音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置
——音声言語認知、音楽認知、両耳分離聴取、クロス・モダリティの研究のために——

内田 照久

I. 問題

今日、コンピュータの発展に伴い、心理学においても認知心理学などの研究領域に中心として、データ処理ばかりでなく、実験刺激の提示や反応時間の測定など、実験操作そのもののためにパーソナル・コンピュータに頼る傾向に用いられるようになってきた。このように心理学の実験において、パーソナル・コンピュータの利用がこれほど進んだ理由としては、パーソナル・コンピュータは大型計算機等と違い、実験場面でCPUを占有し、試行ごとのリアルタイムの処理が可能である点、また、一定の心理学が取り扱うような文字や図形などの実験刺激の表示や制御が可能であることなどに加え、比較的容易に実験に必要な反応スイッチを設けるため、外部機器を連動して制御するような簡便性の高いことが上げられる。また、コンピュータ全般の発展もあって、従来の性能を比較して、格段のコスト・パフォーマンスを得られるようになった点も価値ある。

これまで心理学実験のためのパーソナル・コンピュータの利用については、一般的な心理学実験の実施を意図して、中谷（1985）や阿部（1988）などが活用性の高い利用方法などを示している。

このように、心理学実験においてパーソナル・コンピュータの利用が格段に進んできたが、現在、我が国で主に使われているパーソナル・コンピュータには、音声の入出力が必ずしもサポートされていないこともあり、文字や図形を中心とした視覚的な実験と比較すると、音声刺激を用いた聴覚的な実験には、まだ十分に活用されていない側面がある。

聴覚的実験の例を挙げると、脳半球優位差の研究などで用いられる両耳分離聴取（dichotic listening）などでは、異なる音声を極めて細かいたイニングで制御して、2つのチャンネルから同時に提示することが必要となる。その目的のために従来のテープ・デッキなどのアナログ機器を利用するような場合では、同期して提示を行うために専用のハードウェアを作成する必要があった。これには、専門の知識が必要な上、作成した機器は専用機となっててしまうため、汎用性に乏しいという欠点がある。

また、視覚と聴覚の相互作用を扱うようなクロス・モダリティの研究では、視覚刺激と聴覚刺激の同時提示などの時間制御の方法を工夫することが求められる。さらに音声言語認知や音楽認知に研究においては、音声の音響特性そのものを操作して、人間の認知過程を探究することが求められる。そのためには、音声情報処理技術の利用が不可欠であり、そのような技術を利用できる環境を整えることが必須である。

このように考えると、音声を扱う心理学的な実験においてもパーソナル・コンピュータを活用することにより、実験環境を改善できる点が多い。このような環境に対して、これまでにも河合・吉崎・伊藤（1989）などは、心理学実験において音声を取り扱うための装置を開発している。

ここでは、さらに汎用的に音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置について述べるとともに、音声を扱う上で特に配慮すべき点についても述べていく。まず1章では、名古屋大学の教育心理学教室に設置されている「聴覚及び認知実験用多目的装置」とその周辺機器の概要を紹介しながら、音声を取り扱うための留意点を述べる。次に2章では、心理学実験で音声刺激の選定の際に必要となる、標準化された音声データベースの利用を視野に入れて、本装置での利用のためのデータ互換性を検討する。また3章では、音声を扱うために民生用の音響機器を利用する場合の留意点について検討する。4章では、パーソナル・コンピュータがマルチ・メディア化していく流れの中で、音声を扱う心理学実験装置としての関わりを考える。さらに、5

1) 名古屋大学大学院博士課程（後期課程）研究生
音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置

章では、本装置で心理学の実験に実際に利用されている音声情報処理技術を紹介する。そして、最後の6章では、本装置の今後の進展について検討する。

1. 実験装置の概観

名古屋大学の教育心理学教室に設置されている『聴覚及び認知実験用多目的装置』とその周辺機器を中心に紹介しながら、音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置における留意点を検討していく（図1）。記述にあたっては、心理学実験の実施の手順に沿って行う。まず、実験に用いる音源を記録したり、被験者や学習者が発声した音声を録音する段階に相当するサウンドライノスについて述べる。次に、実験の目的に応じた音声実験刺激を作成し、加工するための音声処理系について触れる。なお、この音声処理系の一部については、後述の別の章で改めて詳細に述べる。そして最後に、実際に聴覚実験を行う際の音声出力系について記述する。

1.1 音声入力系

音声の録音は、静音なところで行われる必要があると考えることは周知の事柄であるが、特に音声を対象とした心理学実験を検討する場合には、さらに細心の配慮が必要となる。実験刺激の作成のために、特に音声情報処理技術を利用した加工を行う場合には、録音時の静粛性は極めて重要な事項となる。

音声情報処理技術の多くは、雑音などが伴わない単一音源からの音声であることが前提とされており、その仮定が満たされた上で、はじめて有効に機能する。一方、例えば蛍光灯などから発生するノイズでも、条件によっては音声情報に悪影響を与えることがある。このようなノイズは、交流電源などの50Hz、または60Hzの周期を持っており、男性の声の高さ（pitch）と重複する部分がある。そのため、音声の加工にあたってpitchを利用するような音声情報処理においては、無音音などを音圧レベルの低い部分では致命的な影響を与えてしまうことがある。

人間であれば、カクテル・パーティー現象で知られるように、自分の周囲の喧騒の中から、自分に必要な情報を持つ音声に、それを何らかの効果をあたえながら選別して聞くことができる。すなわち、外ではセミが鳴き、道路を幾つかの車が通る音がした、夜の上ではラジオが楽曲を奏で、隣の家では子どもが呟いている音が聞こえてくる中でも、階段の下で自分が呼んでいる微かな声を聞き取ることができる。しかし、実際はこのことはあくも家事している。私たちが耳にする現実の音は、今述べたような全てのもののが、実に一次元の空気の振動として戻り込まれて集約されてしまうものである。もちろん、ラジオ聴取による音の位相差や時間差、強弱差などの情報も音の認識に重要な影響を与えている。しかし、人間は戦う聴取者さえも、上述のようなノイズの中から、音圧レベルの上では普通に小さい対象の音声を難なく分離して意識化させることができるのである。

図1 「聴覚及び認知実験用多目的装置」及び周辺機器の概要
このようなノイズと対象音声の分離は、それ自体が工学的な研究の重要なテーマである。例えば、工場内での騒音の中での音声認識の精度を上げたり、車中での音声通信のためのノイズ成分の低減といったことが精力的に検討されている。このことは、逆にノイズと対象音声の分離がいかに困難なことかを示している。

従って、一般的にはノイズと対象音声の分離は、録音後は事実上困難であると考えておるべきである。したがって、心理学実験に用いる音声刺激などの録音時の静粛性の重要性を改めて指摘できよう。

このような録音時の静粛性を達成する上で、本装置では、適音特性の優れた防音室（理研音響：聴力検査室AT-80S）を使用している。この防音室は実験室とし、100Hzの音に対して60dB以上、1kHzの音に対して70dB以上の適音能力を有している。そして、さらに音響処理を施した換気用デフレクトに接続されたエアコンディショナーや音響装置を用いたため、被験者に対して負担の少ない実用上適当な環境をもたらしている。

録音にあたっては、防音室内で発音した音を、まずマイク（ソニー：ECM-23PⅡ）でとらえ、その出力を防音室の外に取り出している。そして、ミキサー（ディアック：M-06）で出力を見積用システム・レルに上げた上で、オーディオ・アンプ（オンディオン：PMA-910V）を介し、ディジタル・オーディオ・テープ・デッキ（以後DATと記す）（ソニー：DTC-300ES）に録音している。

ところで、マイクの出力は信号レベルが極めて低いため、外部環境の影響を受けやすい。一方、コンピュータ関連機器などでは、それから発生する電磁波による電気的ノイズレベルが極めて高い。特にCRTディスプレイなどのノイズは激しい。従って、音声入力時、コンピュータ関連機器は全て停止させた上で、録音作業のみに専念している。そして、コンピュータを使用する次の音声処理系の作業段階には、常に分離させたステップでの実施に留意している。

1.2 音声処理系

音声処理系は、録音された音声から、当該の心理学実験に応じた音声実験刺激を作成し、加工するステップである。

このステップでは、まず録音された原音声の単語や短い文章などをコンピュータに取り込み、実験に応じた必要な単位ごとに切りとれて、それぞれを音声材料の音声ファイルとして保存していく。

音声をコンピュータに取り込むのにあたっては、DATからのアナログ音声信号をA/D変換して行ってある。具体的な装置について述べると、先述のDATからの音声信号を、まずエレアリング・ノイズの除去のため、選音周波数7.2kHz、低減率110dB/octのロー・パス・フィルタ（エヌエヌ回路設計ブロック：RT-8FLB2を使用した自作品）を通過させる。そして、次のA/D変換にあたっては、音声入力ボード（カーナーズ電子：Sound Master）を用いている。このA/D変換は、主にサンプリング周波数16kHz、量子化16bitで行っている。その後、パーソナルコンピュータ（日本電気：PC-9801VX21）に入り込み、音声データ・ファイルとしてハードディスク上に保存している。

ここで、音声信号のデジタル化に関連した事柄について、古井（1985）を参考にして簡単に説明しておく。まず、はじめから考えていくと、音声は空気の振動、すなわち空気の圧力の微細な変化である。そして、これをマイクなどの音響機器でその圧力を強さ、すなわち電気信号としてアナログ量の電圧に変換する。この電圧は、時間軸上で連続値として常に変化するものである。しかしこのままでは、コンピュータに取り込むには適していない。そこで、この電圧の高さを一瞬ごとにデジタル量である数値に変換し、その連続として、電圧の変化、すなわち音声の情報を記録しようとするのが Analog to Digital 変換、すなわち A/D 変換の考え方である。

したがって、特定の一瞬の電圧を数値化するステップをどれだけ短い時間間隔ごとに行うか、そして、その電圧をどれだけ細かいかぎりで数値化するかによって、音声の持つ情報をどれだけ保持できるかが決まる。よって、より繊細な音声情報を保持しようとすると、それに応じて膨大な記憶容量が必要になる。

また、数値化を行う時間の間隔は、サンプリング周波数で選ばれることが多い。例えば、サンプリング周波数が48kHzの場合は、一秒間に48,000回のA/D変換が行われることを示す。このサンプリング周波数は、A/D変換された音声における再生可能な周波数域に影響を与える。結論からいえば、A/D変換される当時の音声に含まれている周波数成分の情報の内、サンプリング周波数の1/2の周波数（特にナイキスト周波数と呼ぶ）までの情報は、A/D変換によるサンプリングによっても完全に保持されることが、情報理論におけるサンプリング定理によって確かめられている。したがって、高いサンプリング周波数でA/D変換を行う程、高域の周波数についてまで忠実な音声情報を保持することができる。この例では24kHzまでで情報が保持され、人間の可聴域である20Hz～20kHzをカバーすることになる。

しかし、A/D変換の対象となっている音声に、ナイ
音声を扱うためのバーソナル・コンピュータを用いた心理学実験装置

キスト周波数以上の周波数成分の音が含まれている場合には、それがエイリアシング・ノイズとしてデータに混入し、元来の周波数域の音声情報に対しての誤差をも含むことが考えられる。したがって、原音声に含まれる不要な高域成分をカットし、エイリアシング・ノイズの影響を受けないようにするため、A/D変換の前段にアンチエイリアシング・フィルタとして、ローパス・フィルタを設置することが必須である。このロー・パス・フィルタの遮断周波数は、常に先述のサンプリング周波数と対応するナイキスト周波数に応じたものを設定しなければならない。

一方、電圧をどれだけの細かいきざみで数値化するかは、量子化 bit 数で示される。例えば、量子化 16bit でサンプリングした場合は、当該の電圧は 2^16 = 65536 段階の数値で表される。この量子化 bit 数は、当該の音声のダイナミックレンジやS/N比に影響を与える。当然量子化 bit 数が大きい程、広いダイナミックレンジと優れたS/N比を得ることができる。

現状、本装置では、A/D変換ボードに接続するためのクオーツ発振による外部クロック・ユニットを自作し、サンプリング周波数16kHzで行うようにして、量子化16bitでA/D変換を実施しているが、音声刺激作成のために適切な条件については、後述の2-4章でも検討する。

さて、上述の条件での実際の作業により、実験刺激の単語や短文が一つ一つの音声ファイルとして保存されることである。例えば、両耳分離聴取などの課題の場合、右と左のそれぞれ耳のチャンネルで、その実験刺激を割り当てて再生するかを定めるだけで、提示にあたっての同期の問題などにも比較的容易に対処できる。

しかし、実験の目として、音声の音響的加工が必要な場合は、さらにまず、対象となる音声の音響的特性を調べる必要がある。すなわち、ここでは実験刺激の作成に先立ち、音声波形、サウンド・スペクトロ・グラフ、パワー、ピッチなどについて分析を行い、実験のための音声の加工すべき箇所をms単位で正確記録している。本装置では、この作業のため、バーソナル・コンピュータ上で動作する音声分析用ソフトウェア（NTT アドバンステクノロジー：音声工房、及び Voice Plotter）などを主に使用している。

このような、アプリケーション的な専用ソフトウェアの使用にあたっては、ソフトウェア各の音声ファイルの形式の相違に留意する必要がある。音声データのファイル形式は、他のアプリケーション・ソフトウェアのファイル形式と比較すると、極めて単純な形式なものであるが、それでもそれぞれ微妙に異なっている。

 même遺憾なことではあるが、音声に関連した様々な学会や研究会においてできさえも、不適切なファイル形式の音声データを分析したものを、そのまま発表してしまっているもののが散見される。そのような基本的な誤りは、そこから得られたデータや結果の信頼性にまで疑問を与えるものであり、十分に配慮がなされるべきことである。本装置においては、各種のファイル形式をデータ形式に対応するためのコンバータを作成し、それを利用することで、このような問題に対処している。

さて、音声の加工すべき箇所が確定された後、実験の目的に応じた音声情報処理技術を利用して、音声刺激の加工を行うことになる。本装置では、音声協の指定時間について時間軸上の区分長圧縮処理が行われることが多い。その詳細については、5章で改めて述べる。

本装置では、上述のような音声情報系の作業を通じて、心理学実験用の音声刺激を作成している。

1.3音声出力系

実際に被験者を迎える、音声を用いた心理学実験を行うスケジュールの音声出力系に取る。この音声出力系では、音声刺激の制御をバーソナル・コンピュータ上で行うことができるため、実験者が実験時のプログラムを作成すれば、実験者の意図に応じた様々な実験条件を実現することができる。

利点としては、例えばカセット・テープ・デッキなどでは、なかなか実現が困難で、被験者が試験の実施などで挙げられる。また、被験者数の刺激提示順序のランダム化も容易である。さらに、既存の反応時間測定プログラムなどを組み合わせることにより、実験自体の時間制御、ms単位での反応時間測定も可能である。また、阿部ら（1988）の音声提示のプログラム・ルーチンを組み合わせることにより、視覚や聴覚のクロス・モダリティの実験などのため、異なる刺激を同期させて提示することも可能である。

さらに、バーソナル・コンピュータからの音声出力に関しては、深田（1992）が、最近多く使われるようにになってきたEMSメモリの利用を試みたり、より長時間の音声の再生をするために、音声データの圧縮を行っている。本装置の構成においても、必要に応じて、このような成果を開発していくことが可能である。

さて、この音声出力系では、音声刺激はまずバーソナル・コンピュータから出力される。そして、次に先述のロー・パス・フィルタを通し、今度はスムージング・フィルタを通して通過する。そして、オーディオ・アンプで音量が調節された上で防音室内に入り、実験に応じてヘッドホン（ティアック：HP-200PRO）またはスピーカー
資料

（三菱電機：DS－11XL）から被験者に提示される。

なお、必要があれば、被験者の反応を得るためのスイッチ・ボックス（自作品）や図形表示用のCRTディスプレイを防音室内に設置し、防音室内に被験者のモニター用のサブ・ディスプレイを設置することもある。

また、実験によっては必ずしも防音室内で実験を実施せず、カセット・テープ・デッキ（ソニー：TC－WR870）を使用して、音声刺激を実験条件に沿って録音したテープを作成し、それを用いて当該の施設で実験を行うこともある。

以上が、実験実施のステップとしての音声出力系の概略である。

2. 各種の音声データベースとのデータ互換性

心理学実験においては、その実験刺激の選定にあたって細心の注意を払い、十分な時系列でのその採用を決定する。例えば、視覚刺激などにあたっては、様々な図形について、その複雑性などの評定により規準データが作成されたものがある（Vanderplas & Garvin, 1959）。そして、そのような規準に基づいて刺激を通じて、困難度の調節を行ったりする。また、言語材料などを使用する場合でも、日常での出現頻度や親和性を基にしたり、さらに無意味語においても、その選択基準を定められた刺激選定の参考にしている（林, 1972；国立国語研究所, 1962）。

したがって、実験刺激に音声を使用するような場合も、視覚刺激や言語材料の扱いと同様に、音声そのものの素性に対して、音響的な特性をはじめとする多面的な側面について標準化されたデータや、経験的な蓄積のあるデータの使用が必要とされる場合が想定される。このような条件が特に求められる場合には、音声認識や音声合成、情報通信の分野で既に構築されてきた音声データベースの利用が得られる、音声刺激に関する様々な統制が比較的容易になると考えられる。

しかし、これまでに様々な研究機関で作成されてきた各種の音声データベースは、それぞれの仕様も動作環境も異なる。ここでは、各種の代表的な音声データベースについて、その仕様やメディアなどについてまとめておく（表1）。

各々のデータベースは、研究目的の用途に限らず、他の研究機関での購入や使用が可能である。もちろん、それらは、それぞれの作成した機関間の用途や目標に応じた構成となっているため、その選択や使用にあたっては、利用者側の十分な知識と理解の上での検討が必要である。

音声データベースの利用にあたっての、現時点での課題は、これまで本装置を利用して作成してきた音声刺激などについて、その音響的特性を当該の音声データベースの音声試料と比較するような場合に、A／D変換時のサンプリング周波数や量化化bit数の違い、その直接的な相違を困難にしている点である。もちろん、その条件に沿ったA／D変換を改めてやり直せば、比較が可能であるが、以前の音声刺激で得られた心理学実験の結果の蓄積については再現性の保証がなってしまいます。また、既存の音声刺激のデータに、アップサンプリングやダウンサンプリングといった、計算機上でソフトウェア的な処理を施した上での比較も可能であるが、計算量が膨大であり、現在のパーソナルコンピュータ単体での性能ではやや荷重が重いため実用的ではない。さらに実験結果の再現性の問題もそのまま残ってしまう。

<table>
<thead>
<tr>
<th>表1 各種の音声データベースの仕様</th>
</tr>
</thead>
<tbody>
<tr>
<td>電  線  研  究  用  音  声  データベース</td>
</tr>
<tr>
<td>サンプリング周波数</td>
</tr>
<tr>
<td>量子化bit数</td>
</tr>
<tr>
<td>ロールオフ・フィルタ 遮断周波数</td>
</tr>
<tr>
<td>メディア</td>
</tr>
</tbody>
</table>


—231—
音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置

したがって、本装置を利用した音声刺激の作成の上で
も、音声集録の段階から、利用する音声デタベースと
対照できる条件で作成しておく配慮が必要となる。本
装置を利用した音声刺激作成では、当初はコンピュータ
の記憶容量の制限や、実験時の音声データの読み出し時
間の節約などもあって、サンプリング周波数を10kHz
で実験していた。しかし、1992年以降は、音質の向上
各種のデータベースとの対照、さらに音声分析用のソフ
トウェアでの利用の便宜を図るため、サンプリング周波
数を16kHzに移行して音声刺激を作成している。今後、
ATR音声・言語データベースや、磁気録の研究用音声
データベースなどの利用が想定される場合には、サンプ
リング周波数20kHzでの移行を検討する必要があるか
も知れない。今後も、様々な音声データベースの特徴を
見守りながら、検討をすすめていく課題であろう。

3. 各種の商用音響機器の利用

ここでは、コストパフォーマンスが高い、多用途に
使用できる商用機の一般的な音響機器において、音声を
扱う心理学実験におもに利用できる機器を検討する。

現在のところ、音声刺激用の原音を録音するにあ
たっては、DATを用い、サンプリング周波数48kHz、
量化化16bitのliner PCMのスタンダードモードでの
録音が望ましいと思われる。そして、その録音音声を
DATのデジタル送受信を利用して、デジタル信号のまま、
DAT・インタフェース・ボード（岩田アイセイ（IS-3690）
等を介し、適切な条件でダウンサンプリング
してパーソナル・コンピュータに直接取り込むのが、現
時点では最善であろう。

一方、カセット・テープでの録音は非常に手軽に行え
るのがだが、現在では残念ながら周波数特性やS/N比、
クロストークなどの観点から、原音の録音には必ずし
も十分であるとはいえ難い。

さて、最近、光磁気ディスクを用いたミニ・ディスク
（MD）や、従来のカセット・テープと同様の形状でデ
ジタル録音を可能にしたデジタル・コンパクト・カセッ
ト（DCC）といった、現在の主流であるコンパクト・
ディスク（CD）やDAT次の世代のメディアを目指
した新しいデジタル録音機器が相次いで発売された。

しかし、これらのメディアでは、前田（1993）や森本
（1993）が述べているように、音声や音楽をデジタル化
した場合の膨大なデータを、人間の聴覚特性などを利用して
し、データを圧縮して記録している。基本原理として
は、主に聴覚心理におけるマッピング効果と、等ラウド
ネス特性を利用し、量子化雑音が人間に認識されにくい
ように、音声の帯域ごとの量子化bit数を適応的に削減
して符号化を行っている。そして、その成果として、デー
タの圧縮率を1/4〜1/5へと高め、少量の記憶容量し
か持たないメディアにおいて、音楽鑑賞などの実用に耐
える高音質を実現したものである。

しかし、このような人間の聴覚特性を利用した音声
データの圧縮は、その評価主体に人間による官能検査を
必要とする。すなわち、その方式に対しての出来不出来
の評価そのものが、心理学的測定の対象となっているの
である。したがって、このような機器の利用については
、実施しようとする心理学的な実験の内容とよく照ら
し合わせて、その利用の適否を判断する必要があります。
例えば、音声の品質を問わない実験の教示などの音声提
示のために用いる場合ならば、カセット・テープなどの
利用と比べて遅かに好ましい。しかし、音声の特性その
ものが実験条件になるような音声言語理解・ヘッドホ
ン提示などによって音質の影響がより顕著に現れる可能
性のあるダイオード・リスニング課題などに用いるの
は、必ずしも好ましくないと考えられる。さらに、
音声情報処理技術を利用して、録音された音声を加工し
て実験刺激を作成するような場合には、データ圧縮時に
おける非線形の歪みが、刺激の処理過程に幾しくない影
響を与える可能性もある。

したがって現時点では、原音の録音にあたっては
DATの利用が好ましい。そして、録音された音声を
DAT・インタフェース・ボードを介して、ハード・
ウェア的にダウンサンプリングを行い、目的のサンプ
リング周波数での音声データに変換する。その上で、
パーソナル・コンピュータに取り込んで、所与の仕様で
管理するのが望ましいと考えられる。

なお、他のオーディオ・アンプやスピーカー、マ
イクやヘッドホンなどについても、業務用、商用用の区
分を問わず、音響特性ができる限りフラットなものを選
択するのが妥当といえよう。

4. パーソナル・コンピュータのマルチ・
メディア化

近年、パーソナル・コンピュータの世界では、グラフィ
カル・ユーザー・インタフェース（GUI）の進展に
みられるように、従来の抽象的な文字列操作によるコン
ピュータ操作の方法から、図形的な手がかりを重視した
直感的理解が容易な操作への指向を強めている。このよ
うな時流に沿って、コンピュータの操作や利用について
は、人間に親しみやすいインターフェースの設計に重点
をおかれてきている。そして、従来の文字や図形に加
え、人間の視聴覚を重視した、音声や映像による
インターフェイスが実現されつつある。もちろん、これ

— 232 —
までにも機種によっては、当初から音声の入出力を前提に設計されたものもあったが、最近、さらに多くのコンピュータでも音声の取り扱いができるようになり、形式も標準化されてきている。

このようなパーソナル・コンピュータのマルチ・メディア化の傾向は、従来よりも一層コンピュータで音声を扱えるようになる点でも、大いに進展が期待できる。そこで現時点において、音声を扱う心理学的実験を行うために配慮すべきだと思われる点をハード・ウェアとソフト・ウェアのそれぞれの観点から検討しておく。

まず、ハード・ウェア上の留意点としては、まず現時点でのマルチ・メディア用の音声ボードは、職場でのビジネス・ユースを前提に設計されており、音声分析用のボードとは用途が異なっている。具体的な問題点として指摘できるのは、アンチ・エイリアシング・フィルタ用のロー・パス・フィルタには、やや低減率の低いものが用いられていることが多い。また、サンプリング周波数に従って逆波周波数が変化するものは、残念ながら少ない。従って、音声データベースの音声試料などの、低サンプリング周波数での音声の取り扱いにおいては、エイリアシング・ノイズへの対処が別途必要になる。また、ステレオ出力のものには、クロストークの大きいかのもあり、厳密なダミオ・テンプレート・リンピング課題には、必ずしも理想的なものばかりではないことが指摘できるよう。

次に、ソフト・ウェアについては、前述の民生用音響機器のMDやDCCのところでも指摘したように、コンピュータにおいても、音声データの圧縮が留意すべき事柄となる。まず、膨大なデータ量になってしまう音声を少しでも圧縮し、可能な記憶容量を有効に利用し、コンピュータを効率的に活用するには極めて大切なことである。そのため、音声データの形式についても、デイタル圧縮のためにADPCM方式などを利用した非線形の圧縮が行われていたり、LPC方式を応用した圧縮方式の実用化をめざして標準化が進められたりしている。このような方式としては、自体が音声情報処理技術の成果である。しかし、その成果はそのもので、心理学的な検査の評価対象であり、限定された用途において、実用で許容される枠内での最善の品質をめざしたものである。したがって、音声を用いる心理学の実験においては、実験の条件に沿った品質のものであるかを、実験者が把握した上で利用が一層肝要となる。

以上、現時点でのパーソナル・コンピュータのマルチ・メディア化を巡る音声の取り扱い上で、配慮すべき点を検討した。今後、マルチ・メディア化の動向も、高機能化とともに、高品質化が進んでいくと思われる。そうすれば、高品質な音声の扱いも容易になり、より柔軟なデザインでの心理学実験を工夫していく上のプラットフォームになることが期待できるよう。

5. 音声情報処理技術の利用

心理学実験において、実験刺激として音声を用いる場合、実験条件に応じて音声を加工することが必要になる場合がある。例えば、外国人日本語学習者における日本語の発音の中で取り扱うとする場合、発音の作成のために音母音の持続時間の伸長や圧縮、促音作成のための無音部の挿入、また発語速度の制御のためにも、音声の時間軸上の伸長と圧縮を行える技術が望まれる。また、ダイナミック・リンピング課題は、左右のチャンネルから同時に提示される音声刺激として、[apa]と[aba]のように、母音子音の連続で、母音部分は共通で、子音部分の変更した音声を使用する場合ももれる。このような刺激として、実際の発話を利用するような場合は、発話者には音声部のタイミングが揺れるように、同じ速度で発話してもらえるのであるが、現実的にはどうしても微妙なタイミングが異なってしまう。そして、その音声を左右から同時に提示すると、子音部分が必ずしも同時に変わらず、そのままでは刺激としては不適切となる。そこで、刺激作成の段階で、はじめの母音部の持続時間を操作して、子音部の開始点を揃えるような作業が必要になる。

このような用途を満たす音声情報処理技術に、森田・板倉（1986）の考案による、音声データの時間軸上の制御のためのボインターステートメントにより重複加算法（Pointer Interval Control OverLap and Add; PICOLA）がある。

ところで、いわゆる音声情報処理技術としては、LPC分析合成やPARCOR分析合成などが有名であるが、これらと比較してPICOLAは、情報通信の観点からの情報圧縮符号化の側面については劣るが、音声データの波形を直接操作するので音質の面や計算量の利さな点で優れており、聴覚実験の用途に合致している。また塚本・東倉（1990）も、乳幼児の泣き声についての研究で、その音質を生かしてPICOLAを制御用に用いている。さらに、本装置では、この方式をもとにして、音声の指定区間のみに処理を行うように設計仕様を変更したPICOLA plusを作成し、音声処理に利用している。

このPICOLAでの音声データの処理方法は、基本的に自己相関法などで抽出できる音声波形の周期性を利用した波形の挿入や削除である。しかし、それは必ずしも単純な波形の繰り返しや削除ではない。まず処理にあたっては、対象を繰り返す2周期間の波形に限定する。
音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置

そして圧縮の場合には、まずその2つの波形の特徴を生かして重ね合わせた1周期分の波形を作り出し、その他の2つ波形の代わりに挿入する処理を行う。また伸長の場合は2周期分の音声波形から、さらに1周期分の音声波形を余分に作り、2つの波形の間に挿入する処理を実施する。さらにこのような処理の後に、次に処理すべきデータ長を示すボインターやの移動を制御することによって、任意の圧縮と伸長を実現するのである。このような技術を利用すれば、従来のようにも不自然な印象が残ってしまう合成音でなく、自然な音声をそのまま生かした実験状況を実現でき、Liberman（1982）などが考案してきたような、実験に先立った合成音声になるための手続きもおくことができる。

PICOLAについて、森田・板倉（1986）をもとに、さらに説明する。音声の時間軸での圧縮・伸長技術として最も単純な方法としては録音テープの再生スピードを調整することであるが、本方法では容易に推測できるように、音声信号の周波数成分も変化してしまう、ピッチの変化。話し声の欠落、さらに音量の明瞭性を含め、わずかな圧縮・伸長により損なってしまう。それに対し、本技術は録音音声のピッチ情報や個人性情報を失うことなく、時間だけを圧縮・伸長する技術であるため、音声情報検索の効率化。外国語教育の補助。聴覚障害者の聞き取りの補助。音声情報圧縮等、さまざまな応用が考えられる。

さらに、具体的なPICOLAの原理を述べる。処理にあたって、まず音声のpitchを抽出するために、入力波形に対して、分析用ポイント（▼で示す）と分析フレームを設定する（図2）。分析フレームの長さは、予想される最大pitch周期とほぼ等価にする。分析フレームに対して自己相関関数を計算し、その最大値をとる時間遅れをTpとし、周期性の強い波形AとBを決定する（図3）。

次に、入力信号の時間長に対する出力信号の時間長の比を伸長・圧縮率と呼びRで示す（R=出力信号の時間長/入力信号の時間長）。この定義によれば、圧縮の場合はR<1、伸長の場合はR>1となる。

まず、圧縮の場合について説明する。まず音声波形Aに対しては1から0ヘ、Bに対しては0から1へ直線的に向かう重みをつけて加え合わせ、長さTpの音声波形Cを作る。これらの重みは、Cの前後の接続点での連続性を保つために設定したものである。次にポイントを、C上でLc=R×Tp（1-R）だけ移動する（図3、▼で示す）。以下、これを次のポイントとみなして同様の操作を行う。以上の操作で、長さLc+Tp=Tp/(1-R)の入力音声波形から、長さLcの出力音声波形Bを生成する（図4）。

図2 自己相関関数を利用した入力音声のピッチ周期の抽出

図3 PICOLAによる音声の時間軸上での圧縮

図4 PICOLAによる音声の時間軸上での伸長
伸縮比も細かく指定するようになっている。しかし、前述のPICOLAの処理方法からわかるように、実際の処理を行う時間の最小単位としては、1ステップ分の伸長・圧縮の処理を行う時間以下の単位での制御はできない。
したがって、指定区間の伸縮処理の終了は、計算で予測される伸縮時間を基準にして、PICOLAの処理を繰り返していく、処理される音声の時間が、基準時間の前後の最も近くなる時点で処理を打ち切っている。ここで現れる誤差は、音声そのもののpitch周期、分析フレーム長、自己相関関数を求める時の時間遅れ下限値、伸縮比の相互の関連により発生する。したがって、PICOLA plusを用いて1msごとに変化するような刺激音体を作成することは現実的ではなく、実際に音声刺激を作成した後で、目的の操作が達成されているかどうかを確認することが重要である。

以上、本装置で主に用いられている音声情報処理技術について説明した。

6. 今後の発展
本装置のメリットは、心理学実験において音声を扱う上で、音声操作のための専用機器ではなく、パーソナルコンピュータを用いた点である。すなわち、今後現れてくるであろう新規の優れたハードウェアを増設するような場合でも、これまでの蓄積を生かしたまま機器拡張をすることができる。また、ソフトウェアの進展により、現時点の本装置ではまだ使われていない優れた音声情報処理技術の利用も可能である。例えば、DSPを利用することにより、PARCOR分析合成をパーソナル・コンピュータ上で実用的な処理速度で実現した今川・鶴谷（1989）の開発による「音声録音鳴」の利用などにも対応できるよう。

今後、コンピュータや音声情報処理の発展に応じて、音声を取り扱うための実験装置の整備も継続していかなければならない。そして、このような装置を活用することにより、従来の視覚優位であった心理学の研究に加え、音声知覚に代表される聴覚からの入力を中心とする人間の認知などの心理事象を、一層明らかにしていく心理学の研究が、今後、精力的に進められていくことを切望している。

引用文献
阿部純一（編）1988 パーソナル・コンピュータによる心理学実験入門プログラミング ブレーン出版
藤本健文 1993 ディジタル・コンバート・カセット
音声を扱うためのパーソナル・コンピュータを用いた心理学実験装置

（DCC）の開発に携わって、日本音響学会誌，49，284-292。
深田昭三 1992 パーソナルコンピュータにおける音声
発生ルーチンの作成 平成4年度文部省科学研究費
重点領域研究（2）研究課題番号04027022 「日本語音
声教育の社会方言学的音響学的研究」 88班研究
成果刊行書
古井真治 1985 デジタル テクノロジー シリーズ⑧
デジタル音声処理 東京大学出版会
森本直子 1976 ノンセンスシラブル新標準表 東京大
学出版会
今川 博・相谷 滋 1989 DSP を用いたビッチ・フォ
ルマン実時間抽出とその発話訓練への応用 電子
情報通信学会技術報告 SP89-36, 17-24.
板橋 晴 1992 文部省「重点領域研究」による音声デー
タベース 日本音響学会誌，48，894-898.
河合 優・吉崎 一人・伊藤 学・1989 マイクロコン
ピュータを用いた関用音声記録・再生装置 心理学
研究，60，113-121.
小林 哲則・板橋 晴・速水 惟・竹沢 寿幸 1992 日本
音響学会研究用連続音声データベース 日本音響学
会誌，48，888-893.
国立国語研究所 1962 国立国語研究所報告21 現代語
表記五十種の用語用字 第一分冊 総記及び語義表

国立国語研究所
Liberman, A. M. 1982 On finding that speech is
前田保雄 1993 ミニディスクシステム 日本音響学会
誌，49，277-283.
牧野正三・二矢田勝行・真倉昌雄・城戸健一 1992 東
北大－松下単語音声データベース 日本音響学会
誌，48，899-905.
森田直孝・板倉文忠 1986 自己相関法による音声の時
間軸での伸縮方式とその評価 電子通信学会技術報
告，信学技報，66(25)，9-16.
中谷和夫（監修） 1975 パーソナル・コンピュータに
による心理学実験入門 ブレーン出版
垣坂芳典・浦谷則好 1992 ATR 音声・言語データベー
ス 日本音響学会誌，48，878-882.
中野眞也・速水 惟 1992 電気装置の研究用音声データ
ベース 日本音響学会誌，48，883-887.
塚本多子・東倉洋一 泣き声の時間構造とカテゴリー判
断の関係 日本心理学会第54回大会発表論文集，
522.
Vanderplas, J. M. & Garvin, E. A. 1959 The Associa-
tion of Random Shapes. Journal of
Experimental Psychology, 57, 147-154.
（1993年8月25日 受稿）
ABSTRACT

A psychological equipment experiment for controlling speech-sound stimuli using a personal computer: for the studies of spoken language cognition, music perception, dichotic listening tests, and cross modality tasks

Teruhisa UCHIDA

These days, personal computers are used frequently in psychological experiments. In Japan, the major personal computers have no sound recording devices, so, thus we have seen few cases using computers in which speech-sound stimuli is controlled. As for the studies of spoken language cognition, it is believed that computers should be used for processing sound data. The purpose of this study is to make contribution in the fields of auditory and cognitive psychology. At first, a multi-purpose equipment used in auditory and cognitive experiments at Nagoya University was introduced and inspected points to be careful in the cases of managing speech-sound stimuli were studied. In addition, the compatibility of data with various speech-sound database was examined, the use of various general audio-visual apparatus was considered from various viewpoints, and the problems and possibilities of using personal computers installed with multimedia expansion introduced. Then the PICOLA plus was introduced. This is a sound processing technique (time compression and expansion of speech sound) used in this type of equipment. In the end, views about using this equipment from now were examined.