音声認識・理解における韻律情報の利用

The Use of Prosodic Information in Speech Recognition and Understanding

電気通信大学 電気通信学部
Faculty of Electro-Communications, The University of Electro-Communications

尾関 和彦
Kauzhiko Ozeki

The aim of the research group is to exploit prosodic information for various tasks related to speech recognition and understanding. Prosody plays a crucial role in human speech communication. However, very little use has been made of prosody in automatic speech recognition and understanding. Although the current speech recognition technology has reached the level of attaining a high recognition rate for speech signals clearly pronounced in an ideal environment like a studio, it still suffers from the lack of robustness against environmental noises and other adverse factors. The use of prosody will be one possible way to overcome such a problem. Four research themes were set up in the group: (1) exploration of the role of prosody in human speech perception and its application to automatic speech recognition, (2) use of prosody in syntactic analysis of spoken sentences, (3) construction of a speech recognition system based on the unified use of phonemic and prosodic information, and (4) automatic speech summarization using prosodic features. Through the four years' research efforts, a number of useful results have been obtained: a novel language model for large vocabulary continuous speech recognition with the use of prosody, a new method of measuring local speech rate that matches subjective speech rate well, a dependency analysis system for spoken Japanese using prosodic information, a speech dictation system having a completion function with the help of prosodic information, and a speech summarization system based on the combined use of linguistic and prosodic information.

Key words: prosody, speech recognition, speech understanding, prosodic features, accent nucleus, language model, accent phrase boundary, perplexity, temporal structure, local speech rate, subjective speech rate, effective number of morae, dependency structure, dependency analysis, dependency distance, pause duration, accent type recognition, speech dictation system, input completion, speech summarization, sentence importance, decision tree, fundamental frequency

1 研究の目的

音声に含まれる情報は音韻情報と，韻律情報に大別することができる．従来の音声認識・理解においては，これらの情報のうち，ほとんど音韻情報のみが用いられて来た．韻律は一見不規則な変動を示すため，音声の自動認識にはむしろ有害なものとして，これをできるだけ排除しようとして来たといっても過言ではない．しかし，人と人のコミュニケーションにおいて韻律が非常に重要な役割を果たしていることは言うまでもない．韻律の中に，発話態度，感情，個人性といったパラ言語情報や非言語情報が含まれていることは日常的に経験することであるが，韻律には更に単語や文の認識に関係する言語的情報も含まれていることが明らかになって来ている．このように豊富な情報を含む韻律は，音声の自動認識においても有用なはずである．現在の音声認識技術は，理想的な発声環境における読み上げに近い音声に対してはかなり高い認識性能に達しているが，雑音や反響が存在するような実環境における自由発話の自動認識・理解は未だ困難な技術的課題である．しかし，そのような問題を解決しなければ，音声認識・理解技術が研究室から出て，我々の生活環境の中で実用されるようにはならないであろう．そのためには，これまであまり利用されなかった，しかも人にとって重要な情報である韻律情報の利用法を開拓することが不可欠であると思われる．

上に述べたように，韻律には発話の言語的情報，パラ言語的情報，あるいは非言語的情報など，人の発話行動のあらゆる面に関わる情報が含まれている．本研究は，これらの中で主として言語的情報に焦点を絞り，韻律と種々の言語的情報との関係を明らか

研究課題番号：12132203

にするとともに，その工学的応用を図ることを目的とした．具体的には，まず認知科学的な観点から人の音声知覚と韻律との関係について研究を行い，工学的な側面からも認識に有効な韻律的特徴を明らかにする．そして，韻律情報を，高精度な音韻認識，認識過程における高速な単語選択，句境界の検出による認識処理の高速化，発話の構文的曖昧性の解消などに応用するための方策を見出す．また，音韻情報と韻律情報の統合により，認識精度と処理効率を改善する方法を明らかにする．さらに，談話情報と韻律的特徴の関係について解析を行い，その結果を講演音声の自動要約に応用するための手法を開発することなどを目標とした．

2　研究テーマと研究体制

本研究課題において下のような４つの研究テーマを設定した．テーマ１とテーマ２にはそれぞれ２つのサブテーマを設けた．当初の分担者であった新美康永は平成 14 年 3 月末をもって京都工芸繊維大学を定年退官したため，その後を同大学の荒木雅弘が引き継いだ．全体の統括は尾関和彦が行った．

1. 音声知覚における韻律の役割の解明と音声認識への利用

   (a) 音声知覚・認識
   分担者：峯松信明（東京大学・大学院情報理工学系研究科・助教授）

   (b) 音声の時間構造
   分担者：吉田利信（電気通信大学・電気通信学部・教授）

2. 発話の構文解析における韻律情報の利用

   (a) 係り受け解析における韻律情報の利用
   分担者：尾関和彦（電気通信大学・電気通信学部・教授）

   (b) 韻律を利用した係り受け解析の多数話者による評価
   分担者：高木一幸（電気通信大学・電気通信学部・助手）

3. 音韻情報と韻律情報を統合した認識システム
   分担者：新美康永（京都工芸繊維大学・工芸学部・教授）（平成 14 年 3 月まで）
   分担者：荒木雅弘（京都工芸繊維大学・工芸学部・助教授）（平成 14 年 4 月から）

4. 韻律を利用した講演音声の自動要約
   分担者：山下洋一（立命館大学・理工学部・教授）

研究を進めるに当って，次のように班会議を開催し進捗状況を確認するとともに，分担者相互の意見交換を行い連携を密にするよう努めた．

年度	開催日	開催場所
平成 12 年	10 月 28 日	電気通信大学
平成 13 年	8 月 21 日	電気通信大学
	11 月 17 日	京都工芸繊維大学
平成 14 年	12 月 7 日	立命館大学
平成 15 年	6 月 12 日	京都工芸繊維大学

3　研究成果

本節では，テーマ毎に得られた成果の概略について述べる．

3.1　音声知覚における韻律の役割の解明と音声認識への利用

音声知覚・認識

大語彙連続音声認識における韻律的特徴（主に基本周波数，F0）の利用を種々の観点から検討した．まず，音響特徴量レベルにおいて F0 が分節的特徴に及ぼす影響について定量的に分析した．音声情報処理では，音源の制御と声道の制御は互いに独立であるとの仮定を置くことが多いが，本研究では，音源特性と声道特性の相互依存性を音声認識応用を念頭において検討した．種々の音韻及び複数の話者の音声を用いて分析したところ，両者の相互依存性は無声子音（この場合の F0 は前後の有声区間からの内挿により求めた）の場合でも観測された．しかし，その依存性の様子は音韻の種類ばかりではなく，話者にも依存する様子が観測され，話者非依存，音韻非依存の形で両者の依存関係をモデル化することは困難であった．

次に，語レベルの韻律利用に対する試みとして，句頭アクセント核による仮説探索制御を実装した．日本語の単語アクセントはアクセント核の位置によって一意に決められるため，その位置が語頭にあるほどアクセント型の同定は早期の段階で行われ，その結果，語彙同定タスクの完了が早くなる．しかし，この傾向は核位置が２モーラ目に移動すると有意には見られなくなる．即ち語頭の核に対する特異的な知覚特性として位置づけることができる．この特性を上手く利用することが考えられるが，文中のアク

セント核とその単語が孤立発声された時のアクセント核位置とはずれることが多い．しかし，句頭に着眼すると，そこに核がある場合，孤立発声時も必ず語頭に核がある．即ち，句頭にアクセント核が存在していた場合，その音声区間に対する仮説探索は１型単語に絞って検索を行なえばよいことになる．また，句頭単語は言語的な予測が難しいため，認識タスクの困難度も上昇する．その意味において，従来用いられていなかった新情報源を考えることは至極妥当な方法論である．以上の考察を連続音声認識システム Julius の第二パスに対して実装し，大語彙連続音声認識をタスクとした実験により，句頭のアクセント核に基づく仮説探索制御が有効に働くことを示した．

句レベルの F0 利用に関して，アクセント句境界に同期しながら仮説探索幅（ビーム幅）を動的に制御する方法を提案し，その有効性を示した．仮説探索において，単語終端から，次単語の始端へと遷移する際に言語スコアが加算される．言語スコアの早期の加算を実現するために，木構造化された単語群にファクタリングを施し，単語末尾に至る前の時点から（即ち単語が一意に同定される以前から）言語スコアを部分的に加算することが頻繁に行なわれる．本研究では，アクセント句境界の位置情報を元に，ファクタリング時においても適切な仮説探索幅（ビーム幅）の制御を行なう方式を提案し，認識精度及び認識速度の点において，従来の手法を上回る性能を示すことができた．また，アクセント句境界の位置情報に基づいて，使用する音響モデル（triphone, biphone）の動的選択についても検討を行ない，こちらも従来の方法と比べて精度の点でその優位性を示すことができた．

最後に，言語モデルにおける韻律利用について検討した．ここでもアクセント句境界に着眼し，句境界を越える遷移と越えない遷移の両者について個別に言語モデルを構築し，認識時に，韻律の情報を元に選択的に言語モデルを採用する方法を提案した．提案手法によってパープレキシティーは低下したが，実際の認識時の性能向上は非常に僅かなものであった．アクセント句境界ではなく，文節境界に焦点を絞り，同様の枠組みで言語モデルの高精度化を行ったところ，パープレキシティが約 10%低下し，有効性が認められた.（峯松）

音声の時間構造

音声知覚における韻律の役割の解明を目指し，特に，局所話速に依存した韻律変化とその韻律変化が知覚に及ぼす影響を調べることを目標とした．韻律的特徴から時間的制約を推定し，単語選択等の手がかりとすることを目指している．

局所話速の変動が韻律や調音に大きく影響を及ぼし，音声認識の効率や認識率を低下させていると思われる．HMM 音素モデルの認識率の話速依存性を調べたところ，話速に依存した HMM 音素モデルによって，若干の音素認識率の改善が見られた．音声知覚においても，局所話速などの時間構造とそれに依存する韻律変化が知覚に影響していると考えられる．局所話速に依存した韻律変化を観察するために，話速感覚にあう局所話速の測定方法の検討を行い，特殊拍に対する換算モーラ数と，基準モーラの継続長に対する伸縮率を導入した．また，話速に依存して韻律を制御した音声を用いて聴覚実験を行うために，基本周波数，音素継続時間，パワーなどの韻律情報を制御可能な音声合成システムを開発した．さらに，音素ラベルのないデータベースに対しても，局所的話速の計測を可能とするために，音素アライメントシステムを作成した．データベースの音声に対して基本周波数生成過程パラメータを推定するシステムも作成した．

換算モーラ数を導入することによって，局所話速の測定値が特殊拍からの影響を受けにくくなることを示した．また，この局所話速の測定値と話速感との対応を調べるために，聴覚実験を行った．その結果，換算モーラ数を用いた局所話速値が話速感をよりよく表現していることが示された．特殊拍以外の拍に対しても継続長は伸縮しているため，局所話速の算出方法の改良を行った．３モーラ平均局所話速に対するモーラ継続時間長の分布を求め，基準モーラに対する伸縮率を求めた．この伸縮率を用いた局所話速の算出法により，前後の環境の影響を受けにくい局所話速を求めることがでるようになった．局所話速に依存した韻律変化を調べるために，(a) 文内の局所話速の変化を表すモデル，(b) 局所話速に依存したモーラ継続長制御モデル，さらに，(c) モーラ継続長に依存した音素継続長制御モデルの３階層からなる時間構造モデルを提案した．また，これらのそれぞれのモデルについて時間構造の知覚実験を行い，その有効性を確認した.（吉田）

3.2 発話の構文解析における韻律情報の利用

係り受け解析における韻律情報の利用

発話文の韻律とその統語構造の間には密接な関係のあることが以前から知られている．音声合成におい

ては，統語構造に応じて文音声に適切な韻律を付与することが極めて重要であるため，発話文の統語構造と音素継続長，ポーズ長，基本周波数などの関係が詳しく調べられている．本研究ではその逆問題，すなわち韻律情報を発話文の統語解析に利用する方法を見い出すことを目的としている．

日本語文の統語解析には種々の考え方があるが，ここでは「係り受け解析」という考え方を採用している．係り受け解析の方法も種々考えられるが，本研究では「総ペナルティ最小化法」を使用した．これは，２文節間に係り受け関係（広義の修飾・被修飾関係）が成立することの困難さをペナルティで表し，総ペナルティが最小になる係り受け構造を動的計画法の原理に基づいて効率よく探索するものである．係り受けペナルティには，品詞属性のような離散情報だけでなく，確率や韻律のような連続情報も言語知識として組み込むことができる．もう一つの重要な要素は文節間の「係り受け距離」である．ここでは，着目文節の係り受け距離が与えられたときの韻律的特徴量の条件付分布を学習データから推定し，それから逆に韻律的特徴量が与えられたときの係り受け距離の条件付確率をベイズの定理によって求め，それに基づいて係り受けペナルティを定めている．

音声言語資料としては，ATR503 文データベースに収められている 4 名の話者の発話文を用いた．まず予備実験を行い，どのような韻律的特徴量が有効であるかを調査した．その結果，文節間のポーズ長と基本周波数の有効性が確認されたので，以後はそれらについて更に詳しく調べた．着目文節（その係り受け距離を問題にしている文節）の係り受け距離と，その直後のポーズ長の関係を調べると，係り受け距離が４まではポーズ長平均値が係り受け距離にほぼ比例して増加することが分かった．従って，着目文節の直後のポーズ長は係り受け解析に有効な情報となる可能性がある．実際，これに基づいて係り受け解析を行うと，約 56.4%の文正解率が得られたが，これは文節を構成する形態素情報による言語情報のみを用いて解析した場合の文正解率より，約 7 ポイント高い．

このように，着目文節の直後のポーズ長 p_1 の有効性が確かめられたが，着目文節を中心とするもう少し広い範囲のポーズ長の有効性も予想された．そこで，着目文節の直後の文節の直後のポーズ長 p_2 も取り上げ，解析を行った．まず，p_1 と p_2 の同時分布の利用を試みたが，よい結果は得られなかった．そこで，p_1 と p_2 のそれぞれによって条件付けられた係り受け距離の対数確率を求め，それらの一次結合を用いると，結合係数を適切に調整したとき，p_1 だけを用いた場合より約 2 ポイントの文正解率の向上が見られた．

着目文節の直前のポーズ長 p_0 は，着目文節を発声する前の事象であるから，有効でないように思われる．しかし，p_1 と p_2 の情報に p_0 情報を加えて実験をしてみると p_1 と p_2 だけを用いた場合より文正解率が 0.8 ポイント向上することが確かめられた．このように，着目文節前後のポーズ長情報はすべて係り受け解析に有効であることが明らかになった．

基本周波数情報の利用も試みた．着目文節とその直後の文節の基本周波数曲線から特徴ベクトル f を抽出し，p_0, p_1, p_2, f を用いて係り受け解析を行うと，p_0, p_1, p_2 だけを用いた場合より文正解率が 0.7 ポイント程度向上した．ポーズ長と基本周波数には相関があるため，向上率は大きくはないが，基本周波数情報は係り受け解析に有効であるといえよう．

ここまでの実験は全て特定話者，すなわち，学習データと評価データの話者が同一という条件下で行われたものである．（尾関）

韻律を利用した係り受け解析の多数話者による評価

音声言語資料として，ATR503 文データベースセット B に収められている 10 名の話者と，セット C に収められている 34 名の話者の発話を用い，学習用と評価用の文セットの組合せの種類も大幅に増やして評価を行った．

特に不特定話者条件においては，着目文節直後のポーズ長 p_1 は，係り受け距離が 2 以上の場合は分散値が大きく，距離 3 以上の分布はほとんど重なっているため，分布関数の学習方法には再考の余地がある．p_1 の分布を，係り受け距離が 1 の場合，2 の場合，3 以上の場合の 3 クラスで学習した場合，距離毎に別個に学習した場合に比べて必要なパラメータは 3 分の 1 以下に減少し，かつ，正解率は向上した．また，ポーズ無し $(p_1 = 0)$ の場合とポーズ有り $(p_1 > 0)$ の場合に分割してモデル化する方法を用いたときは，係り受け距離が 1 の場合と 2 以上の場合の 2 クラスで学習したときに最も良い解析精度が得られた．さらに，ポーズ長を平均音節長で正規化することにより，言語情報のみを用いて解析した場合に比べ，文正解率は平均で 5.8 ポイントから 7.4 ポイント向上した．

基本周波数情報についても，特徴ベクトルの分布の学習を簡単化することによって，文正解率が向上した．特徴ベクトル f の要素のうち，着目文節と直

後文節の中央点の差のみを用い，係り受け距離が 1 の場合と 2 以上の場合の 2 クラスで分布関数の推定を行なった場合，言語情報のみを用いて解析した場合に比べ，文正解率は平均で 3.2 ポイントから 4.8 ポイント向上した．

ポーズ情報と基本周波数情報の一次結合を用いると，結合係数を適切に調整したとき，ポーズを単独で用いる場合に比べて文正解率は平均で 1.2 ポイント，基本周波数を単独で用いる場合に比べて平均で 3.8 ポイントの向上が見られた．結合係数の最適値は話者によって異なる値を示した．これは韻律の生成におけるある種の話者性を示していると思われるが，この点についての考察は今後の課題である．(高木)

3.3 音韻情報と韻律情報を利用した認識システム

日本語の高低アクセントは，同音異義語の区別や発話の意味理解にたいへん重要な役割を果たしており，音声認識においてアクセントを考慮することは有効な手段と考えられる．そこで本研究では，連続音声におけるアクセント型の認識手法の開発を目標とした．また，アクセント型は後続単語によって変化することがあるため，アクセント型の情報は認識だけではなく，後続単語の予測にも利用可能である．そこで，アクセント情報を用いて後続単語の予測機能を持つ音声インタフェースを開発することをもう一つの目標とした．

アクセント型の認識は，基本周波数軌跡をパラメータとする．基本周波数は音韻情報に比べて緩やかに変化するので，特徴量を抽出する単位を通常の音声認識よりも大きくとる必要がある．当初は音声認識により音韻アライメントを行って拍を切り出し，この拍を 3 等分した小拍単位を用いていたが，音韻によって極端に長さが異なるため，拍を 3 分割しその中点を中心とした frames 単位を用いることとした．frames の長さに関しては実験的に 40ms〜50ms 程度が適当であることを確認した．この frame 単位で対数基本周波数とその一次，二次差分をパラメータとした．この特徴量を用いて，HMM の状態数を変化させたところ， 0,1 型のアクセントでは 7 状態，N 型では 9 状態が最も高い認識率を示した．

また，アクセント情報の音声インタフェースへの応用としては，ユーザからの入力単語を最初の数音節から予測するディクテーションシステムの開発を行った．部分的な文字列から補完候補集合を得るために，携帯端末での文字入力に用いられている POBox を利用した．部分的な音声認識結果を POBox サーバへ送り，補完候補集合を得る．この補完候補は文字種区切りで管理されているため,「京都工芸繊維大学」のような長い文字列や住所なども候補として出力される．アクセント情報の認識は，携帯端末への実装を目指して，より高速でユーザに適応した学習が容易に行えることから，ニューラルネットワークを用いたアクセント型適合度の判定器を実装した．補完候補単語のアクセント情報を予め登録しておき，入力された音声と最初の 3 拍に関してアクセント型適合度を判定し，そのスコアでソートして補完候補としてユーザに提示する．

HMM を用いたアクセント型の認識実験を行った結果，前後の環境 (高，低，無音) とアクセント型 (0 型，1 型，N 型) を組み合わせた 28 カテゴリーでの認識精度は約 49% (オープン) であり，中心のアクセント型のみに着目した 3 カテゴリー (0 型，1 型，N 型) の分類で約 66%(オープン) であった．クローズドテストではそれぞれ 10%程度精度が向上すること，およびアクセントは個人差・地方差が大きいことから，特定話者向けのアプリケーションでは，アクセント情報を利用することによる性能向上が見込まれることがわかった．

この実験結果を受けて，予測型ディクテーションシステムではユーザがよく入力する単語を登録単語とし，そのアクセント情報を記録しておくことによって，絞込み候補を削減する手法を実現した．本システムにおける絞込み効果を確認するために，アクセント情報の一致度を実験によって調べた．5 人の被験者に対して，先頭単語を共有する 10 単語グループ (地名から始まり，後続単語によってアクセント型が変化するもの．例：京都駅，京都工芸繊維大学)10 組を発声させ，そのうちの各人 20 単語をテストセットとした．学習に用いたデータでの精度は 91%，テストセットでの精度は 83%であった．これらの精度は同一話者のデータのみで得られたものであるが，本アプリケーションは特定の個人用に使われることを想定していることから，ほぼ実用に耐えうる精度であると言える．

最終的な成果として，アクセント情報の認識・ディクテーション・入力補完機能を統合したデモシステムを実装し，音声入力作業におけるアクセント情報利用の有効性を示した．(荒木)

3.4 韻律を利用した講演音声の自動要約

講演や演説などのように話者が目的を持って何かを伝えようとした音声発話に対して，自動的に要約を生成する手法について研究を行なった．文字テキ

ストで与えられた文章に対する自動要約では，単語出現頻度や手がかり語 (cue word) などの言語的情報しか利用できない．一方，音声データは発話速度，イントネーション，発話強度などの韻律的特徴を持っており，このような情報を言語的情報と合わせて利用することによって文の重要度や話者の意図を精度良く推定できる可能性がある．そこで，韻律的特徴を利用した発話の自動要約手法を開発することを目指した．

人間が文章を要約するときには，まず全文を読んで内容を理解してから重要な箇所を取り出し，それを頭の中で再構成し要約を完成させる．しかし，現在の情報処理研究では，十分な意味理解ができるところまで技術が進んでいないため，重要な文の抽出を要約とみなすテキスト要約の研究が広く行なわれてきている．本研究における講演音声の要約においても，講演音声を文単位に分割し，文ごとの重要度を決定することによって重要文を抽出する処理を要約と考える．このような枠組において，講演音声の文単位への分割，および文単位に対する重要度の予測を言語情報と韻律情報を利用して自動的に行なう手法を検討した．

人手で決定した文単位に対して，言語情報と韻律情報を用いて重回帰モデルによって文重要度を予測する手法を提案した．言語情報としては，公開されているテキスト要約システム Posum が生成する文重要度を言語パラメータとして用いた．韻律情報としては，文ごとの基本周波数，パワー，音素時間長の最大値，最小値，平均，レンジの１２個と文の時間長の中から，人手で決定した文重要度との相関が高いパラメータを選択的に用いた．言語情報だけで文重要度を予測する場合に比べて，韻律情報もあわせて利用することにより，モデルの重相関係数および重要文認定度が改善することを示した．特に，連続音声認識によって文テキストを生成した場合には，人手でテキストを書き起こした場合よりも，韻律情報利用による改善の効果が大きいことを明らかにした．韻律情報を用いた場合の重要文認定度として，0.42 を得た．また，基本周波数に関して F0 モデルを導入した分析を行った．重要な文の発声時には，声が高くなることがある．これは，声の高さそのものを考えるより，通常の発声に比べて声が高くなると考える方が自然である．基本周波数に関する特徴量として，文中での F0 そのものではなく，F0 モデルにより通常の発声における F0 を予測し，それとの差をとることで正規化することを考える．このために，文中の文節の平均 F0 を予測するモデルを用いる．文節平均 F0 の予測は，文節の音節数，アクセント型，品詞，境界クラスなどに基づいて数量化 I 類を用いて行った．その結果，文節平均 F0 モデルを用いて正規化した方が文の重要度との関連性が高くなること，文中の文節単位のピッチ最大値よりもピッチ最大値とピッチ最低値の差やピッチ最低値の方が文の重要度との関連性が高くなることがわかった．

講演音声を 300ms のポーズで区切った区間を発話単位とし，発話単位の境界を文境界とすべきかどうかを決定木によって判定する手法を提案した．決定木の入力として，ポーズの前後の品詞クラスによる言語情報と，ポーズの前後の基本周波数およびパワーのパラメータおよびポーズ長による韻律情報を用いた．CART アルゴリズムによって決定木を生成し，94% の分類率を得るとともに，韻律情報を利用することによって，わずかではあるが分類率が改善することを示した.（山下）

4　まとめと今後の展望

4 年間の研究の中で，音声認識・理解における韻律の利用に関して多くの成果が得られた．それらの主なものは次のようにまとめることができる：

(1) 大語彙連続音声認識のための韻律を利用した新しい言語モデルの開発

(2) 聴覚的な話速感とよく一致する局所話速測定法の開発

(3) 日本語文の係り受け解析における韻律の利用技術の開発

(4) 補完機能を持った音声ディクテーションシステムにおける韻律利用法の開発

(5) 講演音声要約システムにおける韻律利用法の開発

この研究を通して，音声認識･理解における韻律の有効性をかなり明らかにすることができたが，韻律現象の複雑さのために，未だその完全な理解には到っていない．今後は更に韻律現象の解明を進めるとともに，得られた知見を総合的に利用することによって，人にとって使いやすい，そして，実環境においても頑健に動作する実用的な音声言語システムの開発へと発展することが期待される．