音声知覚の頑健性 —前語彙的過程を中心として—

Robustness in Pre-Lexical Speech Perception

筧 一彦 Kazuhiko KAKEHI

アブストラクト 人間の音声知覚の頑健性に寄与していると考えられる前語彙^い的な聴覚特性を解説した。まず、音響処理レベルでの特性としてマスキングレベル差、共変調マスキング特性、及び視覚の影響について述べた。続いて音素修復現象を取り上げ、時間的に分散する音素情報の知覚的統合や知覚的統合における話者性情報の関与、同時に話者が存在するときの分散聴取能力や音素特徴の知覚的統合について述べた。調音結合により分散する情報が物理的な音素境界を越えて知覚的に統合され音声知覚の頑健性に寄与する一方、分散聴取能力は意外に低く、いったん音素特徴が抽出されると話者性にかかわらず知覚的統合が起こることを示した。知覚的統合の起きる範囲と関連する音声の知覚単位の言語依存性についても言及し、高次レベルの処理との関連についても考察した。また、話者性に対する音声知覚の頑健性についても触れた。最後にこれらの結果を総合して柔軟で適応的な人間の音声知覚の特徴について述べ、今後の研究方向や機械音声認識との関連について論じ、幾つかの提言を行った。

キーワード 音声知覚、音素特徴、知覚的統合、頑健性

1. はじめに

計算機による機械音声認識が始まったのは、1950～60年代ごろである。当初は、人間がどのようにして音声認識しているかということと機械音声認識システムの構成を考えることは密接につながっていた。例えば、ブラックボードシステムを取り入れた音声認識システムが構築されていたころは、工学研究者も人間の音声知覚過程に関心を持っていた。しかし、1980年代の半ば以降、主にニューラルネットワークやHMM (Hidden Markov Model) に代表される統計的学習法が利用されるようになってからは、学習に使用する音声データベースと統計的学習アルゴリズムが研究の対象となり、人間がどのように音声を知覚・理解しているかという認知科学的研究に対する工学研究者の関心は薄れたように思われる。

計算機の処理能力、記憶容量は年を追うごとに増大し、統計的学習法と相まって機械音声認識システムの適用し得る範囲は拡大した。例えばATRの10年間の成果である音声翻訳システムでは、ホテルの予約や航空会社のカウンターなどの限られた状況での対話だけでなく、海外旅行において生じるような種々の対話をも取り扱える規模のものとなっている。また、自由な話し言葉に比較すれば言語的にも音声的にも変動が少なくなる独話(多数の人を相手に話されるニュース放送、講演、講義など)を対象とした大規模音声データベースを統計的学習データとして用いた話し言葉音声の認識処理システムでは、講演音声に対して話者適応を行うことによって80%程度の単語正解精度が得られている⁽¹⁾。

これらの研究の結果から、統計的学習法を大規模に行ったときのパフォーマンスについてある程度の見通しが得られたといえよう。とはいえこのような大規模な音声コーパスを用いたとしても人間の音声認識能力と機械のそれの間には現状ではまだ大きな開きがあるのも事実である。特に機械音声認識においては、妨害音と話者の個人性、見方を変えると音声の変動要因に対する頑健性に問題が残っている。そこで、この時点でもう一度人間の音声知覚の特質について考えてみることで機械音声認識の将来を考える上での参考になるかもしれない。無論、音声知覚の研究にも進展があるとはいえ、計算機にインプリメントするのに容易なモデルを示すまでには至っていない。ここでは、妨害音など音声に変動を与える要因に対する人間の音声知覚の諸現象を取り上げて考察を加えるにとどめる。

図1に概念が話し手から聞き手に伝達される過程で加わる情報の変動についての概念図を示す。概念が言語化される過程ではその表現の仕方による変動が生じる。次に人間の聞き

図1 音声生成・知覚過程における変動要因の概念図

筧 一彦 正員 中京大学情報理工学部
Kazuhiko KAKEHI, Member (School of Information Science and Technology, Chukyo University, Toyota-shi, 470-0393 Japan)
Fundamentals Review Vol.3 No.1 pp.9-20 2009年7月

手にとっては通常余り問題とならない話者の調音器官や調音に関する個性により変動が増大し、空間に放射された音声は、その音環境によって大きな変動を受ける。また、聞き手の側でも同様に聴覚器官の特性から始まって種々の変動要因が存在すると考えられるが、一般には記号化の過程でほとんどの変動は抑えられるというように理解されている。概念理解のところで変動要因が増加しているのは、聞き手の知識・経験、状況の理解等が話し手のそれとは一般にかなり異なるからである。本稿においては主として変動の大きさが減少すると考えられる過程について扱うこととする。

2. 雑音環境下での音声知覚 (音響処理レベルでの特性)

妨害音のあるところでの音声知覚は人間にとっても容易ではない。しかし、現段階での機械音声認識と比較すれば、人間の方がはるかに妨害音に対する頑健性が大きい。このことを象徴的に示す現象としてカクテルパーティー効果がよく知られている⁽²⁾。この効果が報告されてから極めて多数の研究が行われているが、その実態を解き明かすには至っていない。この効果には多くの要因が関係しており、高次の知識が関係する文脈情報の影響が大きいものと考えられている。このため低次の処理だけでの説明は困難であるが、既に教科書的知識として定着しているものも含め、カクテルパーティー効果に寄与する聴覚特性について以下に述べる

2.1 マスキングレベル差

音響処理レベルでよく知られている現象にはマスキングレベル差 (masking level difference) がある。一般に信号音に対して雑音のレベルを上昇させてゆくと、あるSN比以下のところで信号音が聞こえなくなる。この現象を、雑音による信号音のマスキングという。ところが、もし信号音と雑音の音源方向が異なれば、信号音と雑音の音源が同一方向の場合よりSN比が小さくても信号音が聴取できるようになる。この信号音と雑音が同一方向の場合と別方向の場合のマスキングが生じるレベルの差をマスキングレベル差という。この効果のデモンストレーションとして以下の例がよく知られている。

まず、片耳に信号音と雑音を加え、雑音のレベルを増加して、ちょうど信号音聞こえなくなるレベルに設定しておく。次いで、もう一方の耳に同レベルで位相も同じ雑音のみを与えると、雑音は正面方向に、信号は元の耳の側に定位するため、再び信号音が聞こえるようになるというものである。信号音が正弦波信号 (純音) の場合、信号音の周波数、二つの音源方向の開きなどによりマスキングレベル差は異なるものの、一般に3~15 dB程度になる。

また、純音だけでなく音声に対する効果の検討も行われており、SN比と情報マスキングの両者の観点からの検討が行われている⁽³⁾。

2.2 共変調マスキング解除

音響レベルでの聴覚情報処理を考える際の基本的な概念はマスキングと臨界帯域である。工学的なイメージでは入力音響信号は一連の帯域フィルタ (フィルタの帯域幅=臨界帯域幅) によって分析を受け、各フィルタの出力がさらなる処理を受けると考えられる。そこで基本的には同一臨界帯域内にある信号は、同じような処理を受けると考えてよい。ただし、この一連のフィルタは中心周波数が固定されたもので帯域フィルタ群によって連続的に可聴信号範囲をカバーしていると考えるのは妥当でない。人間の音響信号処理においては、その目的に応じて帯域フィルタの中心周波数は適応的に設定され、利用されるということである。

マスキングレベルは、ある臨界帯域内にある信号音の妨害音成分が、その帯域内にある信号成分のマスキングに関与すると考えられる。しかし妨害音の性質によっては、このマスキングレベルが小さくなる現象が知られている。ある臨界帯域内にのみ成分を持つ雑音のレベルを上昇させて信号音をマスキングする。このときごくゆっくり (数十Hz程度) の振幅変調を雑音に与えておき、単位周波数辺りの雑音レベルは保ったまま雑音の周波数幅を臨界帯域を超えて拡大する。このとき全体のSN比は低下するにもかかわらず、信号が再び聞こえるようになる。この現象を共変調マスキング解除 (CMMR: Co-Modulation Masking Release) と呼ぶ。同一音源からの妨害音は、多くの場合その雑音の各周波数帯域成分のレベル間に一定の相関関係が存在することが多いので、共変調マスキング解除はその関係を利用して信号音の聴取を容易にするような仕組みである。この共変調マスキング解除量は、種々の条件によって異なるものの大きいときには11~14dBにも達する。

2.3 マルチモダリティ

以上のような聴覚特性はカクテルパーティー効果を代表とする音声知覚の頑健性を実現する機序の一部を成しているものと考えられるが、その一部分を説明するものにすぎない。

例えば、一般的に人間が二つの耳を使うことによって実現する音源方向の定位に対しては、一般的に視覚情報の方が優位に働くことが知られている。広い講演会場で、自分の近くにあるスピーカから講師の声が聞こえているにもかかわらず、演壇上で話をしている講師の方から声が聞こえてくるように感じることはよくあることである。通常前述したマスキングレベル差が生じるためには、実際に信号音と雑音の音源方向が異なる必要がある。しかし、たとえ信号と雑音の音源方向が全く同一場所 (同じスピーカから出る) であっても、話者像を音源と同一位置に呈示する場合よりも音源から離れた位置に呈示の方が音声に対する復唱成績が良くなる⁽⁴⁾。このように、単純な音響信号処理の観点だけからでは解釈することが難しい不思議な現象は、人間の音声知覚の柔軟な適応性を解明する上での手掛かりとなるものであろう。

一般的な会話状況においては、相手の身振りや口の動きといったものも音声の聴き取りに利用されている。音声情報のない口の動きだけでは、会話の内容を把握するには十分ではないが、状況が加われば会話の内容を特定することは容易になる。例えば野球のピッチャーとキャッチャーがマウンド上で会話するような場合、口元が見えないようにグローブで隠しているのはそのためである。このように文脈や状況からそのときに行われる会話の内容が極めて限定されるような場合には、口の動きだけでも（音声）認識が可能となる。

聴覚障害者が用いるリップリーディングは、かなりの学習が必要であるが、そのような学習の経験が全くない一般の人でも口唇の動きの視覚情報は、音声の知覚に大きな影響を与える。これはマガーク効果として知られている⁽⁵⁾。マガーク効果は視覚情報と聴覚情報が矛盾した場合には聴覚による音声知覚が不正確になり異聴が起きる現象をいうが、逆に聴覚情報と音声情報が整合している場合には、雑音下での（音声）認識を強化するものとなる。日本語母語話者を対象にしたマガーク効果の検討結果では音声雑音による妨害を受けていないときには、米語母語話者よりマガーク効果は生じにくい傾向を持つが、雑音に加わった状態では、むしろ生じやすくなることが知られている。これは雑音下では積極的に視覚情報を利用することを意味しているものであろう。

視覚情報と聴覚情報の統合がどのようにして行われるかについての詳細は分かっていないが、マガーク効果については多数の研究が存在する。その一つとして視覚情報と聴覚情報の時間的同期をかなりずらせてもマガーク効果が生起することから統合処理自体はかなり上位のレベルではないかと推測される。

以上述べた音源の方向情報や聴覚の臨界帯域フィルタ間の信号の相関情報は、工学的な音響信号処理の中で利用され得る情報である。工学的には人間のように耳が2個というような制約はないので、3個以上の空間的受音点を持つことにより、優れた目的音の抽出が容易になる可能性は大きい。また、独立成分分析のように優れた雑音抑圧機能を持つような方法も実用化されつつある。しかし頑健な機械音声認識を実現するためには、従来単に前処理として切り離して考えられている音響処理を、より後段の高次処理と融合したシステムとして考える必要があると考えられる。

図2 ボトムアップな音声知覚過程

3. 音素知覚

音素知覚の諸現象を説明するにあたって、大まかな音素知覚過程のボトムアップ過程の概念を図2に示しておく。まず「音声/音響信号の入力」については既に説明したような種々の低次の処理が関係している。次の「音源の同定」には、一部両耳による音空間知覚が関与しているが、音源の同定には単に空間情報のみでなく話者性などの種々の情報が関与している。本章では「音源の同定」以降の、情報の統合と分離に焦点を当て、主に音素知覚における諸現象を取り上げる。

3.1 情報の知覚的統合と分離

複数の音源からの情報が重畳しているとき、それらの情報を適切に切り分けなければ、元の音源情報を正確に知覚することは困難である。ところが、複数の音源から発せられた音情報が重畳されている場合、その分離にはほとんどの場合あいまい性が存在する。我々人間の知覚過程は、現実世界における音入力（解釈が一意に定まらない）から、われわれ人間の生存にとって不利にならないような外界の解釈を作り出す過程であるといえることができる。ここで重要な点は、最もらしい解釈が、物理的に正しいかどうかということと同義ではなく、また処理時間も次に必要な行動を起こすために制限されているということである。

よく知られている例としては図3に示すような上昇FM信号音の $t_1 \sim t_2$ の間を $F_1 \sim F_2$ までの周波数成分を持つ雑音で置換すると、FM上昇音は $t_1 \sim t_2$ 間で途切れたような知覚は起きず、連続して上昇していくように聞こえる（解釈する）現象がある。これを視覚的な図として眺めた場合にも2本の直線があるという見方より、連続した1本の直線がその前に置かれた長方形によって隠されているという見え方の方が一般的である。視覚現象の場合は、前に置かれたものが不透明であることが、聴覚的現象の場合には、雑音成分中に $t_1 \sim t_2$ 間のFM上昇音の周波数成分が含まれていることが、この解釈が生じるために必要である。

上記の現象は、FM上昇音と雑音という二つの音事象に分離解釈される場合に、どの成分をどちらの音事象に組み込んで解釈するかにかかわるものである。一般に音の流れを、一つのものとして解釈した方が妥当性が高いか、二つまたはそ

図3 FM上昇音の連続的知覚

図4 VCV 音声波形 (Cが無声閉鎖子音の場合)

れ以上のものとして解釈した方が妥当性が高いかは、そのま
 とまりの良さに依存している。高音部と低音部のメロディー
 の音符を交互に演奏するとき、高音部と低音部の音の高さ(周
 波数)が近ければ一つのメロディーとして聞こえるが、周波
 数が離れていると二つのメロディーに分離して聞こえる。こ
 のような現象は音の分凝と呼ばれる。妨害音が加わった音声
 の知覚においてはこのような知覚システムが強く働いている
 と考えられる。

3.2 音素修復現象 (Phonemic Restoration)

音素修復現象とは、前項で述べたFM上昇音の連続的知覚
 の場合と類似の現象である。すなわち連続音声の一部を削
 除してしまい雑音で置き換えても(以降雑音置換という)削除
 された部分に相当する音素が聞こえる現象のことである⁽⁶⁾。
 例えば図4 (a) は、/ape/という発声の波形を示している。最
 初の母音/a/に続いて出現する/p/の音素は破裂子音であるの
 で、その発声前にはいったん両唇が閉鎖する。このため破

裂が起こる前の閉鎖区間に無音が生じる。この閉鎖区間の終
 了後(破裂の開始時点)から/e/という母音の定常的な部分に
 至るまでの70msという長い範囲を雑音で置換しても(図5)、
 /ape/という明確な音声知覚が生じる。

雑音置換された部分が修復されて知覚されるという意味に
 おいては、音素修復現象は、FM上昇音の連続的知覚と類似
 であるが、この音素修復現象の例では、/p/という子音に関す
 る情報が物理的には全く取り除かれてしまっているように見
 えるので、全く同じであるとはいえない。言い換えると/p/の
 破裂が起こる前に無音の閉鎖区間があり、破裂の開始時点は
 明確で、その後/p/の情報が存在していると通常考えられる部
 分の情報を完全に雑音で置換したと想定できるにもかかわらず、
 /ape/の知覚が生じるからである。

この現象を解明するために以下の二つの方法を組合わせて
 用いた実験が行われた⁽⁷⁾。一つは上述した雑音置換であり、
 もう一つはクロスプライシング(cross splicing)である。今、
 母音間閉鎖子音として3個の無声破裂子音/p/, /t/, /k/を対象
 とするとそれらの子音すべては閉鎖区間を持つので閉鎖前部

図5 NORMAL 刺激に対する閉鎖後部の雑音置換

と閉鎖後部を閉鎖部(無音区間)で切り離して異なる閉鎖子音の閉鎖後部とつぎはぎした新たな音声刺激を容易に作ることができる(図4(b))。このように“つぎはぎ”することをクロスプライシングと呼ぶ。またこのようにして作成した刺激をVC₁-C₂Vのように書くこととする。VC₁は、後続する子音がC₁である閉鎖前部の母音を意味し、C₂Vは子音C₂+母音Vから成る閉鎖後部を示す。したがってこの場合、元の閉鎖前部に後続していた子音C₁と実際に構成された刺激の閉鎖後部の子音C₂が異なることを示している。このような刺激に加えて、閉鎖前部を単独発声した母音と切り張りした刺激V-CVと閉鎖後部を切り出した刺激CV(post)を加えた実験刺激を用意する。この実験刺激に雑音置換を適用した刺激を作成し、知覚実験を行う。雑音置換長は0ms(雑音置換なし)から70msまで10msステップである(図5)。

知覚実験の結果を図6に示す。クロスプライシングしていない刺激VC₁-C₁V(NORMAL)に対する結果をみると後部閉鎖区間の終了時点から70ms雑音置換を行っても閉鎖子音の知覚はよく保たれており、音素修復率は85%である。一方単独発声の母音と後部閉鎖区間を組合せたV-CV刺激では、その修復率は雑音置換長の増大に伴って急激に低下して行き、雑音置換長が70msでは40%程度まで下がっている。こ

図6 雑音置換長を変化させたときの子音修復率

の結果は、閉鎖前部の母音中に何ら閉鎖子音の情報が含まれていなければ、①閉鎖後部に含まれる子音情報のみに基づいて子音知覚が起ること、②雑音置換長が70msになると閉鎖後部に含まれる子音情報はほとんど失われるということを示している。

閉鎖前部と閉鎖後部に含まれる子音情報が異なると考えられるVC₁-C₂V刺激に対しては、その閉鎖後部を雑音置換すると雑音置換長の増大に伴って、急激にC₂に対する音素修復率は低下する(VC₁-C₂V (C₂))。この低下の度合いはV-CV刺激の場合よりも激しい。しかしC₁と回答した割合が雑音置換長の増加に伴って増大する(VC₁-C₂V (C₁))。このC₁とC₂という回答を足せばNORMAL刺激の場合の音素修復率とほとんど同じになることは興味深い。すなわち、雑音置換が行われると刺激中に存在する他の閉鎖子音の手掛かりが利用されるということを示している。

V-CV刺激に対する音素修復率は、閉鎖後部の情報のみを切り出したCV (POST) 刺激に対する音素修復率より全体的に高い。これは、V-CV刺激では閉鎖子音であることを示す閉鎖区間の情報が残されており、後部閉鎖区間の情報が雑音置換されても無声閉鎖子音であるというコンテキスト(文脈)情報が存在するために、閉鎖後部で雑音置換されていない残りの子音情報がうまく利用可能になるものと考えられる。このように実験の結果をやや詳細に分析してみると雑音などによって妨害を受けた音声の知覚においては、残存する情報を最大限に利用していることが分かる。

閉鎖前部に閉鎖子音情報が含まれていることを子音知覚の結果から間接的に示した。これは、連続的に音声を生成するときに発声器官の調音運動は前の音素と後ろの音素の影響を受けるためである。つまり、単独で音素ないし音節を発声したときと連続発声の中のそれとは音響的特徴は異なる。この現象は調音結合と呼ばれ、よく知られている。

この調音結合があるため、音素情報は比較的広い時間範囲に分布している。図4(a)にみられる母音間閉鎖子音のように、物理的には閉鎖区間という明確な音素境界を持つような

図7 知覚実験結果に基づく閉鎖子音情報の時間的分散

図8 話者情報および音素情報に競合があるような実験刺激

ものであっても、その境界の前にある閉鎖前部の母音の中には、当該閉鎖子音を示す情報が含まれており、知覚的にもそれを利用することができることを示しているのが先の実験結果である。

知覚実験の結果などから求めた母音間閉鎖子音の音素特徴の時間的分布を図にまとめたものが図7である。調音結合の結果少なくとも3音素のスパンにわたって真中の音素の情報が含まれているので、もし妨害音がその一部分をマスキングしても音素修復現象により当該音素の知覚が生じる。したがってこの現象は非定常的な妨害音が存在する場合の頑健な音声知覚に大きく寄与しているものと考えられる。

3. 3 同時的あるいは時間的に分散した情報の知覚的統合

3. 3. 1 同時的に存在する音素特徴の知覚的統合

同時的に存在する音素特徴の統合に関するものにDPS (Duplex Perception of Speech) がある⁽⁸⁾。これは、複数の音素特徴から成る刺激の一部を片耳に残りの部分を反対耳に分けて呈示する (dichotic 呈示) 手法を用いて、音素特徴の知覚的統合が起きるかどうかを検討したものである。例えば一方の耳に /da/ と /ga/ の共通のフォルマント成分 (F1, F2 の遷移部分と定常部分, 第3フォルマントの定常部分) を呈示し、反対耳に第3フォルマントの遷移部分を呈示すると、両者は知覚的に統合されて第3フォルマントの遷移部分が下降するか上昇するかに応じて /da/ または /ga/ が知覚される。このとき、同時に反対耳に呈示されたフォルマントの遷移部分に対応した音がチャープとして知覚される。このDPS現象では、共通成分のフォルマント周波数を構成する基本周波数と反対耳に呈示した遷移フォルマントを構成する基本周波数を異なるものにしても知覚的統合が起り、上記と同じ音素知覚が生じる。また、反対耳に呈示した第3フォルマント遷移部分の終端のフォルマント周波数が第3フォルマントの定常部開始部分のそれと異なっていて、そこに不連続があっても同様の音

素知覚が生じることが示されている⁽⁸⁾。また、/da/ を特徴付ける二つのフォルマントを互いに異なる基本周波数で構成して、それぞれを別耳に呈示 (dichotic 受聴) すると受聴者は多くの場合、別々の二つの音があると報告するが、両者のフォルマント情報が知覚的に統合され /da/ という知覚も生じるという結果もある⁽⁹⁾。

3. 3. 2 継時的に複数話者が存在するときの音素特徴の統合

前項では、時間的に分散する音素情報を統合して妨害音に対して頑健な音声知覚が生じる現象を示したが、このような統合は音素特徴の分析・抽出が行われれば自動的に働くものかどうかということが問題となる。カクテルパーティー効果を取り上げたときに言及したように、複数の話者が存在しても注意を向けた話者の音声だけを聞くことができるのであろうか。

そこで極端な場合として、男性話者と女性話者が発声した母音間閉鎖子音を含む VCV (母音+無声閉鎖子音+母音) を用いて図8に示すような話者情報と音素情報に競合があるような刺激を作成し、雑音置換を用いて知覚実験を行った⁽¹⁰⁾。例えば、刺激 (2) は、前の実験でも用いた子音情報のみが閉鎖前部と閉鎖後部で競合 (Phoneme conflict) するような刺激であるが、刺激 (3) は、話者情報のみが競合 (Talker conflict) するような刺激であり、刺激 (4) は話者情報および音素情報の二つともが競合 (Talker・Phoneme conflict) するような刺激である。

図9は、そのような刺激の閉鎖後部を雑音置換したときの知覚実験結果を示している。音素・話者両方の情報が競合する刺激 (4) の結果と音素情報のみが競合する刺激 (2) の場合を比較すると音素修復の効果に差はなかった。図9では (1) - (3) の結果のみを示している。図9から刺激 (1) Normal と刺激 (3) Talker-conflict の知覚実験結果を見ると音素修復の効果には全く差がない一方、(1) Normal と (2) Phoneme Conflict の間には明りょうな差が見られることが分かる。

この結果は、抽出された音素特徴には話者性のラベルが付

図9 話者情報及び音素情報に競合がある場合の母音間閉鎖子音の知覚

いていない、すなわち話者が異なるような場合でも時間的に分散する音素特徴の統合が行われるということを示している。前項の結果とこの実験結果は合成音声によるものであるが、音源分離の重要な手掛かりである音源方向 (dichotic 呈示) と話者性 (基本周波数) の両方に相違があっても、一旦音素特徴が抽出されると知覚的統合が起こり、音素知覚が生じることを示唆している。

しかし、これはカクテルパーティー効果を説明するという観点からは都合が悪い結果のように見える。そこで、今まで対象としてきた無声閉鎖子音だけではなく有声閉鎖子音及び促音に範囲を拡大して検討を行ってみた⁽¹¹⁾。

/ape/ のような母音間無声閉鎖子音を含む VCV で、その無音の閉鎖区間の長さを縮めて行くと一般には /p/ の知覚が有声化して /b/ の知覚が起きる。一般に /abe/ と発声された音声の閉鎖区間は、/ape/ と発声されたときのそれよりも短く、/ape/ の発声の閉鎖区間を縮めた場合、/abe/ の閉鎖区間の平均的な値の付近で /ape/ の /p/ の有声化知覚が最も生じやすい。この現象を利用して閉鎖前部と閉鎖後部の話者が異なる刺激を作り、閉鎖時間を短くしていったときに有声化知覚が起きるかどうかを調べる。また、/ape/ の閉鎖区間長を長くして行くと /appe/ のような促音化知覚が起こる。これらの刺激は、単に閉鎖区間長をもとの発声から変化させるだけで、音声区間の信号には全く手を加えないで実現できる。

このような刺激を用いた知覚実験の結果を図 10 に示す。図 10 (a) は、話者情報が競合しない、すなわち同一発声者の音声に対して閉鎖区間長を変化させたときの知覚結果である。白抜きの丸印と黒丸印は、それぞれ話者競合の刺激を作るために用いた各話者の音声に対応する。この結果は、上述した閉鎖区間長の伸縮による知覚現象が生じることを示している。次にこの 2 話者の発声をクロスプライシングして話者情報が競合する刺激を作成し、その閉鎖区間長を伸縮して行った知覚実験の結果が図 10 (b) に示されている。図 10 (a)

図 10 閉鎖区間長を変化させたときの有声化及び促音化知覚の変化

と (b) の結果を比較すれば、閉鎖区間長を長くすると促音知覚の生起には両者で変化がないものの話者情報が競合する場合には閉鎖区間を短くしても有声化の知覚が生じていないことが分かる。言い換えれば、有声音と無声音の識別特徴として使われる閉鎖区間の長さの抽出には、閉鎖前部の話者と閉鎖後部の話者が同一である必要があるが、促音化の知覚においては前後の話者に関係せず、ある一定の閉鎖時間長があればよいことを示している。

この結果は一見奇妙であるが、この現象が日本語音声の知覚単位と密接に関係している証拠である。音声の知覚単位については後述するが、日本語音声の知覚単位はモーラ (拍) である。促音は表記の際には小さい「っ」と書かれるがこれは 1 拍を形成している。したがって促音は前後の話者には関係なく、ある一定の無音区間長があればその知覚が起きると考えられる。一方、閉鎖子音が有声か無声かの知覚に利用される無音区間は、1 拍中の音素の決定に関与するため、あくまで閉鎖区間の前後の話者が同一である必要があると考えられる。すなわち、同じ閉鎖区間長を特徴として利用する知覚であっても、知覚単位との関係からその差が生じたものと考えられる。この解釈は特徴の抽出段階における話者性情報の条件と考えれば、先の抽出された音素特徴には話者性のラベルがないように見えるという解釈と一応の整合性があるということになる。

3. 3. 3 同時に複数話者が存在するときの音素情報の統合

話者性情報が音素の特徴抽出に際して利用される場合があ

図11 話者の同時競合刺激

ることが分かったが、いったん抽出された音素特徴の統合にかかわる過程では、音素特徴に話者性のラベルは付与されていないか、付与されていたとしても特徴の統合に際しては、話者性情報は利用されないように見える。しかし、上で紹介した知覚実験では、複数話者が同時に発話している状況ではなく、継時的に話者が異なるという条件であった。これは、カクテルパーティー効果のように同時に複数の話者が発話している中で特定の一人の話者の音声を聴き取るという状況とはやや異なる。同時に複数話者が存在している場合における、時間的に分散する音素情報の知覚的統合について調べれば、カクテルパーティー効果に近い状況の聴き取りを明らかにすることができるはずである。

次に示す実験⁽¹²⁾の基本的枠組みは、これまでと同様の母音間閉鎖子音の知覚実験であるが、図11に示すような刺激を用いたものである。この図では、VCV (Cは閉鎖子音) 音声波形の振幅包絡を模式的に示している。Case1の場合もCase2の場合も閉鎖前部は二人の話者が同時に発話している。Case1では、話者Aについての音素情報は競合していないが、話者Bの音素情報は話者Aによる閉鎖後部の音素情報と競合している。Case2の場合は、話者Aの音素情報は、閉鎖前部と閉鎖後部で競合しているが、話者Bの閉鎖前部と話者Aの閉鎖後部の音素情報は一貫している。また、同一話者の音素情報の方が優先的に統合されるかどうかを見るために、閉鎖前部の話者Aと話者Bの音声レベルの割合を変化させた。ここでは、通常生起する自然な状況では重畳音声の一方のレベルが下がることが多いので、話者Aのレベルを下げるか話者Bのレベルを下げるかして両者のレベルの割合を変化させた。そこで、話者Bが話者Aに対する妨害音になっているとみなして、

(話者Aの閉鎖前部の音声レベル) /

(話者Bの閉鎖後部の音声レベル)

の比を $-\infty \sim \infty$ dBまで変化させる。すなわち $-\infty$ dBでは、話者Aの音声はなく、 ∞ dBでは、話者Bの音声はない。また、0 dBのときは、話者Aと話者Bの音声レベルが等しいことを示している。

このような刺激において、閉鎖後部を雑音置換しない場合には、レベル比が $-\infty$ から ∞ dBまで変化しても、閉鎖子音の知覚はすべて閉鎖後部の子音情報によって決まる。これは、前述のように閉鎖後部の開始点から40ms程度までにある情報が、破裂子音の知覚を決定する最も強い情報になっているからである。そこで、閉鎖後部の開始点から70msを雑音置換した刺激音声を用いて、時間的に分散する閉鎖子音情報の

知覚的統合において話者性情報がどのように関与するかを知覚実験によって検討した。更に話者性の相違を見るために話者AとBが異なる女性話者である場合と話者Aが男性、話者Bが女性話者であるような極端な場合の二つの条件で実験した。

話者AとBが共に女性である場合の結果を図12に示す。四角印(□, ■)が話者Aの音素情報は閉鎖前部と閉鎖後部で一貫し、話者Bの閉鎖前部と閉鎖後部の音素情報が競合するCase1条件のときの知覚結果を示している。また、丸印(○, ●)がCase2の条件の知覚結果である。

この結果は大変明りょうであり、Case1でもCase2でも閉鎖子音情報の知覚的統合は閉鎖前部における話者Aと話者Bの音声レベル差のみによって決まるということを意味している。ただし、閉鎖前部の子音(C₁)の知覚と閉鎖後部の子音(C₂)の知覚が同じになる話者Aと話者Bの閉鎖前部におけるレベル比は、0 dBではなく、数dBだけ正側にシフトしている。つまり話者A優位側にシフトしており、話者性が一致している方の音素特徴がわずかに統合されやすいようにみえるが、その量は小さい。

また、話者性が明確に異なる、男性話者と女性話者をそれぞれ話者A、話者Bとした結果が図13 (a), (b)である。図13 (a)は、Case1の条件で、話者Aが男声、話者Bが女声の組合せとその逆の場合である。(b)は同様にCase2の条件で、話者Aと話者Bの組合せが(a)と同じように入れ替わる場合である。

図13 (a)、すなわちCase1の場合には、子音情報が一貫する話者が、男性であっても女性であっても閉鎖子音情報の知覚的統合にはほとんど影響がない。しかし、図13 (b)、すなわちCase2の場合においては、話者Aが男性、話者Bが女性

図12 2話者重畳の場合の時間的に分散した音素情報の知覚的統合(話者Aと話者Bはともに女性)

の組合せとその逆の組合せの場合で知覚結果にはかなりの差がある。

話者Aが女性の場合には明らかに話者Aの持つ子音情報のほうが知覚的に統合されやすく、逆に話者Bが女性の場合にも話者Bの子音情報の方が知覚的に統合されやすい。このことは、前後の話者性情報の一貫性は子音情報の知覚的統合には余り関係がないということ、男声と女声ではその発声レベルの比が同じであっても女声の音素特徴の方が抽出されやすいことを示している。一般にマスキング現象は低い音のほうが高い音をマスクしやすくなることが知られているが、この場合には当てはまらない。各周波数帯域でのレベル比など詳細を検討する必要がある。

4. 音源の分離と同定

音素知覚における情報の知覚的統合という視点からの検討について述べてきたが、VCV音声という限定された範囲では、音素特徴の抽出においては話者性の情報は利用されるものの、音素特徴の統合過程においては話者性の情報は利用されないように見える。このことはカクテルパーティー効果を考える上でも日常的な音声聴取の直感ともやや異なるように感じられる。

聖徳太子は同時に10人の訴えを聴くことができたといわれるが、複数話者の重畳された単語音声を聞いて、その内容ではなく話者数を何人までなら答えることができるかという検討がある⁽¹³⁾。結果は意外なことに2話者までで、3話者になると正解率は3～4割程度になってしまう。この場合話者性の相違(男声、女声の組合せ)や音声の到来方向はほとんど影

響がないとされている。

この結果は、複数話者重畳音声の中にターゲットとする特定の話者があるかどうかの検出(分散聴取)とは異なるものである。単耳呈示の場合における分散聴取では、単語音声、文音声によらず、知覚の限界はおおよそ話者数2名程度との報告がある⁽¹⁴⁾。ただし、複数話者重畳音声の分散聴取で、事前にターゲット音声を呈示する方が重畳音声の事後にターゲット音声を呈示するよりも正解率が有意に高かったことから、事前に聞くべき話者の情報がある場合は事後に聴取する場合と聴取の方略が異なっていると考えられる。この結果は知覚の初期過程のみでなく高次の処理過程も含んだものであることに注意が必要である。

前章の結果と併せて考えると、複数話者の分散聴取は人間にとっても困難なタスクであり、10人はおろか2人が知覚の限界である。また、音素特徴の統合に当たっては特徴抽出が容易な方が利用されるにすぎないように思われる。しかし、ここで注意すべきことは、実験におけるすべての話者重畳条件において各話者の発話の開始時点がそろっていることである。実際に複数話者が個々に話をしているような状況では、開始時点が全くそろっているようなことはまれである。音楽の合奏の場合に楽譜上では同時に演奏するように指示されているものでも、実際の演奏を分析すると音の立上がりは正確には同期していないとの指摘がある。立上がり時点のずれが30ms以内の2音は同時に鳴り始めたように聞こえると報告されている。音の事象ないし流れの分離にはこの時間のずれが重要な手掛かりとなっているために、これまでの話者重畳条件の実験は音源の分離が一番厳しい条件においてのものであった可能性がある。

5. 音声の知覚単位と言語的制約

時間的に分散する音素特徴の知覚的統合において、直近のモーラに含まれる音素特徴が利用されることを示した。音声学的な単位としては音素が最小の単位であり心理的な実在性もあるといわれているが、果たして音素の系列として連続音声は知覚されているのか、あるいはより大きな単位で知覚が生じるが、そこから二次的に音素系列の知覚表象を作り出しているのかについての検討は、知覚的統合の時間的範囲を考える上でも重要である⁽¹⁵⁾。

1970年代の初頭にSavinとBeverは無意味単語音声の中のターゲット音の検出反応時間を測定し、音節が音声知覚の基本単位であることを報告した。この結果に対しては、実験方法など種々の問題が指摘された。その後1980年代の初めにMehlerらは新しいシラブル・モニタリング手法(pa-pal課題)を提案し、シラブル・タイム(シラブルを単位としたリズムを持つ)なフランス語音声の知覚の基本単位は音節であることを明らかにした⁽¹⁶⁾。この手法を用いて日本語を含む多くの言語音声に対する検討が行われた。その結果モーラ・タイム(モーラを単位としたリズムを持つ)な日本語においてはモーラが知覚の基本単位であることが示されている⁽¹⁷⁾。ちなみに英語で

図13 2話者重畳の場合の時間的に分散した音素情報の知覚的統合(話者Aと話者Bの性別が異なる場合)

は、シラブルが基本単位ではなく、ストレス・タイム(ストレスを単位としたリズムを持つ)が知覚の基本単位に強く関係していることが示されている⁽¹⁸⁾。これらの知覚単位はクロスリンギスティックな実験を通じて聴者の母語に依存していることが分かっている。

この知覚単位に関連した現象として、日本語(東京方言)の無声化(devoicing)子音の知覚現象がある。東京方言では、無声子音に挟まれた母音/u/や/i/は無声化され、その音響的部分はなくなることが多い。しかし聞き手は多くの場合母音が存在しているように知覚している。これは音素修復現象のところで述べた調音結合の結果ではないことが実験的に確かめられている⁽¹⁹⁾。

また、母音間閉鎖子音のクロスプライミング刺激を用い、日本語・オランダ語母語話者に対してVCV音声の知覚実験を行った結果から、日本語母語話者は同じ子音特徴を持つ刺激に対して、その音素環境が日本語のphonotacticな制約(音韻連鎖規則)に合致していれば閉鎖子音の音素知覚を生じ、合致していない場合には音素知覚が生じないという結果が得られている。日本語母語話者は、音素特徴の抽出はできるが、音韻連鎖規則を満たさない場合には、その音素特徴を利用した子音知覚が生じないという点が興味深い。一方、オランダ語においては、VCCVやVCのような連鎖は許容される音韻連鎖規則であるので、日本語母語話者には知覚できない子音もオランダ語母語話者にとっては知覚可能となる⁽²⁰⁾。

無声化母音の知覚は、情報が無いのに音素知覚が生じ、音韻連鎖規則に基づく制約は情報が利用し得るのに知覚が生じないという例であり、これらは前語彙的(pre-lexical)段階で起こることが確かめられている⁽²¹⁾。この現象は一種の錯聴とも見ることができるが、妨害音中にある言語音声を取り出すには、候補の可能性を減少させるという意味で有効に働くものと考えられる。

一方音声の生成と知覚は表裏一体の関係にあると考えられ、最近のサルにおけるミラーニューロンの発見や人間が音声を聞くと発声に関連する運動野が活性化するという観測結果はこのことを裏付けている。単語の読み上げ課題におけるプライミング効果や復唱課題における反応時間から、音声生成の単位が検討されている。フランス語においてはシラブルの効果が大きく、英語では課題によって単位の効果が異なる結果が得られている。日本語においては最近の研究からモーラ単位にプライミング効果があることが確かめられている⁽²²⁾。これらの結果は音声の知覚的側面から見た単位の結果と整合的である。

6. 話者適応

機械音声認識(Automatic Speech Recognition)においては、妨害音があると認識率は大きく低下する。また、話者による個人性は機械音声認識における大きな課題であり続けている。人間はこのような話者性には比較的影響を受けずに音声認識が可能である。実際に母語話者であれば日常生活の中で、

話者の違いによって音声の聞き取りやすさが変化することに気付くことは少ないが、第2言語の聞き取りでは著しく聞きにくい話者が存在することはよく経験することである。

本稿では主として妨害音に対する人間の音声知覚の頑健性に焦点を当てて述べてきた。ここまでのところでは、事前に話者情報がある場合に分散聴取能力が向上することを除いて話者性は妨害音に対する音素知覚の頑健性にはほとんど関係していないという結果であったが、現状の機械音声認識と比較するなら、母語話者の話者性に対する音声知覚の頑健性は驚異的である。

話者の個性に対して人間の音声認識がどのように働いているかについては、speaker normalizationという視点から主として母音を対象として非常に多くの研究が行われてきた。妨害音に対する音声知覚の頑健性という観点とは異なるので本稿ではその詳細には立ち入らないが、話者性に対する多様な実験結果を説明するのにVocal tract normalization theory(声道標準化理論)よりもExemplar-based memory model(事例ベース記憶モデル)の方が優れていると考えられている。後者のアプローチでは、認知的カテゴリーは、カテゴリーに内在する構造を取り除いた抽象的な標準化表象としてよりも経験したカテゴリー事例が認知的表象として蓄積されたものの集合として表されていると考えられている⁽²³⁾。

一方、母音の知覚における話者に対する頑健性(認知の不変性、恒常性)は刺激群のコントラストを用いた情報処理に基づくという考えをベースにして、母音の知覚は、連続かつ可逆な変換に対して不変となるコントラスト量に基づく構造的表象によっているという考えが提唱されている⁽²⁴⁾。

もし、話者への適応効果が顕著にあるならば、この構造的表象の考え方は、話者性に対する頑健性の説明モデルとなる。実際には、話者への適応効果は、かなり妨害音の大きい状況でしか観測できない。一つの実験結果によると、単音明りょう度が70%程度になるまで雑音を加えたような音声では、同一話者の音声を5音節程度聞くと話者適応が起こり、明りょう度は4~5%向上するが、異なる話者の音声を6~7音節聴くといった獲得された話者適応効果は消滅することが分かっている⁽²⁵⁾。この結果は構造的表象の存在を支持しているが、このように聴取条件が悪くなければ、単音節ごとに話者が異なる音声を聞いても明りょう度はほぼ100%になるので、構造的表象のみを利用して音声知覚が行われているわけではないと考えられる。

また、自分の発声を録音再生して聞くと自分の声ではないように聞こえる自声聴覚という現象がある⁽²⁶⁾。これは自分の声は骨導音などの影響により空間に放射されて他人が聞く音声と異なるためである。しかし、録音再生の技術がない時代にも他人の音声を上手にまねすることができる人たちがいたのも事実である。これはいわば直接に「声道の形状」を聞き取る能力を人は持っているということであり、絶対的な音声の聞き方も起こり得ることを示唆している。

人間の話者性に対する頑健な音声知覚の研究は最近少ないが、機械音声認識の不特定話者化とも関連する重要な研究課

題である。

7. 音声知覚の実体と研究の展望

ここまで音声知覚の頑健性という観点から、音響処理レベルの特性と音素知覚における諸現象について述べてきた。後者において観測されているところは、人間の柔軟な適応処理能力を示している。これは最終的に人間にとっての意味の抽出(実験では、実験タスクの遂行)が可能となるような適応的処理が遂行されていると考えられる。そこでタスクないし条件が変化すれば、知覚結果は異なり一つの処理モデルでは説明が困難である。一方、多くの場合知覚の結果は受聴者に共通のものであるので、何らかの不変性(invariant)を音声信号に見いだす試みが行われてきた。前章で述べた構造的表象もその一つといえよう。しかし、不変量の追求は、音声知覚における一側面を示してはいるが、それだけが音声知覚のすべてではない。もし、人間と同じような音声認識能力を持つ機械音声認識システムが実現され、それを一つのブラックボックスとみなすなら、そのシステムがいかに複雑なものであっても音声認識における不変な変換が見いだされたといってもよいだろう。しかし、人間との共通理解性を保証する根拠となる意味を発生させるシステムの実現は実質的に不可能であると考えられる。すなわち認識とはその個体にとっての何らかの生理的、社会的意味すなわち個体にとっての生存上の意味を見いだすことであるので、そのような意味を持ち得ないものが人間と同じような認識を持つことはあり得ないということである(このような議論に興味のある方は、認識論の本や文献(27)、(28)が参考になると思われる)。しかし、このことはあくまで原理的な話であって、今後、機械音声認識の性能の向上が見込めないと機械音声認識は人間生活に役に立たないといっているわけではないので注意してほしい。

それでは、人間を対象とした実験や観測によって音声知覚の実体に迫れるのであろうか。これまで見てきたところでは、人間の処理は極めて柔軟であり、タスクや条件に応じた適応的処理が行われていると考えられる。したがって条件等が変化すれば、その知覚は変動し音声の知覚を調べ尽くすことは困難であると考えられる。正に「分け入っても分け入っても青い山(山頭火)」の状態に近い。しかし、人間の分散聴取能力の意外な低さや、音素特徴の抽出後には音源の情報が利用されないように見えることなどから、処理の進展に応じて情報を圧縮していることや処理のモジュール性の存在をある程度は期待できると考えられる。

人間を対象とした実験・観測手法には極めて制約が多く、従来から手法の貧困がいわれてきた。最近では、PETよりも使いやすいfMRIなどの生理的観測手法の使用が比較的容易になり、これらをEEGやMEGなど他の生理学的観測手法や心理学的実験と組み合わせた研究も多く行われている。また、研究のアプローチとしては、伝統的な健常成人に対する心理的なアプローチのみでなく、幼児の音声言語獲得、第2言語の学習過程、クロスリンギスティックな比較研究などの多面的総

合的アプローチが有効になると考えられる。

新しい研究の切り口としては、これまで余り問題にされることのなかった、知覚過程の個人性や連続音声の知覚を扱う上で音声生成と音声知覚の総合的視点が重要になると思われる。

最後に機械音声認識との関連からの課題についてみると、音声の生成には不可欠である韻律情報に対する取組みを強める必要があると思われる。人間はピッチをフラットにするなど韻律情報をかなり変形してもほぼ正確な音素知覚が可能であるので、韻律情報は不要と考えるのは妥当な戦略であろうか?利用できる情報を使うことで新たな糸口が見つかるかもしれない。連続音声の知覚においては、適切な分節化が極めて重要であり、それには韻律情報の利用が不可欠であろう。また、exemplar basedといわれる音声の話者性に関する頑健性がどのように獲得されるのかも機械音声認識との関連から重要な課題である。

HMMなどに代表される統計的な手法は、音声認識・理解システムの急速な発展をもたらしたが、一方音声科学との間で距離が生じた。これからの研究の新しい展開のためには、もう一度研究者が、相互に関心を持って交流することは意味があろう。

8. まとめ

主として妨害音に対する音声知覚の頑健性にかかわる知覚現象のうち音響処理レベルの特性、更に言語音声の特徴を生かした頑健な知覚、また統合の時間的範囲と密接に関連する処理の単位などについて述べた。実際にはここに述べた範囲だけでなく音声知覚の頑健性には高次の効果が大きく寄与している。語彙の親密度、頻度などの効果を初めとして統語的なものや発話の状況といったものが知覚の頑健性に一層大きな効果をもたらしているものと考えられる。また、ここでは詳細は言及しなかったが、韻律情報の効果も大きい。最終的には「意味」というものが適応的な処理に大きな効果を果たしていると考えられる。

文 献

- (1) 古井貞熙, “音声認識の動向 [I] - 話し言葉音声認識 -,” 信学誌, vol.89, no.8, pp.746-751, Aug. 2006.
- (2) E.C.Cherry, “Some experiments on the recognition of speech, with one and with two ears,” J. Acoust. Soc. Am., 25, pp.975-979, 1953.
- (3) A.Infeld and B.S.-Cunningham, “Spatial release from energetic and informational masking in a selective speech identification task,” J. Acoust. Soc. Am., 123, 6, pp.4369-4379, 2008.
- (4) J.Driver, “Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading,” Nature, 381, pp.66-67, 1996.
- (5) H.McGurk and J.MacDonald, “Hearing lips and seeing voices,” Nature, 264, pp.746-748, 1976.
- (6) R.M.Warren, “Perceptual restoration of missing speech sounds,” Science, 167, pp.392-393, 1970.
- (7) 柏野牧夫, “閉鎖区間の前後に分散する手掛かりに基づく日本語閉鎖子音の知覚,” 音響誌, 48, 2, pp.76-86, 1992.
- (8) A.M.Liebermann, D.Inseng, and D.Rakerd, “Duplex perception of cues for stop consonants: evidence for a phonetic mode,” Perception

- and Psychophysics, 30, pp.133-143, 1981.
- (9) J.E.Cutting, "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," *Psychological Review*, 83, 2, pp.114-140, 1976.
 - (10) 加藤和美, 笥 一彦, "音声知覚における話者への適応性の検討," *音響誌*, 44, 3, pp.180-186, 1988.
 - (11) K.Kakehi, K.Kato, and M.Kashino, "Phoneme/syllable perception and the temporal structure of speech," in *Phonological Structure and Language Processing*, T. Otake & A. Cutler eds., de Gruyter, Berlin, 1996.
 - (12) 加藤和美, 笥 一彦, "話者混合音声の知覚過程における音素情報の抽出と知覚的統合," *音響学聴覚研資*, H-92-55, pp.1-6, 1992.
 - (13) 柏野牧夫, 平原達也, "一度に何人の声を聞き分けられるか?," *音響論集*, 分冊I, pp.467-468, March. 1996.
 - (14) 川島尊之, 佐藤隆夫, "同時複数話者の分散的聴取における知覚限界," *音響誌*, 65, 1, pp.3-14, 2009.
 - (15) 笥 一彦, "言語音声の認知単位," *月刊言語*, 35, 10, pp.28-35, 2006.
 - (16) J.Mehler, J. Y.Dommergues, U.Frauenfelder, and J.Segui, "The syllable's role in speech segmentation," *J. of Verbal Learning and Verbal Behavior*, 20, pp.298-305, 1981.
 - (17) T.Otake, G.Hatano, A.Cutler, and J.Mehler, "Mora or syllable? Speech segmentation in Japanese," *J. of Memory and Language*, 32, pp.358-378, 1993.
 - (18) A.Cutler, J.Mehler, D.G.Norris, and J.Segui, "The syllables differing role in the segmentation of French and English," *J. of Memory and Language*, 25, pp.385-400, 1986.
 - (19) E.Dupoux, K.Kakehi, Y.Hirose, C.Pallier, and J.Mehler, "Epenthetic Vowels in Japanese: A perceptual Illusion?," *J. of Experimental Psychology HPP*, 25, 6, pp.1568-1578, 1999.
 - (20) M.Kashino, A. van Wieringen, and L.Pols, "Cross-Language differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners," *Proc. ICSLP 94*, pp.1079-1082, 1992.
 - (21) E.Dupoux, C.Pallier, K.Kakehi, and J.Mehler, "New evidence for prelexical phonological processing in word recognition," *Language and Cognitive Processes*, 16, 5/6, pp.491-505, 2001.
 - (22) Y.Kureta, T.Fushimi, and I.Tatsumi, "The functional unit of phonological encoding: Evidence for moraic representation in native Japanese speakers," *J. of Experimental Psychology LMC*, 32, 5, pp.1102-1119, 2006.
 - (23) K.Johnson, "Speaker Normalization in Speech Perception," in *The Handbook of Speech Perception*, D. B.Pisoni and R. E.Remez, eds., Blackwell, Malden, 2005.
 - (24) 峯松信明, "音声言語運用が要求する認知的能力と音声言語工学が構築した計算的能力," *信学技報*, NLC2008-29, SP2008-84, pp.1-6, Dec. 2008.
 - (25) 加藤和美, 笥 一彦, "音声知覚における話者への適応性の検討," *音響誌*, 44, 3, pp.180-186, 1988.
 - (26) 笥 一彦, "録音した自分の声って自分の声とは思えませんが、なぜ?," *心理学ワールド*, 34, p.35, 2006.
 - (27) 渡辺 慧, *知ること - 認識学序説*, 東京大学出版会, 東京, 1986.
 - (28) アントニオ・R・ダマシオ, *Looking for Spinoza*, 田中三彦 (訳), *感じる脳*, ダイヤモンド社, 東京, 2005.
(平成21年4月15日受付 平成21年5月15日最終受付)

笥 一彦 (正員)

昭40早大・理工・電気卒。昭42同大学院修士課程了。同年日本電信電話公社(現NTT)入社。通話品質、聴知覚の基礎研究に従事。平6名大大学院人間情報学研究科教授、平16中京大・情報科学部・教授、現在同大情報理工学部・教授。工博。音声知覚・理解・文処理・コミュニケーションの研究を行う。昭58年度本会論文賞、昭60年度科学技術庁長官賞、日本音響学会、日本認知科学会、日本音声学会等会員