4.53)	-0.91(3.41)
NotLabeled	0.05(1.83)	-0.33(4.34)	0.19(2.72)
Miss	0.36(1.53)	0.45(3.77)	0.23(2.68)

4 考察

音声資料に対して付けられたラベルの作業者間での一致率は、声の高さ、大きさ、発話速度のいずれについてもほぼ 40% 程度であった。これらから、同程度の変化を聴取したとしても、その変化が周辺言語的情報を担うのに十分なものであると判断するかどうかは聴取者によって異なると考えられる。また、周辺言語的情報は、仮に発話者がそれを含めて発話を行なったとしても、聴き手や場合により、周辺言語的情報であるととらえられるとは限らないということを示している。

韻律のラベル付けの場合であれば、作業者間での一致率として、約 80% という割合が報告されている [1]。これは、ラベル付けの対象や環境、集計方法などが異なっているため、単純には比較できない。

だが、このように被験者間におけるラベルの一致率が異なった理由として、以下のように考えられる。

第 1 に、 F0 を示すグラフを示してのラベル付けか否かという、環境の違いが影響していると考えられる。

第 2 に、韻律としてのピッチ変化の場合、アクセントなどと関連して多くは判別しやすいと考えられる。対して、周辺言語的情報としては、現時点においては単に聴き手の主観に頼るしかない。このため、例えば F0 のグラフを示したとしても意味はあまり無いと考えられる。

ラベルの正解率において、図 1 から 3、および表 4より、声の高さ、大きさ、発話速度のいずれも、およそ 400 msec. 付近の窓幅において最も高い正解率を得られていた。これに関しては、今回対象としている音声資料の平均文節長が、 423msec. であることから、文節開始点をラベル付けを行なう位置の基準にしたことの影響も考えられる。

しかし、図 4 から 9 に見られたように、音声資料に含まれる文節数に対して、付けられたラベル数は少数であるため、単に文節毎にラベルを付けたことにより、このような結果になったわけではない。これが文節という単位と何らかの関係が有るのかどうかについては、今後の検討が必要である。

また、図 1 から 3 に見られるように、今回の評価においてほぼ全体を通して平均値による比較が、最大値や最小値による評価よりも良い結果であった。これにより、平均文節長に近い時間幅内における 平均的な値が、周辺言語的情報の判断に影響していると考えられる。この点において、周辺言語的情報の判定は実際にある程度の時間幅内の変化によると考えられる。

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かった。しかし、大きな変化の場合にはラベルが付けられる割合も増加しており、プロミネンスの判断において報告されている傾向と同様の傾向となっている [3]。 この傾向は声の高さのみでなく、声の大き

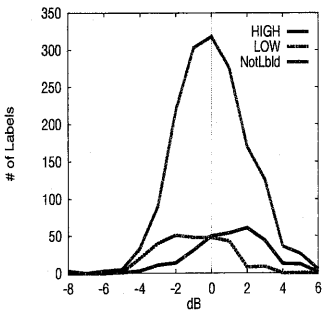

図 4: ラベル前後の変化 (声の高さ)

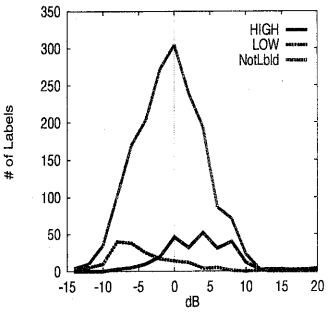

図 5: ラベル前後の変化 (声の大きさ)

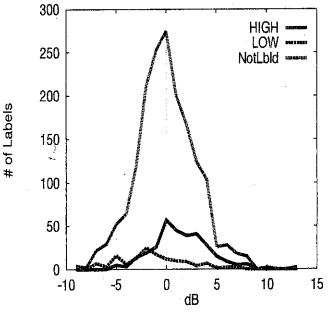

図 6: ラベル前後の変化 (発話速度)

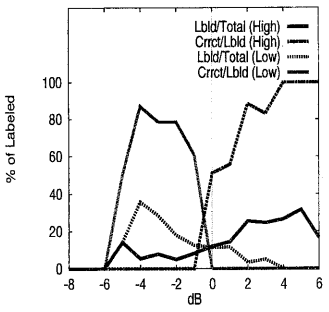

図 7: 変化の大きさとラベル付けされる割合 (声の高さ)

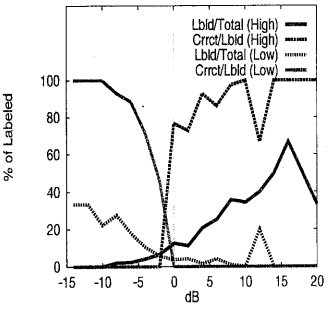

図 8: 変化の大きさとラベル付けされる割合 (声の大きさ)

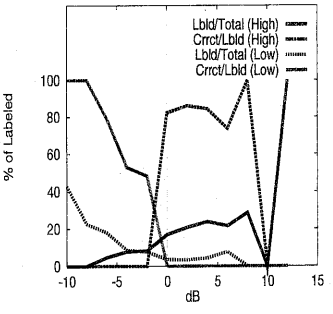

図 9: 変化の大きさとラベル付けされる割合 (発話速度)

さおよび発話速度についても同様に見られた。また、単純に変化の大きさのみでは判断できないという点も同様であった。

変化の大きさ以外の要因としては、出身地など被験者の属する様々なグループの差異など多くの要因考えられる [4]。また、F0 やポーズなどの韻律的特徴が、発話の意味的および構文的要因と関係すると言われているが [5, 6, 7]、周辺言語的情報の判断にもそれらが影響を与えていることが考えられる。

5 まとめ

今回、韻律と周辺言語的情報についての考察を行ない、周辺言語的情報は韻律に比べてより長い時間的範囲内において変化が持続することによって表現されるという仮定を行なった。

また、周辺言語的情報を受けとる場合には、聴き手はそこに明確な変化が有ると判断すると想定し、声の高さ、大きさ、発話速度の変化に対してラベル付けを行なった。そのラベルを元に、どの程度の大きさの変化によって人間がそれを意味の有る変化であると判断するのかについて調査を行なった。その結果、聴き手が意味の有る変化であると判断する変化の大きさの傾向が得られた。

今後、周辺言語的情報が含まれるかどうかについての判断に関して、より精密な分析を行なうとともに、発話の意味的および構文的要因による影響についても調査を行なっていきたい。

参考文献

[1] Kim Silverman, Mary Beckman, John Pirelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, Julia Hirschberg: "TOBI: A Standard for Labeling English Prosody", Proc. of ICSLP, Vol.2, pp. 867–870, 1992.

[2] 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.

[3] Jacques Terken: "Variation of Accent Prominence within the Phrase": Models and Spontaneous Speech Data", Computing Prosody chap. 8, pp. 95-116, Springer (1996).

[4] 武田 晶一, 横里 恵, 海老 義人, 鈴木 修平: "座談会及び落語における日本語会話音声の韻律的特徴の解析", 音響学会誌 Vol. 54, No. 3, pp.199-206 (1998)

[5] 山下 洋一, 作田 瑞, 溝口 理一郎: "対話音声合成のための2段階予測に基づく韻律規則の生成と評価", 音響学会誌 Vol. 53, No. 2, pp.103-109 (1997).

[6] 川端 弘道, 広瀬 啓吉: "対話音声の韻律的特徴の定量的分析による韻律規則の作成", 音講論 平成9年 春秋, 1-7-4, pp.207-208 (1997).

[7] 安藤 ハル, 額賀 信尾, 北原 義典: "文の多義性解消におけるピッチとポーズの関係について", 音講論 平成 10 年 春, 2-P-1, pp.295-296 (1998).