

論文 / 著書情報
Article / Book Information

論題(和文)	音声知覚研究とその音声情報処理への応用
Title	
著者(和文)	古井 貞熙
Author	SADAOKI FURUI
出典(和文)	日本音響学会誌42巻12号, Vol. 42, No. 12, pp. 953-958
Journal/Book name	, Vol. 42, No. 12, pp. 953-958
発行日 / Issue date	1986,

音声知覚研究とその音声情報処理への応用*

古井 貞 熙**

(NTT 電気通信研究所)

1. はじめに—音声生成と音声知覚—

音声は人間が用いるコミュニケーション手段の中で、最も基本的かつ重要なものである。音声は聞き手に理解されることを目的とし、それを前提として発声される。従って、音声の発声方法は、人間の音声知覚能力と極めて密接に関連している。音声が進ぶ情報は、言語（意味）情報、個人性情報、アクセント情報などに大別することができ、これらは相互に関連を有しているが、一般に音声の知覚というときには、言語情報を意味することが多い。

人間が言語音声を生成するプロセスの第1段階は、相手に伝えたい内容を決め、語句を選択して文法規則（構文論規則）に合うように並べることにより、それを言語形式に変えることである。これができると、音形論規則に基づいて、脳から発声器官に運動神経指令が発せられて、発声器官の種々の筋肉が動き、空気振動としての音波が作り出される。話しかけられた相手は、聴覚によってそれを知覚し、内容を理解（認識）する。相手が、基礎となっている言語体系を知らないと、会話の内容を知覚するが理解には至らない（これは文字の場合も同様である）。例えば連続音声では、単語境界でも前後の単語の音響特徴が連続的に変化し、物理特徴としての単語境界は存在しないため、人間が音声を知覚する場合は、聞き手の能動的機能により、高次の情報に基づいて top-down 的に単語境界が仮定されていると考えられる。音声の聴取過程は、単に音響的刺激を聞くという受動的なものではなく、言語音として聴き取るという能動的なものである。いかなる音も、言語音又は非言語音のいずれかとしてはっきり区別して知覚され、この中間はない。音声には種々のあいまい性が含まれているため、相手の人の音声を理解するためには、相手の声の個人性、現在の話題は何か、等の知識を活用することが必要な場合が多い。

図-1 に示すように¹⁾、音波は相手の耳に伝わるだけでなく、話し手自身の耳にも伝わり、話し手はこの自分の音声のフィードバックを聞きながら、絶えず発声器官のコントロールを行っている。このフィードバックが重要な役割を果たしていることは、聴覚障害を生じてから数年経たずの人の音声の例や、音声の遅延フィードバック効果、すなわち自分の音声に時間遅れを付与して聴きながら発声しようとするとき極めて発声が困難になる現象からよく分かる。この音声の生成と聴き取りの密接な結びつきは、「ことばの鎖 (speech chain)」²⁾ と呼ばれている。ことばの鎖は、話し手と聞き手でその順序は逆であるが、いずれも言語学的、生理学的、及び物理学的（音響学的）段階の3段階からなっている。

2. 音 韻 知 覚

2.1 音韻知覚の手掛かり

音韻性の知覚には種々の手掛かり (cue) が複合して寄与しており、それらの手掛かりの相互作用が重要であることが知られている。同じ音素でも、前後の音素関係等によって、用いられる手掛かりやその重要さが異なる場合もある。それだけ音声の音響的性質には冗長性があるということもできる。

1) 母 音

母音は（準）定常的な音であり、第1～第2ホルマント (F_1 , F_2) (共振周波数) によって、知覚がほぼ決定される。ただし連続音声の中では、母音の音響特徴も常に変化しており、後述するように動的情報が母音の知覚においても重要な役割を果たしている。

2) 子 音

子音は非定常的な過渡音であり、短時間のスペクトルあるいは波形の固有の性質³⁾ と、前後の母音へのホルマントの遷移が重要な役割を担っている⁴⁾。実際に観測されるホルマント軌跡の外挿 (50 ms 程度) に基づく仮想的な遷移開始起点であるローカス (移行開始周波数) に着目して、 F_1 のローカスは主として調音方法に、 F_2 のローカスは主として調音位置に関連していると言われていた⁵⁾。最近ではそれよりも、ホルマント遷移の方向と速度がより重要とする考え方が強い⁶⁾。単音節を発声

* Speech perception research and its application to speech information processing.

** Sadaoki Furui (NTT Electrical Communications Laboratories, Musashino, 180)

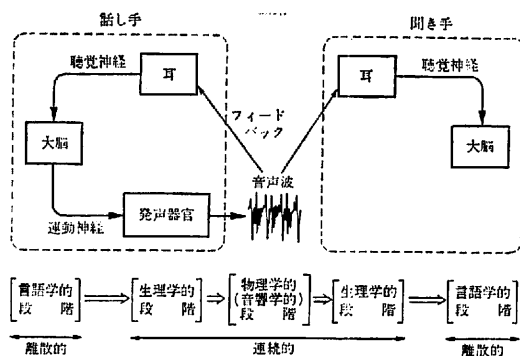


図-1 ことばの鎖

した自然音声进行分析合成系を通して、スペクトルの時間的な変化速度を速くしたり遅くしたりして行った明瞭度試験により、無声音相互間の区別、有声・無声破裂音の調音位置の識別、摩擦音と摩擦音の識別、弾音の知覚などにおいて、時間情報が重要な役割を果たしていることが確認されている⁷⁾。スペクトル遷移部の長さを長くすると /b/ が /w/ に知覚されるようになるという実験結果も報告されている⁸⁾。また、音素の手掛かりには相対的な性質があるので、一部分だけを他の部分と違う割合で時間伸縮すると、このために異聴を生ずることがある。

子音の種類別の特徴的な手掛かりを以下に示す。

a) 摩擦音：比較的に長い定常的雑音区間における一對の共振・反共振周波数と、後続母音への F_2 の遷移。

b) 破裂音：破裂時点 (noise burst, release) のスペクトル概形 (中心周波数)、前後の母音への F_2 と F_3 の遷移の相対関係、遷移の速さ、破裂時点から有声音の立ち上がり時点までの休止区間の長さ (VOT: voice onset time)、破裂の前の休止区間等。

c) 鼻音：2 kHz 程度以上の高域のエネルギーの低下と、前後の音素へのスペクトルの遷移。

2.2 動的音響特徴の役割

音声は本質的に動的性質を有し、特に連続音声においてはほとんど定常的な部分を持たない。極端に言えば、“変化のみが情報を担っている⁹⁾”とも言える。日本語 100 音節のそれぞれの音声区間の初めや終わりの一定区間を除去して、受聴試験を行った結果によれば、音節の知覚に最も重要な情報は、母音立ち上がり部付近の、スペクトル変化が極大となる位置に存在することが確認されている¹⁰⁾。また、FM 音を用いたマスキング実験によれば、周波数変化時にマスキングパターンが大きなオーバーシュートを示し、FM 音に対する人間の聴感度が高いことが確認されている¹¹⁾。母音への遷移区間のスペクトル変化極大点が重要ということは、子音は子音単独でなく、CV (C: 子音, V: 母音) 又は VC を単位として知覚される傾向が強いことを示している^{6), 12)}。連続音声中

の波形でも、CV 音節を単位として切り出せば、調音結合の影響を受けてはいるが、それだけで可成り正しく聞き取れることが確認されている¹³⁾。

音声の動的性質には、上述のように動特性それ自体に音素知覚の手掛かりが含まれているという側面のほかに、調音結合の結果として物理特性が連続的に変化するために、そこに音韻特徴が埋め込まれてしまうという側面がある。調音結合の知覚的側面は、連続音声中である音素を発声するとき、本来の調音の目標値まで達しないうちに次の音素の発声に移る (中性化) にもかかわらず、その音響的特徴の動特性から目標値が知覚される現象であるということが出来る。そこで、調音結合のメカニズムをモデル化し、調音結合の逆変換、すなわち種々に変形した音響のパターンから、話者が意図した本来の音韻情報へ変換する試み (聴覚的補償機構の解明) が幾つか行われている。これまでの試みは、ホルマントの時間パターンに 1 次系又は 2 次臨界制動系をあてはめ、モデルのパラメータを推定するものが多い^{14), 15)}。知覚的瞬间に対応する 50 ms という比較的短いスペクトルの時間パターンに基づいて、その変化の目標値を推定する聴覚の数値モデルも研究され、聴覚との良い対応が得られている¹⁶⁾。推定した声道断面積にカルマンフィルタを適用し、時間軸を逆にたどってホルマントローカスに相当する声道断面積を推定して /b, d, g/ を区別する試みも行われている¹⁷⁾。

ホルマント周波数パターンをオーバーシュートさせて知覚される音韻を推定し、更に対比効果により音素判断の境界が移動する現象をとり入れた母音の知覚モデルも検討されている¹⁸⁾。破裂子音を含む VCV 音節の調音結合 (ホルマント軌跡と声道形) を、V-V の遷移に C 固有の特徴 (invariant) が重ね合わされたモデルで説明する試みも行われている¹⁹⁾。

2.3 範ちゅう的知覚

子音に関しては、類似した音同志の相互分離の能力 (感度) は、同一音素内より音素と音素の境界のほうが著しく高いという結果がある。 F_2 ローカスを連続的に変化させて有聲破裂音を合成した場合、聴取率は /b/ → /d/ → /g/ とほとんど all-or-non の形で変化し、音素差以上の音色の差はほとんど感じられない。このことから、子音に関しては範ちゅう的知覚がなされていると考えられる。

一方、母音に関しては、実験の方法にもよるが、一般に音素境界において相互分離能力が高まるということではなく、同一母音の音響的な要因の変化に応じて音色の差異を認めることができる。これから、母音に関しては連続的知覚がなされていると考えられる。母音の知覚では、聞き手はすでに聞き終えたものとの対比で判断する、すなわち、聞き手がその状況に応じて適当な判断基

準系を形作って知覚するという、相対的、対比的現象（対比効果）が顕著に見られる²⁰⁾。

ただし、母音でも刺激の長さを非常に短くした場合には範ちゅう的知覚となることから、言語音に関しては基本的にまず音素に関する範ちゅう的知覚が行われ、連続的知覚が併存できるか否かは与えられる情報量によるという考え方もある^{21), 22)}。すなわち記憶のメカニズムと関連づけて、範ちゅう的でない記憶は感覚記憶に蓄えられるか容量が小さいので短時間に失われやすく、連続的知覚には多量の情報を必要とする。ところが子音や連続音声の中の母音は情報が少なく失われやすいので、主として短期記憶に蓄えられている言語符号が判断に用いられて、範ちゅう的に知覚されるという考えである。

3. 音韻知覚メカニズムのモデル化

3.1 聴覚末梢系の機能モデル

これまでに得られている種々の知見を考慮すると、聴覚系の機能モデルが基本的に備えているべき条件として次のようなものがあり²³⁾、このような機能モデルに基づく音声処理の試みが幾つか行われている^{24)~26)}。

- 1) 2~3 kHz 付近の周波数領域で最大となるエンフエンス特性（外耳と中耳の特性）。
- 2) Mel 又は Bark スケール上で十分オーバーラップして等間隔に並んだ臨界帯域フィルタ群。位相情報は除去する（基底膜の特性）。ここで、1 Bark は臨界帯域幅に対応し、周波数 [Hz] から Bark への変換については、幾つかの公式が提案されている²⁷⁾。
- 3) 半波整流（有毛細胞の特性）。
- 4) 時定数の短い低域通過フィルタ（10 ms 程度の時間分解能と 50 dB 以上のダイナミックレンジ）による時間平滑化。
- 5) スペクトルのピークを強調する側抑制回路。
- 6) フィルタ出力の立ち上がりとしち下がり（立ち下がり）を強調する微分又はこれに類似した計算。
- 7) ラウドネスの phon スケールを近似するフィルタ出力の対数変換。
- 8) スペクトルのピークの位置を重視し、そのレベルの絶対値やスペクトルの全体的傾斜の変動を吸収する能力。

3.2 音韻知覚モデル

2.3 節で述べた母音と子音の一般的な知覚形態の相違は、調音と対応づけて考えることができる。すなわち、母音の調音は連続的に考えることができるが、子音の調音は、/b/ と /g/ の中間の子音を調音することはできないように、範ちゅう的すなわち不連続にしか変化できないことと対応づけられる。このことから音響刺激の音声としての聴き取りの様式と、その音素の調音の様式の間には密接な関連があるのではないかという考えが生ま

れ、「音声聴取の調音参照説」²⁸⁾ が出された。これは、聞き手は、自分がその音を音声として発声するとしたらどのような調音運動を行うであろうか、という自分の運動感覚とでも言うべきものを参照して、音声の聴取を行っているという説である。

これを更に進めた説として、「音声聴取のモータ理論」あるいは「運動指令説」²⁹⁾ がある。これは、音声の生成過程を更にさかのぼって、音素とのより本質的で簡明な 1 対 1 対応を示すものは調音器官への神経制御指令（運動神経指令）であると考え、音声の聴取がこの段階で行われるとする説である。この説に対しては、白らの発声器官で生じることのできる音声は非常に限られているにもかかわらず、生まれて 1 か月の子供でも（生まれつき）大人の音素を弁別可能であるし、外国語が発声できなくても区別できるように、発声することができるものだけを聞くことができるわけではないという批判もある。

音声生成過程を更にさかのぼって、聞き手は抽象的で離散的な表現から音声が発声するときの規則に従ってパターン（音響信号）を合成し、この内部的に作られた合成パターンを話し手の音声（聴覚パターン）と比較照合してパターンマッチングを行うとする、「合成による分析 (analysis-by-synthesis, A-b-S) モデル」も提案されている³⁰⁾。この説に対しても生成を根源的なものと考えすぎるのではないか、もっと簡単な音響特性比較が中心ではないかという批判がある。

4. 音声知覚メカニズムのモデル化

4.1 音声知覚過程のモデル化

言語処理過程を含む音声知覚過程は、図-2 のようにモデル化することができる。音声波に対して、外耳から基底膜に至る聴覚器官によってまず周波数分析がなされ、基底膜上の有毛細胞から発する聴覚神経系においてスペクトル包絡、強度、基本周波数等の時間パターンが抽出される。これらのパターンは神経系の階層を高次へと進むにつれて非線形変換、動的特徴の抽出などがなされる。ここまでの過程は受動的かつ連続的な過程であり、音声でも音声以外の刺激でも変わらない。その後これらの音響的特徴は大脳中枢における複雑な処理である言語理解過程の入力情報として用いられる。

言語理解過程は、言うまでもなく人間だけが持つ処理過程であり、音声に対する言語理解過程は書かれた文字の場合とは異なった特徴がある。その一つは、音響刺激が文字と違って逐次的にしか入力されないため、基本的に left-to-right の処理であること、2 番目は入力情報が調音結合、個人差、雑音などによって常に、言わば草書体のような大きな変形を受けていること、3 番目は書き言葉と違って多くの言語的あいまい性を含んでいることである。音声言語理解過程では、図-2 に示すような種

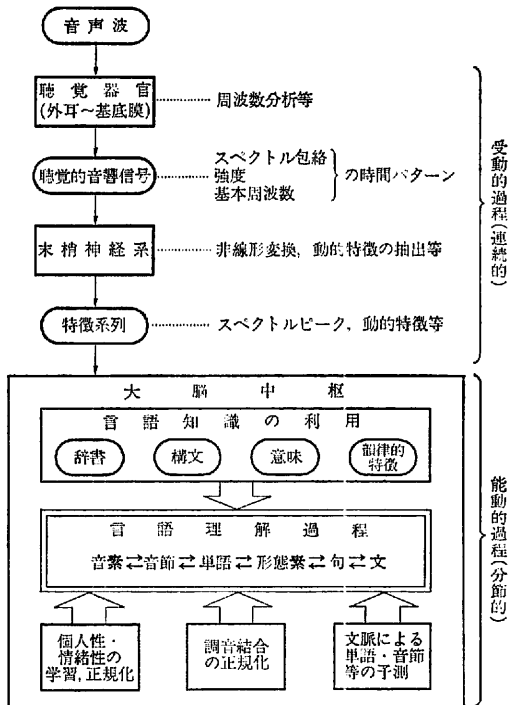


図-2 音声知覚過程のモデル化

種の処理が、再符号化（分節化）された様々な情報を用いて並列的に実行され、音響特徴、構文、意味等にあいまい性があると、そのあいまい性を補うように他の情報が組み合わせて用いられる。音声の知覚においては、音響特徴の動特性に基づく予測のほか、音節、句、文のレベルでも top-down 的な予測が重要な役割を果たしていると考えられる。聞こえと意図したもの（無意識であっても）だけが聞こえるということもでき、この意味で言語理解過程は能動的過程とすることができる。

4.2 単語知覚モデル

音声知覚における時間的まとまりとしての処理単位は、文章音声にクリック音又は hiss 雑音を重畳してその位置を同定させる実験によって確認されている^{31), 32)}。実験によれば、固定した認知的処理単位というものではなく、主要な文法的切れ目を中心に音節から句まで種々に変化するが、単語が最も重要な単位である^{33), 34)}と考えられる。R. A. Cole と J. Jakimik は、これまでに得られた種々の実験の実験結果をもとに、次のような単語知覚モデルを提案している³³⁾。

- 1) 単語は音と知識の相互作用によって認識される。
- 2) 音声は1語1語連続的に処理される。各単語は、
 - a) すぐ次に続く単語の始まりの時点を示す、と共に、
 - b) すぐ次に続く単語の認識に用いられる構文的及び意味的制約を与える。
- 3) 辞書（記憶）中の単語はその始まりの音によってアクセスされる。

4) 単語が認識されるとは、その音響的構造の分析によって候補単語が1単語に絞られることである。

音響特徴と種々の知識の bottom-up 及び top-down 処理の組み合わせにより単語が一つに決定される過程の心理言語学的モデルとしては、logogen モデル³⁵⁾、cohort モデル³⁶⁾、phonetic refinement モデル³⁷⁾、analysis-by-synthesis モデル³⁸⁾、LAFS (lexical access from spectra) モデル³⁹⁾ などが提案されている。これらの多くは、音声だけでなく書かれた文字の知覚などにも適用できるモデルである。LAFS モデルは計算機による音声認識の影響を受けた比較的簡単なモデルで、単語は音素でなく二つの音素の結合“diphone”スペクトルのネットワークとして記憶されていると考え、音声のスペクトル系列とこのネットワークとの比較で単語認識が行われるとしている。

4.3 個人性情報の知覚

音声に含まれる情報の中で、言語（意味）情報について重要な個人性情報は、声の質の違い、高さ、大きさ、発声速度、テンポ、イントネーション、アクセント、語彙などに現れる。これらの特徴に関連する音声の物理量には種々のものがあり、これらが複雑に寄与していると考えられるが、聴覚的に最も重要な特徴である声の質と高さは、主としてスペクトル包絡情報と基本周波数（ピッチ）の平均値及び動特性として捕えることができる^{39)~41)}。

人間が種々の人の声を聴取する場合、人によっては声質が大きく異なって、最初は何を言われたかはっきり分からない場合があるが、すぐにその声に慣れて（適応して）よく分かるようになる能力がある。種々の条件による多数話者の音節呈示による明瞭度試験の結果、新しい話者に対する適応は4~5音節程度の受聴で達成され、適応に用いられる音節の種類には依存しないことが確認されている⁴²⁾。

大勢の人が話している中で、ある1人の声に注目して聞き取ることができる、いわゆる「カクテルパーティ効果」には、両耳による音像の定位が重要な役割を果たしているが、人による声質の違いを聞き分ける能力も大きな役割を果たしていると考えられる。

5. 音声情報処理への応用

音声情報処理の研究と実用化は、一見近年大きく進展しているように見えるが、例えば不特定話者が発声した大語彙の連続音声を認識しようとするときに問題になる、個人差、調音結合、言語処理などの本質的問題は未解決のままと言っても良い。鳥と飛行機のたとえを持ち出すまでもなく、工学的な音声認識のアルゴリズムに人間と同じメカニズムを採用しなければならない道理はないが、これまで力づくや思いつきと批判されがちであっ

た音声情報処理研究の新たな進展へのヒントあるいは指導原理を音声知覚研究に求めるのは、一つの有望なアプローチと考えられる。

音声認識における音声知覚の知見の応用としては、音声知覚モデルによる音声認識の試み²⁴⁾、人間がスペクトルの動特性に敏感な性質に基づいて、ケプストラムの回帰係数として動的特徴を抽出して用いる方法⁴³⁾ などがある。中性化しているホルマントやスペクトルの時間パターンに、変化を強調するように変形を加える試みも行われている^{44), 45)}。その他に、聴覚のスペクトルの傾斜に対する感度が低く、ピークの位置のずれに敏感な性質に着目した方法として、スペクトルの周波数軸方向の局所的な傾斜を認識に用いる方法⁴⁶⁾、これらと等価な重みつきケプストラムによる方法⁴⁷⁾、ピークの位置ずれに敏感な距離尺度 (WLR)⁴⁸⁾ 等が提案され、それぞれ有効性が示されている。線形予測分析法もこのような意味で聴覚にマッチした分析法であると言える。

音符号化法では、聴覚のマスクング特性を利用して、量子化歪のスペクトル特性を変形し、聴覚的に検知されにくくするノイズシェイピングの技術⁴⁹⁾ が広く用いられている。また、聴覚が位相特性の変化に鈍感であることを利用して、位相特性を変え、符号化効率を上げる位相等化処理⁵⁰⁾ も有効性が確認されている。

6. む す び

聴覚の心理及び生理の研究は、H. von Helmholtz をはじめとして 19 世紀からの長い歴史を有し、音声知覚の研究も、1920 年代の H. Fletcher の研究をはじめとする歴史があるが、音声知覚のメカニズムに関しては、まだほとんど明らかになっていないと言ってよい。近年の音声情報処理技術の中にも、すでに人間の聴覚の機能に関する知見が幾つか取り入れられているが、今後の飛躍的發展のためには、音声生成及び音声知覚の研究がますます重要になってくると考えられる^{1), 51)}。その際、心理学、生理学、及び工学的研究の協力関係がいっそう必要になってくるであろう。ただし、性急に音声情報処理への応用を考えるのではなく、むしろ「何のために」などと考えずに純粋に科学としてじっくり研究する姿勢が重要である。近年「科学」として音声を研究する姿勢が乏しくなっているが、特に大学や公立研究機関ではこのような姿勢がもっとほしいように思う。

これまで合成音の聴取による音声知覚の研究が種々の重要な知見を明らかにしてきたが、合成音による研究の結果には、単純化しすぎたモデルに起因する誇張、不自然に独立に、あるいは広範囲に変化された要因の影響、用いられた合成方式の持つ欠点による不完全さなどが現れることがしばしばあり、結論を導くに当たって十分注意する必要がある。合成音による結果が、聴覚のあ

る面を表現しているのは間違いないが、それが音声知覚の一般的・本質的現象であるかどうかは分からない。過去に研究された弁別素性も、生成機構からみた理想的姿であり、人間の音声に対する多面的な感覚の一側面に過ぎないと言えよう。合成音でなくとも、例えば連続音声の中から音節や単語を切り出すと、切断による歪を生じているため、そのまま受聴試験を行ったのでは、本来それらの音声区間が有している量よりも低く情報量が見積もられてしまう¹³⁾。目的にあった適切な実験方法を開発することも重要な研究課題の一つである。

文 献

- 1) 古井貞照, デジタル音声処理 (東海大学出版会, 東京, 1985).
- 2) P. B. Denes and E. N. Pinson, *The Speech Chain* (Bell Telephone Laboratories, Inc., Murray Hill, 1963).
- 3) K. N. Stevens and S. E. Blumstein, "The search for invariant acoustic correlates of phonetic features," in *Perspectives on the Study of Speech*, P. D. Eimas and J. L. Miller, Eds. (Lawrence Erlbaum Associates, Hillsdale, 1981).
- 4) R. A. Cole and B. Scott, "Toward a theory of speech perception," *Psychol. Rev.* **81**, 348-374 (1974).
- 5) D. W. Massaro and G. C. Oden, "Speech perception: A framework for research and theory," in *Speech & Language*, Vol. 3 (Academic Press, New York, 1980), pp. 129-165.
- 6) J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, 2nd ed. (Springer-Verlag, New York, 1972).
- 7) 片桐 滋, 東倉洋一, 古井貞照, "単音節知覚における時間情報の役割," 音響学会誌 **42**, 97-105 (1986).
- 8) J. L. Miller and T. Baer, "Some effects of speaking rate on the production of /b/ and /w/," *J. Acoust. Soc. Am.* **73**, 1751-1755 (1983).
- 9) L. C. W. Pols, "How to make more efficient use of the fact that the speech signal is dynamic and redundant," *Proc. IEEE ICASSP* **86**, 37.3 (1986).
- 10) 古井貞照, "音声知覚におけるスペクトル変化情報の役割," 音響学会聴覚研資 H-85-6 (1985).
- 11) 伊福部達, "FM 音によるマスクング," 音響学会誌 **29**, 679-687 (1973).
- 12) F. S. Cooper, P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman, "Some experiments on the perception of synthetic speech sounds," *J. Acoust. Soc. Am.* **24**, 597-606 (1952).
- 13) 東倉洋一, "言語音中の音韻知覚," 音響学会聴覚研資 H-83-65 (1983).
- 14) B. Lindblom, "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.* **35**, 1773-1781 (1963).
- 15) 藤崎博也, 吉田 賢, 佐藤泰雄, 田辺吉久, "ホルマント周波数領域における調音結合のモデルとその母音連続音声の認識への応用," 音響学会音声研資 S73-03 (1973).
- 16) 赤木正人, 古井貞照, "音声知覚における母音ターゲット予測機構のモデル化の検討," 音響学会聴覚研資 H-85-35 (1985).
- 17) 石崎 俊, "カルマンスムーザを用いた調音点の推定," 音講論集 2-2-7 (1976.5).
- 18) 金森吉成, 城戸健一, "母音知覚における音素環境の影響," 音響学会誌 **32**, 3-11 (1976).
- 19) S. E. G. Ohman, "Numerical model of coarticulation," *J. Acoust. Soc. Am.* **41**, 310-320 (1966).

- 20) D. B. Fry, A. S. Abramson, P. D. Eimas and A. M. Liberman, *Lang. Speech* 5, 171-189 (1962).
- 21) 藤崎博也, 川島崇子, “合成音声の弁別と言語音知覚機構のモデル,” 音響学会誌 27, 453-462 (1971).
- 22) B. H. Repp, A. F. Healy and R. G. Crowder, “Categories and context in the perception of isolated steady-state vowels,” *J. Exp. Psychol.: Human Perception & Performance* 5, 129-145 (1979).
- 23) D. H. Klatt, “Speech processing strategies based on auditory models,” in *The Representation of Speech in the Peripheral Auditory System*, R. Carlson and B. Granstrom, Eds. (Elsevier Biomedical Press, New York, 1982).
- 24) E. Zwicker and E. Terhardt, “Automatic speech recognition using psychoacoustic models,” *J. Acoust. Soc. Am.* 65, 487-498 (1979).
- 25) S. Seneff, “Pitch and spectral estimation of speech based on auditory synchrony model,” *Proc. IEEE ICASSP* 84, 36.2 (1984).
- 26) B. Delgutte and N. Y. S. Kiang, “Speech coding in the auditory nerve: I, II, III, IV, V,” *J. Acoust. Soc. Am.* 75, 866-918 (1984).
- 27) S. Seneff, “A computational model for the peripheral auditory system: Application to speech recognition research,” *Proc. IEEE ICASSP* 86, 37.8 (1986).
- 28) A. M. Liberman, “Some results of research on speech perception,” *J. Acoust. Soc. Am.* 29, 117-123 (1957).
- 29) K. N. Stevens, “Toward a model for speech recognition,” *J. Acoust. Soc. Am.* 32, 47-55 (1960).
- 30) M. Halle and K. N. Stevens, “Analysis by synthesis,” *Proc. of Seminar on Speech Compression and Processing*, AFCRC, Vol. 2 (1959).
- 31) P. Ladefoged and D. E. Broadbent, “Perception of sequence in auditory events,” *Q. J. Exp. Psychol.* 13, 162-170 (1960).
- 32) J. A. Fodor and T. G. Bever, “The psychological reality of linguistic segments,” *J. Verbal Learn. Verbal Behav.* 4, 414-420 (1965).
- 33) R. A. Cole and J. Jakimik, “A model of speech perception,” in *Perception and Production of Fluent Speech*, R. A. Cole, Ed. (LEA Publishers, Hillsdale, 1980).
- 34) J. Morton and J. Long, “Effect of word transition probability on phoneme identification,” *J. Verbal Learn. Verbal Behav.* 15, 43-51 (1976).
- 35) J. Morton, “Interaction of information in word recognition,” *Psychol. Rev.* 76, 165-178 (1969).
- 36) W. D. Marslen-Wilson and A. Welch, “Processing interactions and lexical access during word recognition in continuous speech,” *Cognitive Psychol.* 10, 29-63 (1978).
- 37) D. B. Pisoni, H. C. Nusbaum, P. A. Luce and L. M. Slowiaczek, “Speech perception, word recognition and the structure of the lexicon,” *Speech Commun.* 4, 75-95 (1985).
- 38) D. H. Klatt, “Speech perception: A model of acoustic-phonetic analysis and lexical access,” in *Perception and Production of Fluent Speech*, R. A. Cole, Ed. (LEA Publishers, Hillsdale, 1980).
- 39) 伊藤憲三, 齋藤収三, “音声の音響的特徴パラメータが個人性の知覚に及ぼす影響,” 信学論 J65-A, 101-108 (1982).
- 40) 桑原尚夫, 大串健吾, “声質の類似性とその物理的対応,” 音講論集 1-2-9 (1981.10).
- 41) 古井貞照, 赤木正人, “音声における個人性の知覚と物理関連量,” 音響学会聴覚研資 H-85-18 (1985).
- 42) 加藤和美, 古井貞照, “話者の個人性に対する聴覚の適応性,” 音響学会聴覚研資 H-85-5 (1985).
- 43) S. Furui, “Speaker-independent isolated word recognition using dynamic features of speech spectrum,” *IEEE Trans. ASSP* 34, 1, 52-59 (1986).
- 44) H. Kuwabara, “An approach to normalization of coarticulation effects for vowels in connected speech,” *J. Acoust. Soc. Am.* 77, 686-694 (1985).
- 45) 古井貞照, “スペクトル変化強調による単語音声認識,” 音響学会音声研資 S85-77 (1986).
- 46) N. Nocerino, F. K. Soong, L. R. Rabiner and D. H. Klatt, “Comparative study of several distortion measures for speech recognition,” *Proc. IEEE ICASSP* 85, 1.7 (1985).
- 47) Y. Tohkura, “A weighted cepstral distance measure for speech recognition,” *Proc. IEEE ICASSP* 86, 14.17 (1986).
- 48) 杉山雅英, 鹿野清宏, “ピークに重みをおいた LPC スペクトルマッチング尺度,” 信学論 J64-A, 409-416 (1981).
- 49) R. E. Crochiere and J. L. Flanagan, “Current perspectives in digital speech,” *IEEE Commun. Mag.*, 32-40 (1983.1).
- 50) 菅田雅彰, 守谷健弘, “位相等化処理を用いた音声符号化,” 音響学会音声研資 S84-05 (1984).
- 51) 中田和男, “音声認識の基礎,” 第8回応用情報学研究センター・シンポジウム予稿集「音素を認識の基本単位とする音声自動認識」, 東北大学, 1-8 (1983).