# 音声情報工学雑感

高 橋 秀 俊*

音声というものがこれほど複雑微妙なものであろうとは，専門外の者には容易に想像できないことであろう。こういう私も学生の頃手軽に音声の合成を試みて失敗した一人である。それは縁に波形を切った円板をスリットの前で回転させて，脈動する光を光電管で受けるという極めて幼稚なものであったが，ほとんど同じ原理と考えられるトーキーフィルムがうまく行くのに，これがなぜ成功しないかは，ずっと後になるまでわからなかった。音声の認識の問題にしても昔はフォルマントがわかればよいぐらいに簡単に考えていたようである。

要するに音声というものの物理的実体は生身の人間が自然的な状況において発する言葉という具体的な形においてのみ存在するものであり，母音とか子音とかいう我々が日常使う音素の概念は，もっと抽象化されたものだということである。そもそも音声が発せられるまでには，まず脳の中で言葉を発しようとする意志がはたらき，その言葉の発音に対応して筋肉への指令が発せられ，それによって動力学的に時々刻々の発音器官の形がきまり，それが更に音の波形をきめるという何段かの段階がある。そうして音素の概念は物理的な音波や発音器官の形と 1 対 1 に対応するものではなく脳の中での事象に対応するものだという認識がなかったのが失敗の原因である。このことは一部の先覚者は昔から意識していたのだろうが，これが広く認識されるようになったのは最近のことであろう。

音声の人工発生，自動認識を中心とする音声情報工学はこれから大発展が予想されるが，それは結局このような事情を考慮した複雑な情報処理機能を含む逐次制御，逐次判定の装置が今日の電子計算機の技術の応用により可能になったからにほかならない。そこで今後の人工音声技術は子音，母音の波形のみならず抑揚，リズムその他考えられるすべての要素を加味したものであるべきであり，そのようにして聞いて自然な音声を発生させることに努力すべきであろう。そのような技術は決して単なるいたずらではなく，電子計算機の出力の一つの形として，オペレーターその他の人間への速やかな情報伝達に利用されようとし，そのほか，駅や空港でのアナウンス，試験場での呼び出し，その他比較的限られた範囲で変化する内容を声で伝達するのに有効な手段となるであろう。

音声(音語)の自動認識の場合は尚更のこと人間が自然の状態で発する言葉を認識できなければならない。その場合はいわゆるパタン認識であって，合成よりはるかに困難である。そこで必要な複雑な情報処理に電子計算機が使われることはいうまでもないが，その際もやはり音声のフォルマントのような波形に関する情報のほかに強弱，高低，継続時間等の

* 東京大学教授，理学部物理学科

我々人間に意味をもち得る情報はすべてを利用することを考える必要がある。また，人間は話をきく場合，どのようなことが話されるかについて，かなりの予想をもっていて，比較的少い可能性のうちからどれであったかを判定するという場合が多いことに注意する必要がある。したがって，たとえば voice dialing のような具体的な応用を考える際も，前述のような多種多様の情報の中から，何がその場合の discriminatory feature であるかを解明することが本質的であろう。しかし，そのような discriminatory feature は単一ではなく，redundant にしておくことが判定を確実にするために必要である。また，そのような判定のための raw data に対して情報処理をする際，できるだけ analog 的に，つまり連続変数的な処理を進め，yes-no の判定は最終段階で行うようにすることを，この種のパタン認識の問題を扱う際の指導方針とすべきであるというのが私の考えである。

以上，音声に関して私は現在何もしていないが，傍観者としての感想を述べた。