4. プラグマティック・イントネーション: 韻律情報の機能的役割

ニック キャンベル

要旨

従来の研究において既に、文法と韻律(イントネーション)には明瞭な関係のあることが明らかにされてきた。しかしながらこれまでの研究では、それらは発話文の構文的意味理解を対象としており、話者の意図のレベルにおける意味理解という段階には及ばなかった。そこで本章では、韻律情報の機能的アプローチとして有効なPragmatic Intonation理論という観点に基づいた多層的韻律分析により、新しい意味認識手法を提案する。

1. 機能的理論からみた韻律

ここで提案するプラグマティック・イントネーション理論(以下、PI理論という)は音声言語における意味認識手法の機能的アプローチを指し、発話行為が影響を及ぼす聞き手の反応まで含む。韻律の役割は多層的であり、PI理論は、韻律とそれが表わす情報(意味)の種類を分類したイントネーションの階層的構造である。

表1で示すように韻律情報は7階層からなり、その最下位のもの(Pro.L0)は音韻的層、最上位のもの(Pro.L7)は話者的層である。この2層だけは韻律以外の要素と複雑に関係しており、前者は音声の調音的特徴(Seg)と、後者は話者の生理学的特徴(Ind)と切り離して考えることは困難である。

＜表1　意味層と韻律の関係＞

level	example	function(機能)	realisation(実現)
Seg/Pro.L0	おじさん／お爺さん	segmental(調音的)	phone(音韻特徴)
Pro.L1	はし(端／箸／橋)	lexical(辞書的)	accent(アクセント)
Pro.L2	(年輩の男)と女／年輩の(男と女)	syntactic(統語論的)	chunking(韻律区切り)
Pro.L3	田中さんの本／田中さんの本	semantic(意味論的)	prominence(強調)
Pro.L4	そうですね／そうですね？	attitude(態度)	speech act(発話行為)
Pro.L5	「嬉しいわー」	emotion(感情)	interpersonal(信頼性)
Pro.L6	「すきですね」	sincerity(誠実性)	commitment(人間関係)
Pro.L7/Ind	「(電話で：)もしもし」	individuality(個人性)	voice quality(声質)

まず、それぞれの階層について簡単に述べる。Pro.L0は、音韻情報識別レベルであり、/hashi(橋)/と/kawa(川)/の意味の違いの識別のように、音韻記号列で記述されるため原則として韻律情報を必要としないが、/ojisaN(おじさん)/と/ojiHsaN(お爺さん)/あるいは/eki(駅)/と/eHki(英気)/などのように母音の引き伸ばしという韻律的特徴での区別を必要とする場合もある。ただしここで、/H/は引き音を、/N/は撥音を表わす。

Pro.L1は、単語アクセントのレベルであり、/hashi(端)/(平板型)と/hashi(箸)/(頭高型)と/hashi(橋)/(尾高型)等の同音語をアクセントによって識別する。Pro.L0およびPro.L1は韻律の辞書的レベルと見ることができる。

Pro.L2は構文のレベルである。特に同一単語列からなり、係り受け関係に複数の可能性がある文の区別に有効で、統語論的構造で意味を示す。例えば、/年輩の男と女/の場合の「(年輩の男)と女」と「年輩の(男と女)」の曖昧性を解消することができる。

Pro.L3は、フォーカスのレベルである。例えば、「田中さんの本です」という文は、一つの文でありながら、それを問いの答えと考えた場合、「だれの本ですか」という問いと「田中さんの辞書ですか」という問が想定できる。従来の研究において既に、これらの各要因が韻律(イントネーション)に与える影響については明らかにされてきた[1-4]。

Pro.L4は、話者意図の違いを示す。例えば、「そうですか」は言い方によって、賛成、反対、躊躇の意味を伝える。これらの違いは通常発話アクトとして分類されるが、これについては6.で詳述する。これより上位のレベルとして、話者の心的状態や発話の信憑性あるいは信頼性など心理的領域までを含むものである。ただし、Pro.L5-Pro.L7は理論的に区別できるが、自動処理は現時点では困難なため、別の機会に譲ることとする。

先にも述べたとおり、これまでの研究は統語論的あるいは意味論的アプローチが中心的課題であったが、現在音声翻訳の実現といった自然対話における自動音声情報処理の必要などを背景に、さらに高度な意味理解の機械処理システムの構築が求められ、そのアプローチの展開には新たな進展が見られる。

2. 韻律特徴の計測と表現：ピッチ、音韻時間長、パワー

韻律情報の機能について述べる前にまず韻律特徴をどのように計測し表現するかを述べる必要がある。音声言語には様々な韻律情報が含まれているが、主なものとしては、(1)声の高さ、すなわち基本周波数(通常、F_0と略記する)、(2)声の大きさ、すなわち音声パワー(または単にパワー)、(3)それぞれの音の長さ、すなわち音韻時間長などがある。この3つの物理量は、音声波形から容易に計測することができるが、求められた実測値をそのまま比較することはできず、これらを正規化した後、比較する必要がある。以下ではまず、各韻律特徴の正規化方法について説明する。

2.1 基本周波数の正規化

一般に、男性の声よりも女性の声の方が、また大人の声よりも子どもの声の方がF_0は高いが、同性同年齢でも一人一人違った声の高さを持っている。このようなピッチレンジの相違は、発声内容に依存する

イントネーションの違いとも別のものであるが、人間はこれらの相違も容易に聞き分けることができる。

人間は自然のうちに声の性質の個別性を知覚的に正規化するのであるが、この過程を機械が行なう場合、生データから基本のイントネーションパラメータを抽出する必要がある。話者による個別性を除くために、正規化を行なう。つまり、ピッチの高低は楽音の音程と同様にF_0の対数で決まるので、まずF_0の対数をとる。次に、話者のピッチレンジの個別性を減少させるために、それぞれの話者の平均を減算し、さらに標準偏差で割ることによってピッチレンジの幅を正規化する。このようにして求められた値をF_0の「Zスコア」と呼ぶ。

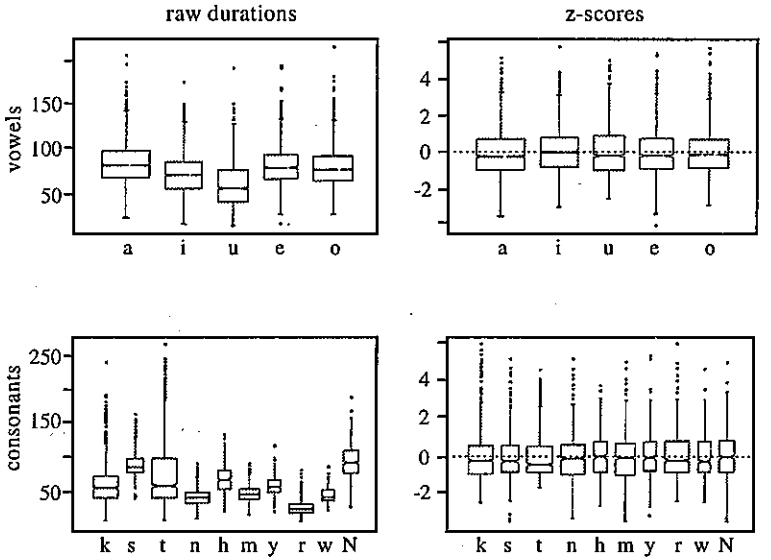
この変換を用いることによって、次の3つの情報が音声波形から抽出される。すなわち、話者の平均ピッチ、文のピッチレンジの幅、話者に依存しないF_0 contourである。話者の平均ピッチそのものは韻律情報としての価値が少ないのに対して、ピッチレンジは発話スタイルと相関があることが報告されている[5]。一方、正規化されたF_0 contourは、話者依存性が取り除かれ、平均のF_0値が0として表わされる。また、F_0値が正規分布を示すと仮定すると平均値より標準偏差分だけ高いF_0値が1に、反対に標準偏差分だけ低いF_0値が-1として表わされる。

2.2 音韻時間長の正規化

音韻時間長の違いは、主に音韻の種類によって決まる。音韻の種類による音韻時間長の差の例として、例えば、母音/a/は、母音/u/よりも平均的に長い音韻時間長を持つ。これは顎の動きで説明できるものと考えられる。/a/を発声するためには顎を開かなければ適当な音は出せず、顎を開くために時間がかかり、このために母音/a/は母音/u/よりも平均的に長くなるものと考えられる。また、一般に子音/s/は子音/r/よりも長い。これは子音/s/の生成には顎だけではなく、舌の設定も必要としていることによるものと考えられる。口蓋と舌先との間の間隔が

適切でなければ、正しい/s/の摩擦音は生成できない。さらに、ある時間摩擦を持続させなければ摩擦音としては知覚されない。それに比べて、/r/の発音はより単純で生成に時間はかからず、子音/r/は子音/s/よりも短くなるものと考えられる。このような音韻毎の特徴を取り除くため、前述したF_0値と同様にZスコアを用いた正規化を音韻の種類毎に適用する。それぞれの音韻の時間長の正規化前の値と正規化後の値を図1に示す。

＜図1　各音韻の時間長。上段は母音、下段は子音で、左が実測値、右がZスコア＞



基本的な考え方は前述したF_0の場合と同様である。正規化の結果、それぞれの音韻時間長のZスコアはほとんど-3から＋3の範囲におさまる（音韻時間長の分布が正規分布であるとすると全体の99.7%）。このように実測値を正規化することによって、音韻固有の時間長特性に依存しない各音韻の時間長伸縮を見ることができる[6]。

2.3 パワーの正規化

上述の音韻時間長に加えて、抽出が容易であり、かつ明瞭に識別できる韻律的パラメータとしてパワーがある。音声波形のパワーの違いや変化は、主に2つの理由による。ひとつは録音マイクとの距離であり、もうひとつは声の大きさそのものである。マイクとの距離の影響を除外すれば、プロミネンスと声の大きさの間には相関関係がある。このようなパワー情報を使用することによって、発話におけるさらに堅固な指標を読み取ることができる。パワーの正規化は、音韻時間長の場合と同様の手法により音韻毎に行なう。

3. レキシカル イントネーション(単語アクセント)

レキシカル イントネーションは、単語の辞書的性質によって決まる単語のアクセントの違いを示し、Pro.L1、すなわちイントネーションの最も低次のレベルである。しかし、この種の区別は文脈だけでは意味が曖昧なときに限って意識され、通常、文脈から意味を区別する。

レキシカルイントネーションは高次のレベルの要素と異なり、方言の影響が最も出やすい。レキシカル イントネーションはイントネーションを論ずる上で重要な要素ではあるが、非常に多くの研究が既に発表されており[7-10]、しかも本論文の主題が、より高次のイントネーションの機能を論ずることにあるので、ここでは以上の言及に留める。

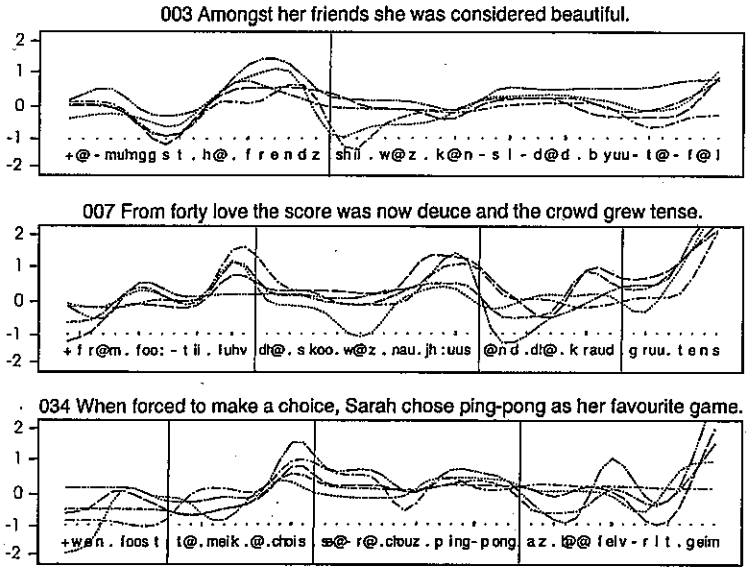
4. チャンキング(韻律区切り)

発声された文の意味を正しく理解するためには文を構成している要素に分ける必要があり、分け方を誤ると意味内容まで変わってしまうケースがしばしばある。従来の研究によって、英語韻律境界についてZスコアで正規化した音韻時間長とパワー情報を用いて韻律境界を検出する手法が提案されており、その有効性が示されている[11,12]。手法として3段階の処理を行なう。

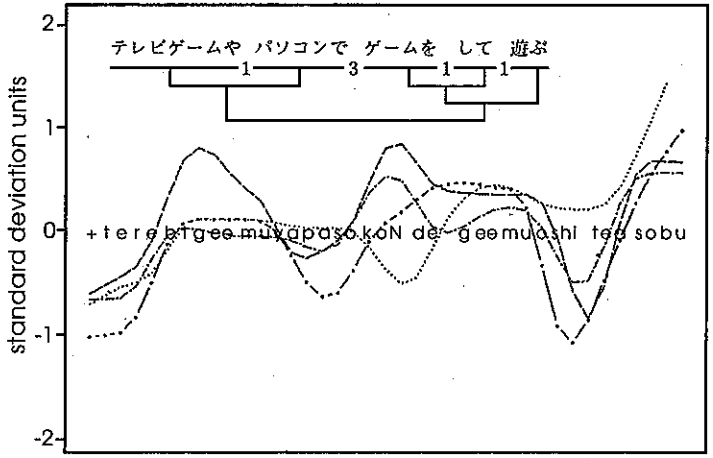
a) 正規化した音韻時間長(Zスコア)を求める。
b) 音韻時間長伸縮の傾斜を算出する(タイミングprofileを作成する)。
c) タイミングprofileの傾斜から韻律境界を推定する。

この実験は、4人の話者によって読まれた200文をデータとして用いた。その結果、すべての話者の音声が比較的似通った音韻時間長曲線を示し、韻律境界が精度良く抽出できることが示された。この実験の結果より、音韻時間長と韻律境界には有意な相関関係があることが明らかになった。なお、図2には英語3文の正規化音韻時間長曲線とそれによって推定された韻律境界を示した。図3に示すように本正規化手法を用いることにより、英語だけでなく日本語についても同様の傾向が見られている。日本語のfinal lengtheningに関しては文献[13]を参照されたい。

＜図2 英語話者4人が発声した英語の文音声の正規化時間長(Zスコア)の変化パターン＞



＜図3　日本人4人話者が発声した日本語の文音声の正規化時間長(Zスコア)の変化パターン＞



5. フォーカルプロミネンス(意味的強調)

ピッチの高低が同一の文であっても、さまざまな発話の仕方が考えられる。「Please take the subway to Kita-Oji station(地下鉄で北大路駅まで行ってください)」を例にあげると、英語でも日本語でもニュートラル、つまり特別な情報を含まない場合には自立語はそれぞれほぼ同等のプロミネンスとなる。これに対して、「どこで降りればいいですか」という質問の答えの場合では「Kita-Oji(北大路)」にフォーカスが置かれ、「タクシーで行くべきですか、それともバスの方がいいですか」という問いの場合には、「いいえ、タクシーやバスではなくて」という意味を表わすために「subway(地下鉄)」にフォーカスが置かれると言えよう。

5.1 プロミネンスとピッチ、音韻時間長

上記の3文の比較の例について、韻律特徴の変化パターンを図4に示す。図4に示す3曲線は、同一文をそれぞれ異なった意味を表わすように、3つの異なるフォーカスパターンで読んだ音声を分析したものであ

る。例とした、"Please take the subway to Kita-Oji station."のフォーカスパターン1(図中では太めの一点鎖線で表示)はフォーカスなし、パターン2(図中では破線で表示)はsubway(地下鉄)に、パターン3(図中では細めの一点鎖線で表示)は、Kita-Oji(北大路)に強調がある。

＜図4　フォーカスの抽出法。図a(左上)は正規化音韻時間長の変化パターンを、図b(右上)はパワーの変化パターンを、図c(下)は両者の和を示す。＞

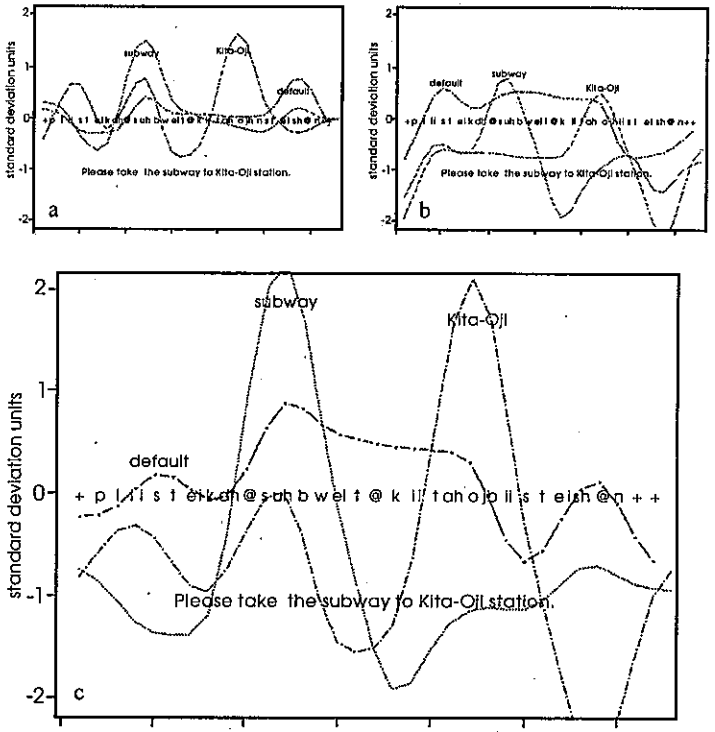


図4aは、それぞれについて音韻毎の正規化時間長を表わす。プロミネンスが置かれている部分は一般に音韻時間長が伸びているものと考えられ、この曲線の山が発話におけるプロミネンスを示していることが期待されるが、この音韻時間長の情報のみでは、これらが強調によっ

て生じた音韻時間長伸長か、あるいは韻律句末の影響による音韻時間長伸長かを区別することができない。図4bは、同一文例のパワーの変化を表わしたものである。パワーの変化パターンは、正規化音韻時間長の変化パターンが山を示した箇所のうち強調個所では正の値、韻律句末では逆に負の値となる。そこで、これらの両者の情報を同時に用いることを試みる。図4cは、図4aと図4bの2つの情報を加えたものである。図4aにおいて、区別できなかった2種類の音韻時間長伸長(共に正の値を示す)にパワー(強調か韻律句末かによって正の値と負の値を示す)を加えることによって、明確に両者を区別することが可能となる。

このようにすることによって、フォーカルプロミネンス(この例では「フォーカスなし」、subway、Kita-Ojiのいずれか)はさらに顕著な山として捉えられる。なお、本方法を300文に適用した実験では、フォーカス抽出は発話スタイルによって74～79%の精度で可能であることが確認されている[12]。

5.2 プロミネンスと音源パラメータ

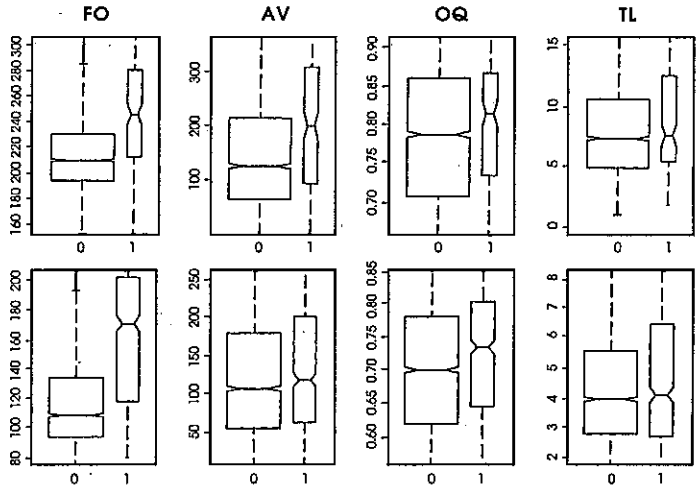
以上では、基本周波数やパワーなどの基本的な韻律パラメータを用いることによるフォーカスの抽出について述べてきたが、ここではさらに、プロミネンスと、音源パラメータとの関係について触れておく。ここで述べる音源パラメータは、これまで述べてきたF_0やパワーなどの基本的な韻律パラメータと異なり、スペクトル特徴に現われる韻律の影響を表わすものである。このため、発声された音声のスペクトルを、主に韻律の情報を表わす音源の特徴と、主に音韻の情報を表わす声道の特徴とに分離することが必要となる。このうち音源の特徴は強調や発話スタイルに応じていろいろに変化することが知られており[14,15]、音声波形から音源パラメータを抽出することでさらに強調箇所の検出精度を精度良く推定できる[16]。

このような観点から、英語および日本語の対話文を用いて文中のプロミネンスと音源特性の関係について調べた。この実験では、これまで用いてきたF_0に加えて、声帯音源波形振幅、声門開放率、音源スペクトル傾斜などの音源パラメータを用い、単語フォーカスを含んだ対話文のプロミネンスの抽出を行なった[17]。分析した音声データは、英語女性話者1名と日本語男性話者1名が、それぞれ英語[18]と日本語(翻訳したもの)で会議登録の対話文を読んだもので、各話者につき98文である。対話文は、例えば、

a) 案内書に 記載されている 口座番号に 振り込んで下さい。
b) **案内書**に 記載されている 口座番号に 振り込んで下さい。
c) 案内書に 記載されている **口座番号**に 振り込んで下さい。

のように同一の文字列から構成され、プロミネンスの位置だけが異なるものである。図5に出力の例を示す。実験結果より、それぞれの音源パラメータは、プロミネンスと相関関係があり、フォーカス検出の実験では、英語では80%、日本語では74%の正解率が得られている。

<図5 音源パラメータとプロミネンスの関係。FO、AV、OQ、TLはそれぞれ声帯振動周波数、声帯音源波形振幅、声門開放率、音源スペクトル傾斜を表わし、各図の横軸の0はプロミネンスなしを、1はプロミネンスありを表わす。なお、上段は英語、下段は日本語の結果である。>



6. 発話アクトの識別

次の段階は発話アクトを用いて発話意図の分類を行なうことである。発話アクトは、発話行為(Communication Act)の種類を示すものであり、統語論的および意味論的には違いがないが、語用論的な違いが存在するものを表わすことになる。例えば、以下に示す「そうですか」は、賛成、疑問(反対を意味する場合もある)、躊躇の意味をイントネーションの違いによって表現する。

発話アクトの識別における韻律利用の可能性を調べるために、まず、同一表記で複数の発話アクトの可能性がある発話の韻律的な違いの分析を行ない、さらに、これらの結果をもとに任意の文字列の発話に対して適用可能な、韻律に基づく発話アクトの識別方法を考案し、その方法を用いた発話アクトの識別実験を行なった。

表2(右ページ)に本実験に関係する発話アクトだけを列挙する[19]。wh-question、YN-question、confirmation-questionの3つの発話アクトとacknowledge、yesの2つの発話アクトはこれらの発話に対するシステムの動作という観点から見るとそれぞれ同じであり、発話意図も類似しているので、ここでは1つのグループとして扱い、それぞれquestion、acknowlegeとする。

6.1 同一表記文の聴取による発話アクトの識別

まず、人間が韻律の違いのみに着目してどの程度正確に意図が抽出できるかを検討した[20,21]。実験に用いた会話音声データについて述べる。実験には2種類のデータを用いた。1つは自由発話データベース[22](以下、旅行会話とする)であり、もう1つは、くだけた発話の収集を目的として新たに作られたものである。後者は2人の親近度の高い話者(関東地区出身)に雑談する時のような口調で構わないと指示を与えた上で、表裏にまたがる迷路の表側のみの情報を片方の話者に、裏側のみの情報をもう一方の話者に与え、お互いに情報交換しながら迷路を解いてもらい収録した(以下、迷路会話とする)。旅行会話には

<表2 発話アクトの一部(本実験に関係するもののみ)>

発話アクト名	説明	システム動作
inform	話し手は聞き手に情報を与える	応答／待機
wh-question	話し手が聞き手にいつ、どこ、だれが、いかに、だれにを尋ねる	応答
YN-question	話し手が聞き手に「はい」「いいえ」で答えられる質問をする	応答
confirmation-question	話し手が聞き手に確認をする	応答
confirmation-question-to-self	話し手が知り得たことを一人つぶやく	待機
acknowledge	聞き手が話し手に談話を継続するためにあいづちをうつ	応答／待機
yes	YN-questionの肯定応答	応答／待機
temporizer	話し手が躊躇を示す	待機

temporizerの発話(躊躇を示す発話)は出現するが、話者間の親近度が低いために終助詞を伴った丁寧な表現の疑問の発話しか出現しなかった。一方、迷路会話にはtemporizerの発話は出現しないが、終助詞「か」を伴った疑問だけでなく終助詞を含まない様々な疑問の発話が出現した。したがって、temporizerの識別実験には旅行会話を、questionの識別実験には迷路会話を用いた。

人間の識別能力を調べるために、文脈を伴わない単独の発話を聞いて発話アクトを識別する実験を行なった。被験者は10人(男性3名、女性7名)である。結果を表3に示す。「そうですね」以外はどれも70～80%程度は正しく発話アクトを識別することができ、聴覚的に区別可能な違いがあることがわかった。

＜表3　聴覚による発話アクト識別の正解率＞

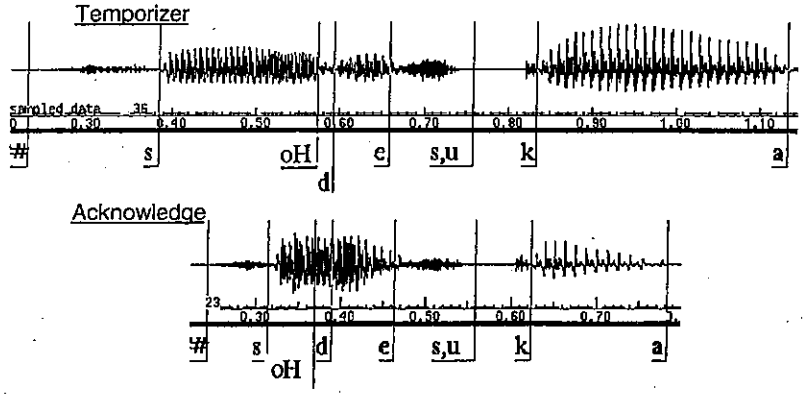
発　話	平均正解率	発話アクト	正解率
そうですか	82.2%	acknowledge	80.5%
		temporizer	91.2%
～ですか？	76.8%	wh/YN/confirmation-question	90.0%
		confirmation-question-to-self	61.8%
そうですね	49.0%	acknowledge/yes	82.0%
		temporizer	38.0%
～ですね	76.8%	confirmation-question	90.5%
		inform	76.7%

6.2　韻律を用いた発話アクトの自動認識

6.2.1　躊躇やためらいの発話の検出

次に、個々の韻律パラメータとしてF_0と音韻時間長に着目して違いの分析を行なった。「そうですか」では、図6に示す例のようにtemporizerの場合はためらいの気持ちがあるためにacknowledgeに比べて/oH/と/a/の音韻時間長が長くなることがわかった。また、「そうですか」以外の「ですか」を文末に持つ発話でも/k/の音韻時間長が長くなることがわかった。

＜図6　「そうですか」の発話アクトの違いによる時間長の違い＞



以上のことから、躊躇やためらいがある発話には音韻時間長の長い音素が出現する傾向があると考えられる。そこで、テキスト音声合成システムによって予測された音韻時間長を標準として各音素毎に標準との音韻時間長の比を求め、その最大値を入力として判別分析によりtemporizerの識別を行なう。テキスト音声合成システムとしてはATRν-Talk音声合成システム[23,24]を用いる。ただし、ATRν-Talkの音韻時間長の予測値は朗読調の場合のものなので、全発話の各音素ごとに実際の発話の平均値とATRν-Talkの予測値の平均値が一致するように係数を掛けて補正を行なう。また、temporizerは躊躇やためらいの発話であり、長い発話は出現しないことが予備的検討で確認されたため、発話の長さが10モーラ以下の場合についてのみ識別を行なうこととした。

temporizerの識別実験の結果を表4に示す。「そうですね」の場合には32.4%しか識別されていないが、表3に示した人間による文脈なしでのtemporizer識別の聴取実験の正解率を考慮すると決して悪い数字とは言えない。なお、表4からもわかるように今回使用したデータベースには「そうですか」、「そうですね」以外にtemporizerはなかった。

＜表4　音韻時間長を用いた発話アクトの識別結果＞

正解率		
Temporizer		Temporizer以外
そうですか	84.0%(21/25) [91.2%]	78.6%(246/313)
そうですね	32.4%(11/34) [38.0%]	

(注)[　]内の数字は人間による文脈なしでのTemporizer識別の聴取実験における正解率

6.2.2　疑問の発話の検出

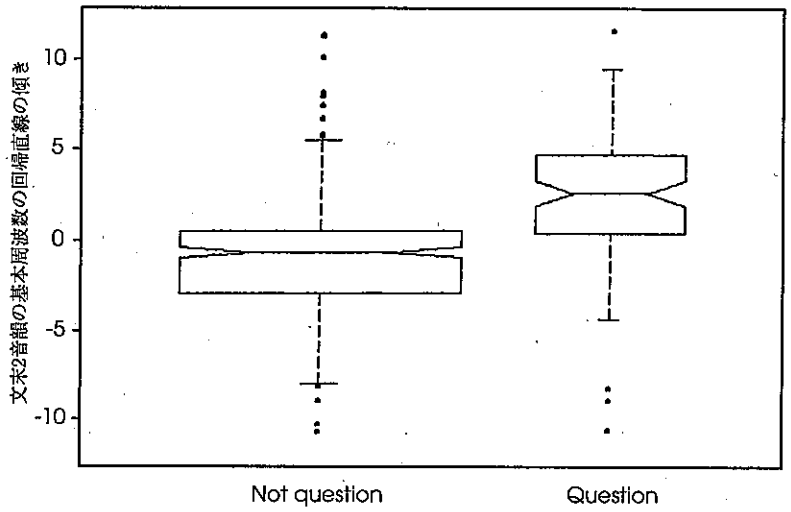
前述したとおり旅行会話中には丁寧な表現の疑問の発話しか出現し

なかったため、以下では迷路会話を用いて疑問の発話の検出実験を行なった。文末の韻律を表わすパラメータとして

a)　文末の2音韻の区間内で抽出されたF_0の回帰直線の傾き

を用いる。図7に示すように文末の2音韻のF_0の回帰直線の傾きに違いが見られ、これを入力として識別した場合、約70%のquestionの発話を識別できた。

＜図7　文末2音韻の基本周波数の回帰直線の傾き＞



次に、a)のパラメータに加えて

b)　文末の1音韻の音韻時間長のZスコア

c)　文末の1音韻の区間内で抽出されたパワーの回帰直線の傾き

を用いて決定木を作成し、questionの識別を行なった。この時、文末2モーラのピッチの高低(LH型(例えば「行く」)、LL型(例えば「見てる」)またはHH型(例えば「してる」、HL型(例えば「見える」))の違いによる識別精度への影響を調べるために、ピッチの高低で分類を行なった場合と、分類を行なわなかった場合の両方について識別をする。

表5に示す結果のように約80%のquestionの発話が識別され、3つのパラメータを用いたことによる精度の向上が確認された。また、ピッチの高低を3種類に分類した場合と、分類せずに識別した場合では精度の差はあまりなかった。この結果から、questionの文末のF_0パターン等に文末のピッチの高低が大きな影響を与えていないと考えられる。したがって、questionの識別過程において文末のピッチの高低による分類は必要ないと思われるが、これは2名の話者のデータでの実験結果であるため、さらに話者を増やして分析および実験する必要がある。

＜表5　questionの識別結果＞

文末ピッチの区別	ピッチの高低	正解率	
		Question	Question以外
あり	HL	76.9%　(10/13)	98.1%(154/157)
	LL or HH	85.2%　(52/61)	84.8%(195/230)
	LH	82.8%　(24/29)	92.3%　(36/39)
	平均	83.5%(86/103)	90.4%(385/426)
なし	——	89.3%(92/103)	86.4%(368/426)

7.　結論

現在ATR音声翻訳通信研究所では音声翻訳システムの構築を進めており、音声翻訳の場合、入力文のイントネーションによって最適な出力文は異なってくる。本章で述べた一連の研究はこのようなアプリケーションに有効であるものと考えられる。現在のところ、完全とは言えないまでもPI理論のPro.L4までの自動識別が可能であることを確かめたが、Pro.L5以降については、まだ検討の段階である。

本章でははじめに、意味認識の構造化を目的として、PI理論に基づいて韻律の階層別の役割について述べた。すなわち、韻律情報の機能を、辞書的レベルから話者性レベルまでの7つの段階に分類し、その階層性を明らかにした。

次に、入力の音声波形の情報を音響的に分析して得られるパラメー

タを韻律の分析という観点から有効に利用するための正規化手法について述べた。また、韻律的情報として、音韻時間長、ピッチ、パワーだけではなく、音源パラメータとプロミネンスとの関係について明らかにした。Zスコア正規化によって、話者性や発話スタイルの影響を除き一般化したメジャーを用いて、それぞれの段階の意味の情報抽出を検証した。

PI理論が決定するように、イントネーションは意味の層の段階毎に個別の機能を持つ。筆者は、韻律区切りおよびプロミネンスの抽出に対してイントネーションが有効であることを述べるとともに、今回は特に発話行為の意味理解レベル、つまり、話者意図を理解するための、発話アクト識別手法について述べた。特に、音韻の時間長の伸長度がtemporizerの識別のために有効であり、これを用いることにより約8割の精度で識別可能であること、文末のF_0パターンの傾きが疑問の識別に有効であり、これを用いることにより約7割の精度で識別可能であること、さらに文末の音韻の時間長および文末のパワーの傾きを入力として加えることにより、識別の精度が約8割に向上することを示した。

韻律によって示されるものは、現在自動的に測定あるいは抽出可能ではないものも含めると、さらに多くの情報があると言える。例えば、人間は相手の発話から「うれしい」、あるいは「悲しい」といった話し手の気分を区別したり、「疲れている」、「酔っぱらっている」といった話し手の状況を判断することができる。またさらにはどの程度真剣かなどといった、話者の関与の程度について聞き分けることができる。

現在準備的段階として、大規模な音声データベースに対して音響情報のラベリングを開始し、これらのデータベースをもとに、発話タイプ別音響的特徴および発話の個人性の識別の研究段階に入っており、今後さらに高次の韻律処理が可能になるものと期待される。

謝辞: Acknowledgements

I would like to express my gratitude to Norio Higuchi and Kumiko Hayakawa-Campbell for their help with the production of this paper, and to Shigeru Fujio and Wen Ding for their contribution to its content.

参考文献

[1]藤崎博也・須藤　寛 (1971)「日本語単語アクセントの基本周波数パタンとその生成機構のモデル」日本音響学会誌 vol. 27, p.445 - 453.
[2]H. Fujisaki and K. Hirose (1984) "Analysis of voice fundamental frequency contours for declarative sentences of Japanese", J. Acoust. Soc. Jpn (E), vol.5, p.233-242 .
[3]箱田和雄・佐藤大和 (1980)「文音声合成における音調規則」電子通信学会論文誌 D Vol.J63-D, p.715-722.
[4]河井　恒・広瀬啓吉・藤崎博也 (1994)「日本語音声の合成のための韻律規則」日本音響学会誌 vol.50, p.433-442.
[5]N. Higuchi, T. Hirai and Y. Sagisaka (1994) "Effect of Speaking Style on Parameters of Fundamental Frequency Contour", Proc. 2nd ESCA/IEEE Workshop on Speech Synthesis, p.135-138.
[6]W. N. Campbell and S. D. Isard (1991) "Segment durations in a syllable frame", Journal of Phonetics, vol.19, p.37-47.
[7]金田一春彦 (1967)『日本語音韻の研究』(東京堂出版).
[8]服部四郎 (1960)『言語学の方法』(岩波書店).
[9]杉藤美代子 (1982)『日本語アクセントの研究』(三省堂).
[10]杉藤美代子 (1996)『大阪・東京アクセント音声辞典』(丸善).
[11]W. N. Campbell (1992) "Prosodic Phrasing from Normalised Acoustic Measures", 日本音響学会春季研究発表会講演論文集, p.243-244.
[12]W. N. Campbell (1993) "Durational Cues to Prominence and Grouping", Proc. of ESCA Workshop on Prosody, p.38-41.
[13]W. N. Campbell (1992) "Segmental Elasticity and Timing in Japanese Speech", in Speech Perception, Production and Linguistic Structure edited by Y. Tohkura, E. Vatikiotis-Bateson and Y. Sagisaka (Ohmsha).
[14]W. N. Campbell and M. Beckman (1995) "Stress, Loudness, and Spectral Tilt", 日本音響学会春季研究発表会講演論文集, p.279-280.

[15]W. N. Campbell (1995) "Loudness, Spectral Tilt, and Perceived Prominence in Dialogues", Proc. of The XIIIth Int. Congress of Phonetic Sciences, vol.3, p.676-679.
[16]W. Ding, H. Kasuya and S. Adachi (1995) "Simultaneous Estimation of Vocal Tract and Voice Source Parameters Based on an ARX Model", IEICE Trans. on Information & Systems, vol.E78-D, p.738-743.
[17]丁　文・W. N. Campbell (1996)「プロミネンスと音源パラメータの関係について」日本音響学会秋季研究発表会講演論文集, p.197-198.
[18]W. N. Campbell (1993) "Automatic detection of prosodic boundaries in speech", Speech Communication, vol.13, p.343-354.
[19]M. Tomokiyo (1995) "Segmentation and Aggregation of Utterances by Using Speech Act Labels", 電子情報通信学会研究技術報告, NLC95-23.
[20]藤尾　茂・W. N. Campbell・樋口宜男 (1995)「韻律を用いた発話タイプの識別」日本音響学会秋季研究発表会講演論文集, p.199-200.
[21]藤尾　茂・W. N. Campbell・樋口宜男 (1996)「韻律を用いたテキスト非限定型発話アクト識別方式」日本音響学会春季研究発表会講演論文集, p.245-246.
[22]T. Morimoto *et al.* (1994) "A speech and language database for speech translation research", Proc. of Int. Conf. on Spoken Language Processing, vol.IV, p.1791-1794.
[23]匂坂芳典・海木延佳・岩橋直人・三村克彦 (1992-10)「ATR ν-Talk 音声合成システム」情報処理学会「音声言語情報処理と音声入出力装置」研究グループ、電子情報通信学会「音声認識の実用化をめざす新手法」時限研究専門委員会研究会資料.
[24]Y. Sagisaka, N. Iwahashi and K. Mimura (1992) "ATR ν-TALK Speech Synthesis System", Proc. of Int. Conf. on Spoken Language Processing, p.483-486.