4　心理言語学、音響音声学からの知見

音響音声学: 音声信号の音響的性質の研究

- 人間の音声の処理は極めて速く、一秒に 25 から 30 個の音素を認識できる。これは、たとえば楽器の音の認識と比べるとはるかに速い。
- **音響スペクトログラム (sound spectrogram)**: 音声の音響的エネルギーを周波数と時間軸のパラメタとして表示したもの (図 4-2 参照)—横軸が時間軸、縦軸が周波数、エネルギの強さは色の濃さで表されている。音響スペクトログラムを得るには、**音響スペクトログラフ (sound spectrograph)** という機器を用いる必要があったが、今では、Praat [7] というソフトウェアなどで簡単にできるようになっている。

  以下の図は、「あめんぼ あかいな あいうえお」の発話を Praat で表示させたもので、上が音の強さ (intensity)、下がスペクトログラム。。

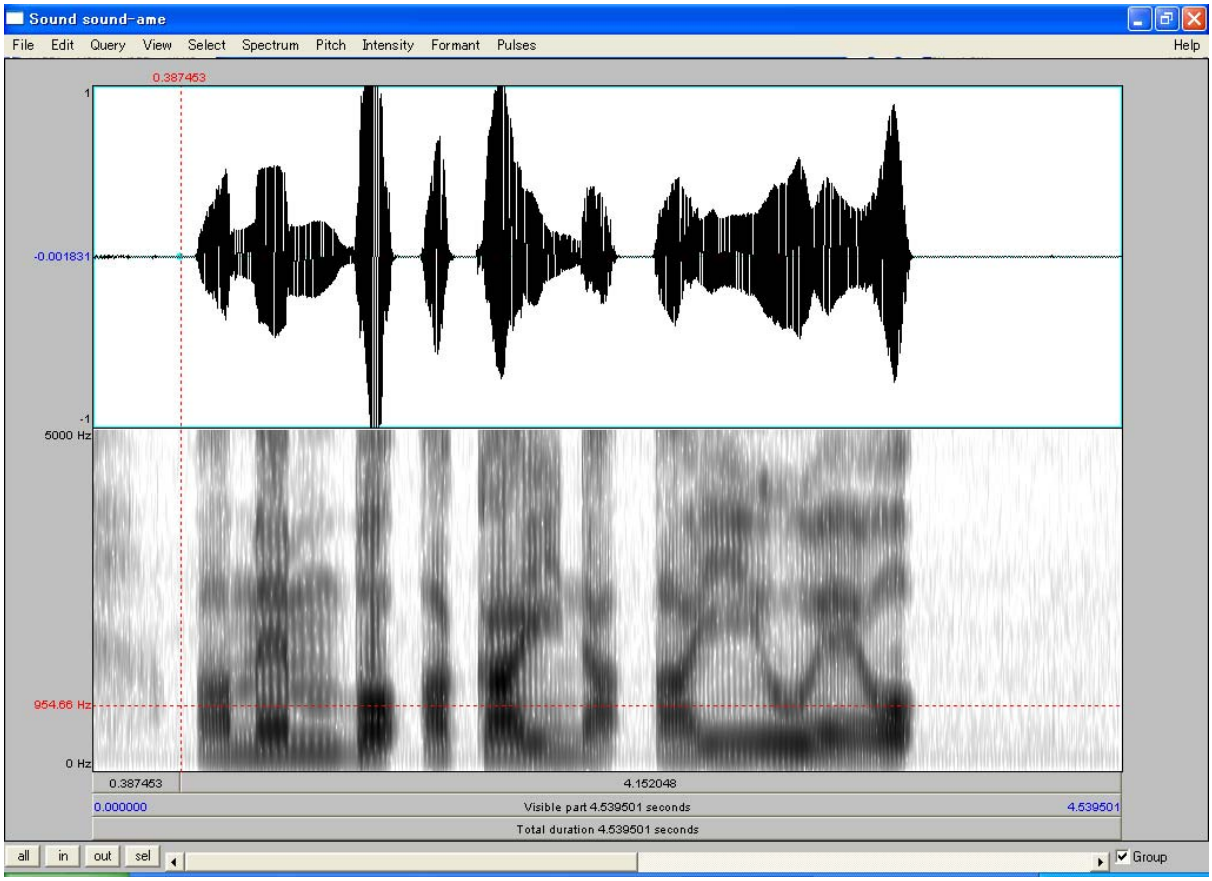

図 4-2. 典型的な音響スペクトログラム (Praat を使って発声を分析)

[7] Praat の公式サイトは `www.praat.org`

- **フォルマント (formant)**: スペクトログラム上に水平方向に色の濃い部分が見られる部分 (つまり、エネルギーが集中している)。下から第一フォルマント (F1)、 第二フォルマント (F2)、. . . という。

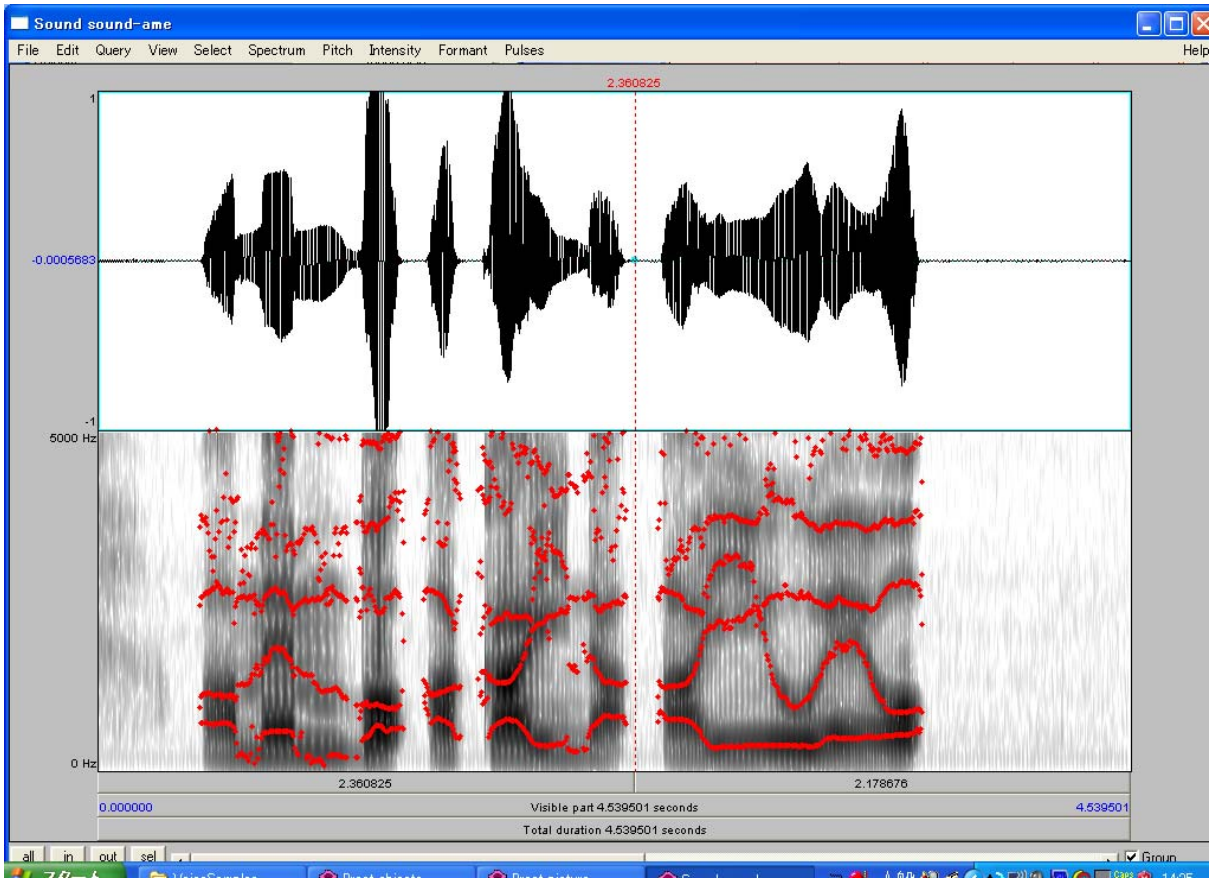

図 4-3. 図 4-2 のスペクトログラムでフォルマントを強調したもの

- 音声知覚において重要なフォルマントの二要素: フォルマント周波数が比較的安定した**定常状態 (steady state)** と、短時間にフォルマントが変化する**フォルマント遷移 (formant transition)**。

  フォルマント遷移は子音や異なる母音の連続部分、定常状態は同じ母音にほぼ対応する。

- **文脈に依存した多様性 (文脈多義性, Context-conditioned variation)**: 同じ音素でも音声的な文脈によってスペクトログラム上では異なってみえる、という現象のこと。例えば、子音がそれに続く母音によって変化する、というのはこの明らかな例。

- 文脈に依存した多様性は**調音の様態 (manner of articulation)** と密接に関係。一時に複数個の音声音を生成するという現象は、**同時調音 (coarticulation)** と呼ばれ、生成がもつ音韻的文脈で変化する傾向を表すものである。

- 韻律的な要因がさらに音声信号の多様性に加わる— 強勢 (stress)、イントネーション、発話の速さのようなプロソディ (prosody, 韻律) の要因も、音声信号の性質に貢献する。

  Miller (1981): 話す速度を上げると、母音の持続時間が減少し、いろいろな子音のための手がかりも変化する

- 音韻的手がかりにおけるこのような多様性にも拘わらず、人間は容易に音声音の同定が可能であるが、コンピュータはそうではない。最良のプログラムも二流の速記者に適わない (Pinker, 1994)。

- 音声知覚は不変的な手がかりと文脈依存的な手がかりの両方に基づいている。

  その例— Cole & Scott (1974): 鼻音 (nasal) の子音の [m] と [n] は高周波数帯のエネルギーを欠き低周波数帯のエネルギー一本からなるという特徴をもち、いろいろな母音が現れる文脈でも特徴的であるようにみえるが、[m] と [n] の区別には母音情報 (つまりフォルマント遷移) が必要。

- **範疇的知覚 (categorical perception)**

  刺激の弁別が，物理的な差異は等しくても，同一の範疇に属する場合よりも異なった範疇に属する場合の方がはるかによくなるような知覚様式のこと

  物体やできごとを普通に知覚する場合、色同士や匂い同士の間の細かな識別が可能。この識別能力は、連続体をなす刺激の量的変化を知覚するもの。音声知覚における聞き手のタスクはこれとは別物—入力信号の物理的特性を相対的に識別するのではなく、入力信号に対し絶対的、範疇的識別をしなければならない。つまり、入力信号の強さや周波数が高いか低いかではなく、それが [p] なのか [b] なのか—入力信号がどの範疇に属すかを判定。

  [ba] と [pa] は、唇から音が発せられた時間と声帯が振動した時間との時間差 (VOT) により識別されている — VOT が 0s ならば常に [ba] と聞こえ、40ms ならば [pa] と聞こえる。音声合成器により、その中間の刺激に対する知覚を調べることが可能。VOT を連続的に変化させると、40ms のところで明瞭に音の認識が変化する (一番目の実験)。

  被験者に三種類の刺激を与え、3 番目の音が先行する二つの音のうちどれと同じかを答えさせる (二番目の実験)。先行する二つの音が異なる範疇の場合、成績は非常に良いが、同じ範疇だと偶然レベルに落ちる。したがって、範疇的知覚を二つの基準が決定する: 鋭敏な同定機能が存在すること、また、同じ範疇内の音は識別が不能であること。

4.1 連続音声の知覚

現実には連続的な文脈の中に音声信号は埋めこまれている。音声の同定に、隣接するシラブルや節と言うような幅広い文脈が重要な役割を果たしている。

Pollack & Pickett (1964): 防音室で心理実験に参加するため待っている女性の会話を録音し、単語をつなぎ合わせて被験者に提示。単語は理解できるものの、個々に提示された場合は半分だけが正しく同定された。したがって、音響情報だけでは音声信号の同定には十分ではなく、文脈が必要であることが示された。

ここでは、文脈を構成するものとして次の二つの要因が重要: 音声知覚における韻律、および高次の意味および統語的要因。

- **文脈と音声認識**
  Pollack & Pickett (1964): 文脈から切り離された音声は認識が困難。ということは、文脈の意味的、統語的要因を換えれば、音声の知覚可能性に影響するはず。

- Miller らが、音声認識において高次の文脈要因が関与していることを示す。

  Miller, Heise, & Lichten (1951): 白色雑音 (white noise) の中で、単語を孤立させた提示と、5 単語からなる文中での提示を比較し、すべての白色雑音のレベルにおいて、文中条件がよりよい成績が得られた。これは、被験者が高次の情報を用いて可能性を制限できたからのようにみえる。

  Miller & Isard (1963): 統語情報と意味情報の影響を分離した実験。連続音声に対し以下の三つのタイプの文が提示された:(1) 文法的、(2) 文法的な語順であるが変則的、(3) 非文法的。

  (1) Accidents kill motorists on the highways. (事故により自動車を使う人がハイウェーで亡くなっている)

  (2) Accidents carry honey between the house. (house ではなく houses が正しい—偶然により家の間を蜜が運ばれた)

  (3) Around accidents country honey the shoot. (事故のまわり、田舎の蜜、射撃)

  その結果は、文法的な文が最も正確、ついで変則的であるが文法的語順、最も成績が悪いのが非文の場合。この実験からは、文の予測がしやすければ、認識の精度もよくなる、といえそう。

- **音素修復 (phonemic restoration)**
  (4) 中の単語 *legislatures* の最初の/s/ を取り去り、咳払い (cough) で置き換える:

  (4) The state governments met with their respective legislatures convening in the capital city. (州政府はその州都で開かれたそれぞれの州議会と会った)

  被験者は削除された/s/を聞いたと報告し、どの音が削除されたかと聞かれても、ほとんどの被験者が答えられない。音声信号を置き換えずに雑音が追加された他の手続きでも同様の修復が報告されている (Samuel, 1981)。

それに対して、咳払いのような「広帯域の雑音」以外で音を置き換えた場合、音声が中断したことは明らかに検知され、その部分に相当する音素を正しく答えることが難しくなる。

- 音素修復に対する文脈の要因が、その後の研究で示されている。Warren & Warren (1970): 次の四つの文を提示し、音素修復が構造句の文脈と関連していることを示す。

  (5) It was found that the *eel was on the axle (車軸). (wheel、車輪)

  (6) It was found that the *eel was on the shoe (靴). (heel、かかと)

  (7) It was found that the *eel was on the orange (オレンジ). (peel、果物の皮)

  (8) It was found that the *eel was on the table (テーブル). (meal、食事)

- 音素修復は、我々がいろいろなできごとが起こっている中で音声を聞き取っているという事実と密接な関係。他の音の存在や不明瞭な発音などの多くの要因により、多くの音の断片は個別には聞き取り不能。しかし、高次の文脈要因を能動的に用いて、一般に認識が可能。上昇型の情報が不在にもかかわらず知覚が可能であることを示しているので、音素修復は下降型処理が行われていることを明確に示す例となっている。

- **発音誤りの検出 (Mispronunciation detection)**
  以下のように、普通の文にちょっとした発音誤りを含む場合

  (9) It has been zuggested that students be required to preregister. (zuggested → suggested) (学生に事前登録を求めるべきであると、提案されていた)

  (を言おうとしているかが容易に推測される時は特に) 発音の小さな誤りは無視されがち。しかし時に検出される場合もある。

  Cole(1973) の発見: 発音の誤りを発見する可能性は、その誤りの位置 (単語内もしくは文内) による—単語の先頭の方が最後のものよりも、また文内の最初の方が後ろのものよりも発見されやすい。

- Marslen-Wilson & Welsh (1978): 発音誤りの検出と**追唱 (shadowing)**[8] タスクとを組みあわせて実験して、発音誤りを復唱する条件を調査。その結果、流暢性が修復と関係していること、またさらに文脈が予測しやすい場合に修復が起こりやすいが、そうでなければ復唱が起きやすいことをみいだす。

[8] 文章などの音声を被験者に呈示し，これを追いかけるように同じ内容をできるだけ早く発声させる実験課題．単語などの開始時点に着目し，呈示から発声までの時間遅れおよび発声内容を分析し，音声認知の諸性質の解明に用いる．追唱の性質は，被験者によって大きく異なるが，熟練者では約 200ms の時間遅れで発声することが可能となる．また，呈示音声に誤りがある場合にも，これを訂正して発声する傾向がある．追唱実験で得られたこれらの事実から，単語や文の音声認知過程では，統語情報や意味情報を利用したトップダウン処理が存在すると考えられている．

- 何をいおうとしているかを「知っている」場合は実際に発話される言葉をほとんど聞かず、予想とのおおまかな整合検査だけを必要としているようである。それに対し、不確実さが高ければ、修復を行うちゃんとした根拠がない。修復が流暢性と関係していることから、言語処理において意味的、統語的制約を入力信号と自然に統合していることが示唆される。これらは「推測」ではなく、あたかも発話されたかのように聞こえる。したがって、入力信号に対して即時的に認識されたものは、入力信号の分析と意味、統語制約の適用の組合せのように思われる。