＜情処論投稿原稿＞

［題名］

日本語音声におけるポーズ位置および アクセント位置の持つ情報量

Information of Pose and Accent in Japanese Speech

［著者］

村上仁一*　:非会員　Jinnichi Murakami

池原　悟**　:会員　Satoru Ikehara

荒木哲郎***:会員　Tetsurou Araki

＊　:ＮＴＴヒューマンインターフェース研究所

＊＊　:ＮＴＴ情報通信処理研究所

＊＊＊:福井大学工学部

連絡先：

〒238-03　横須賀市武１丁目２３５６

ＮＴＴ情報通信処理研究所

ＭＳ部　池原　悟

あらまし

自然言語の音声では音節と同時に、アクセント、ポーズ、イントネーションなどの韻律も用いられる。特に、日本語は音節の種類が少ないこともあって同音異語が多いため、音声言語における韻律情報の役割は大きく、その扱いは日本語音声の認識においても考慮すべき重要な課題と考えられる。

本論文では、韻律情報が音声認識において認識率の向上にどの程度役立つかを定量的に把握するための第一歩として、韻律情報の中の基本的な情報であるポーズ位置及びアクセント位置が持つ情報量を、実験的に求めた。

すなわち、音節情報から漢字かな混じり日本文を生成する際、韻律情報の存在によって解釈の曖昧さがどの程度削減できるかという観点から、韻律情報の持つ情報量を求める簡易な方法を提案し、それに基づいて、１６万語の単語辞書を持つ大量語彙の環境下で、新聞記事文を対象に実験を行い、韻律情報の有無による多義絞り込みの効果を求めた。その結果、ポーズ位置の持つ情報量は文節当たり平均３．２１ビット、ポーズ位置とアクセント位置が合わせ持つ情報量は平均５．１６ビットであることが分かった。前者は、曖昧な候補数を平均で１０．８％にまで絞り込むことに相当し、後者は候補数を２．８％にまで絞り込むことに相当する。これらの結果、韻律情報は日本語音声認識における曖昧さを解消する上で有効な手がかりになり得ることが確認された。

１．まえがき

一般に書かれた自然言語の文では、話者の認識した内容は、言語の規範（文法）に基づいて文字と関係づけられるため、読者はその規範を手がかりに表現に結び付けられた話者の認識を追体験する。日本語は特に音節の種類が少ないこともあって同音異語が多いが、書き言葉では漢字やかな文字等多くの字種が使用されるため、使われた文字によって単語が特定され易く[1,2]、著者の記述内容を理解する上で助けとなっている。

これに対して、音声化された自然言語の文ではアクセント、ポーズ、イントネーション等の韻律を付与することによって、少ない音節から生じる解釈の困難さを避け、聞き手の理解を助けている。これは、音声の場合における前述の同音異語の問題の解決の手段ともなっている。

ところで従来、音声合成の分野では、漢字かな混じり日本語文を対象に、アクセント、ポーズ、イントネーションなどの情報を自動付与する方法について研究が行なわれてきた[3,4,5]。そこでは、一般名詞、固有名詞、数量表現など多様な語彙と表現を持つ日本文に対して、複数の読みや特殊な読みが存在すること、また単語が複数連接して構成される複合語や文節に対して、アクセントやポーズの位置が変化することなどを考慮して、それらの付与方法が提案されている。これらの結果を見ると、日本語文に対するアクセント、ポーズ及びイントネーションなどの韻律情報は、単に単

語レベルで決まるものではなく、単語間の文法的及び意味的な接続関係などを背景とした規則性を持っていることが認められる。

これに対して、このような韻律情報が持っている文法的及び意味的な情報を、解釈の曖昧さの解消に利用して、音声認識の精度を向上させる研究[6]については、まだ見るべきものがない。この理由としては、日本語文の自然な連続音声に対して韻律情報を正しくとらえることが物理的に困難なこと、韻律情報は個人差や地域差があり統一的な取扱いが難しいことなどが上げられる。

本論文では、「情報量は、その情報が存在する場合と存在しない場合での解釈の可能性（曖昧さ）の差で表される」という点に着眼して、日本語音声のポーズやアクセントなどの韻律情報が持つ情報量の評価法を提案する。

すなわち、発話された日本語文に対して正しい音節情報と韻律情報が得られた時に、音節情報のみに基づき音節漢字変換されて得られた漢字かな混じり文候補[7]（文法的、また意味的な誤りを持つ候補を含む）の数と、韻律情報をも用いて得られた漢字かな混じり文候補の数を比較することによって、韻律の持つ情報量を求める。その際、音節と韻律情報から漢字かな混じり文を、如何にして正確に生成するかが問題となるが、この問題は従来、開発されてきた日本文音声出力プログラム[4,8]の機能を用いることにより解決できること

を示す。日本文音声出力システムでは、種々の韻律情報を出力するが、ここでは最も基本的な情報である第一ポーズ位置、及び第一アクセント位置が持つ情報を対象とする。

２．ポーズ位置及びアクセント位置が持つ情報の定義とその測定法

２．１　情報量の測定に関する基本的な方法

情報は確率事象に対して定義され、確率事象Ｅが生起したこと知った時、

$$I = -LOG2P(E) \qquad \cdots\cdots\cdots(1)$$

の情報量を受け取ったという。但し、P(E)は事象Ｅの生起確率を示す。

このことは、一般に事象Ｅが生起したことを知ることによって、対象とする事象系に関する解釈の曖昧さがＩビットだけ減少することを意味し、この解釈の曖昧さの減少の度合いが、事象Ｅの情報量にあたる。この考え方によって、ポーズ位置やアクセント位置が持つ情報量を求めることができる。つまりポーズ位置やアクセント位置の情報を知ることによって、解釈の可能性（曖昧さ）が減少するような事象系を一つ設定し、ポーズやアクセントの情報が存在しない時に得られる解釈の数と、それらの情報が存在する時に得られる解釈の数を比較すればよい。

図１のような音節漢字変換モデルを考えると、韻律の持つ情報量は以下の手順で求めることができる。

［処理１］音節の情報から得られる漢字かな混じり文を全て出力する。

与えられた音節列を漢字変換Ａにより、漢字かな混じり文に変換し、得られた文の集合をS_iとする。集合S_iの要素をβ_i（１≦ｉ≦ｍ）とする。

［処理２］音節、ポーズ位置及びアクセント位置の情報から得られる漢字かな混じり文を全て出力する。

上と同様、与えられた音節列と韻律情報を漢字変換Ｂによって漢字かな混じり文に変換し、得られた文の集合をS_2とする。集合S_2の要素をβ^*_i（１≦ｉ≦ｎ）とする。

［処理３］両方の漢字かな混じり文の比を計算してそれを確率Ｐとする。

集合S_1の要素数ｍと集合Ｓ２の要素数ｎの比をＰ（＝ｍ／ｎ）とすると、韻律の持つ情報量は、(1)式よりＩ＝－Log_2Ｐと表すことができる。

しかし、ここで示した方法を直接実行するには以下の問題があり、容易でない。

①漢字変換Ａの正確さ

漢字変換Ａの方法としては、単語の文法的意味的関係を考慮せず、音節情報と一致する漢字候補（漢字の読みが等しいもの）を単語辞書より全て抽出し、これらを組み合わせて出来る漢字かな混じり文を求める方法が考えられるが、このように

して求められた漢字かな混じり文候補は、それを逆に読み上げたとき元の日本文の音節情報と一致するとは限らない。もとの音節と一致しない候補は、もとの音節情報を使い切ったとは言えない。

② 漢字変換Ｂの困難さ

漢字変換Ｂでは韻律情報を活用して漢字かな混じり文を生成するため、①に加えて韻律情報生成ルールを逆に適用するシステムを作成する必要があるが、韻律を含むかな漢字変換のアルゴリズムは確立しておらず、精度良い実験を行なうのは困難である。

２．２　漢字音節変換を利用したポーズ位置及びアクセント位置の持つ情報量の測定法

（１）実験の手順

前節で述べた問題は従来開発されてきた日本文音声出力システムを図２のように用いることによって容易に解決することができる。また、このシステムを使用することによって、「音節列」と「音節列＋韻律」の試験標本を作成するのも容易となる。その手順は、次の通りである。

［手順１］漢字かな混じり文を音声変換によって、音節、ポーズ位置及びアクセント位置の情報に変換する（試験標本の準備）。得られた音節列をαとし、音節列＋韻律をα^*とする。

［手順２］次に音節列αに対して音節漢字変換Ａを行

い、漢字かな混じり文β_i（１≦ｉ≦ｍ）を出力する。ここで用いる音節漢字変換Ａとは、通常のワードプロセッサに用いられているかな漢字かな変換機能に、鼻音、長音及び「は」、「を」などの表記が異なる部分に対する変換機能が追加されたものに相当する。

なお、図には明示していないが、音節情報からの変換精度を上げるため、得られた漢字かな混じり文候補は日本文音声変換プログラムを用いて再度音節列に変換し、もとの音節列に一致した候補のみを拾いだして、集合S_1とする。

［手順３］以上で得られた漢字かな混じり文β_iに対して、再び音声変換を行って、ポーズ位置及びアクセント位置の付加された音節列の集合S_3を求める。（この音声変換は上記①の音声変換と共通化できる。）

［手順４］最後に、集合S_3の音節＋韻律候補を①で求めた元の正しい音節＋韻律であるα^*と比較して、音節列とポーズ位置、アクセントの位置が一致する漢字かな混じり文候補のみを拾い出す。このようにして得られた漢字かな混じり候補をS_2の要素とし、その数ｎと集合S_1の要素数ｍとの比から、Ｐを求める。

（２）評価法と実験方法の等価性

漢字変換と音声変換が、いずれも正しく行なわれる

と仮定すれば、図２の方法は図１の方法と等価であることを示す。集合S_1の求め方は図１と図２で変わらないから両者の集合は一致する。従って、評価法の等価性を示すには、図１と図２で集合S_2の要素が一致することを証明すればよい。集合S_2の一致は以下の通り証明される。

＜ステップ１＞

図１において、S_2はS_1の部分集合であること、及び図１と図２の集合S_1は等しいことは明かである。また仮定より音声変換は正しく行なわれるから、図２においてS_1の要素とS_3の要素は１対１に対応する。従って、図２におけるS_3はS_2の各要素に対応する要素を漏れなく含む。

＜ステップ２＞

図２において、仮定より漢字変換Ｂは正しく行なわれるから、集合S_2の要素$\beta_i{}^*$を音節韻律列に逆変換すれば、変換結果はいずれもα^*に一致することが保証される。従って、S_3の要素の内、α^*と一致したα_iを抽出し、それに対応するS_1の要素β_iを拾い出せばそれがS_2の要素である。

＜ステップ３＞

以上から、集合S_3の要素の内、α^*に一致するものの数を求めるとその集合は図１の集合S_2と一致しその要素数はｎである。

図１の代わりに図２の方法を使用しても結果は保証

されることが証明された。現状技術からみて、技術的に未解決な漢字変換Ｂのプログラムを使用する図１の方法より、技術の進歩によって精度の向上した日本文音声変換プログラムを使用する図２の方法の方が、正確な結果が得られるものと期待される。

３．実験の条件

３．１　入力文の種類と入力単位

文単位の連続音声認識に比べて、当面、文節単位の音声認識が急務と考えられるので、実験では入力単位を文節とする。前章では文単位の韻律情報量の評価方法を示したが、入力単位を文節とする場合も、まったく同様である。

実験に用いる日本語文節は、新聞の一般記事（１９８２年の日経新聞）より任意の連続した文から５０文節を選択した。実験で使用した文節を表１に示す。

３．２　漢字音節変換

漢字音節変換では、日本文訂正支援システムＲＥＶＩＳＥの[8]音声出力機能（日本文音声出力システムＪＴＯＳを内包する）を用いる。本システムの音声出力機能では、３種類のポーズ長や、２種類のアクセントなど多様な韻律情報が出力されるが、本実験では第一ポーズ位置と第一アクセント位置の情報だけを用いる。また、新聞記事を対象としたＪＴＯＳの品質実験の結果からみて、変換精度は、漢字かな混じり文の音節へ

の変換では９９．８％（文字単位）、韻律の付与では９５％（文節単位）の精度が期待できる。

３．３　音節漢字変換

音節漢字変換では文節数最小法[9]を使用した。ただし生成候補の漏れを防ぐため、分割数が「最小分割数＋１」までの候補を抽出した上で、ＲＥＶＩＳＥによって逆変換し、もとの音節と一致するものを集めて集合S_1の要素とした。これにより、読み上げたとき音節がもとの音節列に一致する漢字かな混じり候補は、ほぼ漏れなく抽出できると期待されるが、分割の違いから同じ表記の漢字かな混じり文が重複して出力されることが考えられる。このような場合は、重複する漢字かな混じり文を一つ残し、残りは削除した。

なお、音節漢字変換に使用した単語辞書は約１６万語である。

３．４　情報量の種類

アクセントは二つのポーズに挟まれたアクセント区間の音節列に対して付与されることに留意し、実験ではポーズの情報量とポーズとアクセント両方の情報量との二つの場合について求める。

すなわち、韻律の情報量I_iを、

$$I_i = -LOG_2(P_i)$$

とし、以下のｉ＝１，２の場合についてI_iを求める。

① ポーズ位置の情報量：ｉ＝１

$$P_1 = \frac{\text{音節及びポーズ位置が一致する漢字かな混じり文}}{\text{音節情報が一致する漢字かな混じり文}}$$

② ポーズ位置とアクセント位置の情報量：ｉ＝２

$$P_2 = \frac{\text{音節、ポーズ位置、及びアクセント位置が一致する漢字かな混じり文}}{\text{音節情報が一致する漢字かな混じり文}}$$

４．実験結果

４．１　漢字かな混じり文節候補の数の比較

例として、標本内の第１文節「大蔵省は」に対して得られた漢字かな混じり文節候補の一部を図３に示す。図では音節、韻律を比較して集合S_1とS_2の要素を決定する手順を示している。

同様して得られた標本文節５０件に対する実験結果を図４、図５に示す。図４では標本の各文節に対して得られた漢字かな混じり文節候補の数を、また、図５では、音節の一致する漢字かな混じり文節の数に対する各漢字かな混じり文節の数の比を示している。ただし、文節番号３４以降は省略した。また、グラフ中で空白になっている文節は漢字変換されて出力された漢字かな混じり文の数がスケールを超えるため記載しなかった。

これらの結果を用いて、標本文節５０件に対する漢字かな混じり文節候補数の平均値を求めると、表２、

図６の結果を得る。また、表２から（１）式を用いて、最終的に韻律の持つ情報量は表３の通りと求まる。

４．２　結果の考察

以上の結果から、アクセント、ポーズをあわせた韻律情報量は５．１６ビットと大きく、音節のみから音節漢字変換するのに比べて、候補数を１／３６に絞り込む効果が期待できることが分かる。

なお、文節毎に見れば、漢字変換によって得られる漢字かな混じり文節の候補数と、その韻律の一致する候補数の割合は大きなばらつきがあるが、文節を主語（主題文節）、述部、その他の文節に分けて集計した結果では、文節の種類の違いによる韻律情報の差は余り認められなかった。

５．追加実験と結果の検討

５．１　漢字の読みの知識の情報量

日本語の漢字には複数の読みが存在し、前後の漢字や、意味によって読み方が変化する。このため音節を漢字辞書を用いて漢字かな混じり文に変換しても、これを再び音節に変換した場合、元の音節には戻るとはかぎらない。そこで実験では日本文音声変換システムを用いて逆変換を行ない、もとの音節に一致したものをS_1の要素としたが、ここでは漢字かな混じり表現の読みの知識の持つ情報量を推定評価するため、逆変換によってもとの音節列に一致するものの割合Ｐを求め

る。

$$P = \frac{\text{音節の一致する漢字かな混じり文の数}}{\text{文節漢字変換が出力する漢字かな混じり文の数}}$$

とすると、Ｐ＝０．２０９となるから、読みの知識の持つ情報量Ｉ＝－LOG$_2$（Ｐ）は２．２６ビット、また分散は０．８８となった。

これは韻律の持つ情報量に比べて若干小さいが、無視出来ない情報量である。すなわち、単語が連なっているときは、人は読みの知識として、単語辞書に記載されている単語読み方以外の知識も使って、それを読んでいるのであり、その知識も無視できない情報量を持っていることが分かる。

５．２　文節内文法情報との比較

前章で得られた韻律の情報量と日本語の持つ文節内の文法情報とを比較するため、図７の方法で日本文の誤り検出プログラム（ＲＥＶＩＳＥ）を用いて、文節内文法情報による漢字かな混じり文節候補の絞り込み実験を行なった。ＲＥＶＩＳＥは文法情報や誤用語辞書を使って単語レベルでの日本文の誤りを検出するプログラムであるが、入力を文節単位とした場合は文節内の文法ルールが適用されて、誤字脱字などの誤りを検出する機能を持つ。誤り検出の精度は約９０％である。

図２の実験で得られた漢字かな混じり文候補の集合S_1を入力し、誤り無しと判定された漢字かな混じり文をの集合をS_4、誤りありと判定された漢字かな混じり文の集合をS_5とし、それぞれの要素数をm_4，m_5とすると、文節内文法の情報量は$P = m_4 / m$ としてＰを求め、これを式（１）に代入して求めることができる。

この実験結果をまとめると以下に通りである。

① S_1の内、明確な誤りとしてS_5に分類されるものの数＝１４．５％

② S_1の内、誤りは検出できないとしてS_4に分類されたもの＝８５．５％

なお実験では、S_4の内、７３．５％は誤りの可能性があることが指摘されたが、指摘内容の大半は同音異語の存在であった。

この結果から文法情報の持つ情報量を計算すると、情報量は小さく、Ｉ＝０．２３ビットとなり、絞り込みの効果は小さい。このことは、解釈の曖昧性としては同音異語の問題が大きいのに対して、同音異語の曖昧性は単語接続規則などの文法規則では余り期待できないことを示している。

以上の実験から、音声認識においては文法情報よりも韻律情報の方がより重要であることが分かった。

４．むすび

日本語音声におけるポーズ、アクセント等の韻律の持つ情報量が、日本文音声出力プログラムを応用して

求められることを示し、文節単位に入力された日本語に対して、実験的にポーズの位置とアクセントの位置の持つ情報量を求めた。

その結果、ポーズ位置が持つ情報量は文節当たり平均３．２１、ポーズ位置及びアクセント位置を併せた情報量は５．１６になることがわかった。これは音節から、漢字かな混じり文への変換において生じる解釈の多義を、２．８％にまで正しく絞り込むことに相当する。同時に行なった文節内文法情報による解釈絞り込み実験の結果と合わせてみると、音節情報からの漢字かな混じり文生成においては、同音異語の問題が大きなウエイトを持つが、文節内文法情報はこの問題に無力であるのに対して、韻律情報はかなり有効であると言える。

なお実験では、文節単位の音声入力を考え、文節単位入力の場合の韻律情報量を求めたが、今後は文単位入力の場合についても検討する必要がある。その場合は、音節漢字変換における漢字かな混じり文の候補数が膨大となる点で工夫が必要となるが、日本文の文節相互間の係り受け関係によって、もたらされる情報量などについても、調べて比較することが期待される。

また、今回は入力として正しい音節列を仮定したが、音節自身に複数の認識候補のある場合も含めて、音節マルコフ連鎖や漢字マルコフ連鎖などのマルコフ連鎖情報とアクセント位置、ポーズ位置の情報の組み合わせ方等についても検討することが必要がある。

謝辞

本研究では日本文音声出力システムＪＴＯＳとこれを応用した日本文訂正支援システムＲＥＶＩＳＥを実験に使用したが、これらのシステムは著者の一人が宮崎正弘新潟大教授、安田主幹研究員、高木、島崎両主任研究員等の方々と共に開発したものである。本研究の手段を実現されたこれらの方々と日頃議論して頂く自然言語処理関連グループのみなさまに深謝する。

参考文献

(1) 池原,白井:″2次3次混合型マルコフモデルによる日本文誤字訂正候補の抽出″,第31回情処全大,7h-5,pp.1405-1406(1985)

(2)池原,白井:″単語解析プログラムによる日本文誤字の自動検出と二次マルコフモデルによる訂正候補の抽出″,情処論, Vol.25,No.2,PP.298～305(1984)

(3) 箱田,佐藤:″文音声における音調規則″,信学論,D-104,Vol.J63-D,pp.715(1980)

(4) 宮崎,大山:″日本音声出力のための言語処理方式″,情処論,Vol.27,NO.11(1986)

(5) 鈴木,斉藤:″文構造に応じたポーズ長の制御″,日本音響学会講演論文集,2-7-15,pp.199(1989)

(6) 村上,荒木,池原:″音声におけるポーズ長およびアクセント位置の情報量の考察″,日本音響学会講演論文集,3-3-11,pp.89(1989)

(7) 荒木,村上,池原:″2重音節マルコフモデルによる日本語の文節音節認識候補の曖昧さの解消効果″,情処論,Vol.30,No.4,pp.467～447(1989)

(8) 池原,安田,島崎,高木:″日本文訂正支援システム″,NTT研究実用化報告,Vol.36,No.9(1987)

(9) 長尾:日本語情報処理,電子通信学会誌 (昭和59年)

英文アブストラクト

[Title]

Information of Pose and Accent in Japanese Speech

[Authors]

Jinnichi Murakami* :Non-Member

Satoru Ikehara** :Member

Tetsurou Araki** :Member

---

* NTT Human Interface Laboratories

** NTT Communications and Information Processing Laboratories

*** Fukui University

Abstract