健康文化　12 号
1995 年 6 月発行

健康文化

音の世界

若栗 尚

最近、ある委員会の議事録をつくる羽目になって、久しぶりにテープに録音された話声を聞いて書き取る作業をやった。

録音がモノラルでされていたことと、委員の声にそれほどなじみがなかったせいもあって、少々、てこずった。

さらに、用語、特に略語の知識が、まだ、薄いことも影響した。

単独で話しているときはさほどでもないが、２人、３人と同時に発言が重なると、聞き取りにくくなる。現場で聞いているときよりもモノラルのテープをヘッドフォンで聞くときのほうが聞き取りにくい。特に、声質のよく似た人の場合には、なおさらである。

周囲雑音などスペクトルの違いがはっきりしているものに対しては、フィルターで相当な改善が得られるが、よく似た話声同士では、フィルターの効果は、あまり期待できない。話速や話し方の癖がはっきりとわかっている人が相手なら、まだ、区別し易い。

以前に、同じ様なことをやっていたときには、もっとはっきりと聞き分けられた様な気がして、なにが原因だろうかと考えた。

第一に、収音の方法が違っている。使用するマイクロホンの数が多く、話す人の近傍で収音しているので、明瞭度がある。さらに、それをミキシングして、左右のチャンネルに振り分けて、定位（発音源の位置）がはっきりしたステレオ収音としていたことである。聞き取るときに、空間的な方向の情報は、大変、役に立つ。ヘッドフォンによる頭内定位（ヘッドフォンでステレオを聞いたときに頭の中に音像ができて、頭蓋骨の中に、へばりついている様な感じになる定位）でも、方向性はあるので識別しやすくなる。右から聞こえる話と左から聞こえる話とは、両方が一カ所から聞こえるのに比べると、ずっと区別しやすいものである。これは、カクテルパーティ効果といわれるものの一種であり、発言者の位置がわかるので、誰の発言か確認できる。

多数のマイクロホンを使ってミキシングして、それを左右に振り分けるという面倒な方法によらなくても、普通のステレオマイクロホン（市販のどこでも手に入る安価なものでも）を１本だけ使うことでも、十分に効果は得られる。普通は、互いに向き合った２列の座席を設け、その間にステレオマイクロホンを設置して、座席の長手方向を左右に分けるように収音する様にしている。

以上のように、位置、方向の情報は、このような音源を確かめようとする時には、大変役に立つものといえる。

よく見かける情景だが、電車の中などでラジオ放送を聞くようなときにも、この効果は有効である。

第一に、イヤフォン（インナーイヤフォン）（ステレオ用またはモノラル用両耳イヤフォン）を両耳に装着することで、外部からの雑音に対して遮音効果が得られる。このあたりのことを考えて、実際に、多くのメーカーから、モノラル用の両耳インナーイヤフォンが発売されている。

また、頭内定位の現象も、ある意味では、周囲の雑音に対して放送の音を区別するのに有効に働いているといえる。これは、信号がモノラルの場合でも効果はある。

最近は、電車内に受信した放送波を増幅、再送信する設備や、地下鉄や、地下街、地下駐車場などでの放送波再送信設備が増えてきていて、ＡＭ、ＦＭラジオ放送がいろいろな所で、明瞭に受信できるようになってきている。AM放送のステレオ化も進んでいるので、ＡＭステレオ放送の受信できる受信機を持つ意味も増えてきていると思っている。まして、雑音の中で聞きやすいとなれば、なおさらである。

ステレオ・カセット・テープレコーダー（ウォークマン等）が、屋外や電車内などでも使われることも、このような特性による所も大きいのかもしれない。もっとも、カセット・テープレコーダーを聞いている人の多くは、あんな大きな音量で聞いていると、そのうちに難聴になるのではと他人事ながら心配になるぐらいの音量で聞いているので、あまり関係のないこととも考えられる。

話は替わるが、以前は、ＦＥＮの放送などを聞いていても、相当に内容が掴めていたように思っていた。しかし、いつのまにか、だんだん、わかりにくく

なって来てしまったような気がする。英語を聞いたり、正確ではないにしろ、話したりする機会が、だんだん、少なくなってきたせいかと思っていたが、どうもそれだけではないようである。特に、話速の速い話ほど単語そのものが聞き取れていない気がする。まだ、そんな年でもないのにと思っていたが、先日、古巣のＮＨＫ技研の宮坂さんの書いた文章をみていて、ガックリとした。

それによると、ここの所、数十年間の間に、放送などでの話し言葉の話す速さ、即ち、話速がどんどん速くなってきていて、数十年前には、毎分２５０語であったものが、最近では、毎分３２０語から４００語になってきていて、速いキャスターでは、毎分５６０語にもなるというデータがあるとのことである。

それで、若者には、こうした早口が好まれるが、「早口で聞き取れない」、「落ちついて聞けない」などと言う人がいて、話速の遅い「ラジオ深夜便」の放送が、お年寄りに受けているというのである。
これでは、まるで、「年をとったから、速い話に付いていけなくなったんですよ」といわれているようなものである。

確かに、人の聴覚機能は、年とともに低下する傾向があることは、事実であるし、単に、高音域の周波数特性が劣化することや、最小可聴限が高くなるだけでなく、中耳以降の内耳の蝸牛内の有毛細胞やそれに続く神経系、中枢系の全ての部位で劣化があり、統計的には、６５歳を過ぎると劣化が大きくなるとも言われている。
内耳以降の機能が低下すると、語音識別速度（言語の識別に要する処理時間）が小さくなる（処理時間が長くなる）、雑音の下での言葉の聞き取りの困難さが増す、短期、長期記憶が低下する等の現象が現れてくる。
これでは、ただ単に音声の入力レベルを上げてやったり、周波数特性を変えてやっても改善には繋がらないことになる。

話し言葉の識別に時間が掛かると言うことは、順次に入力される音声に対して、中枢での言語演算処理に時間が掛かることであるとして、この識別速度に合わせるように、入力音声の話速を制御することを考えて、

（１）　変換後の音声の品質劣化が極めて小さいこと
（２）　変換後のピッチや個人性が変換前と同じであること
（３）　ユーザーが自分で最適話速に設定できること

（4） 操作性がよいこと
（5） リアルタイムで変換できること
などを条件に話速の変換の研究を行っている。

簡単に説明すると、この方法では、入力音声信号を無音区間（息継ぎのポーズなど)、無声区間（子音など)、有声区間（母音など）に分けて、このうち有声区間と無音区間の波形を細工して、話速を変換している。
無声区間は発声者の個人性と音韻性を保つために細工しない。

有声区間では、ピッチを抽出し、その有声区間の音声波形をピッチ周期毎に分割する。さらに、このピッチ周期ごとに波形の補間（話速を遅くするとき）や間引き（話速を速くするとき）をして、もとのピッチを保ったまま話速のみを変換する。
その後、無音区間、無声区間、変換された有声区間を元の順番に合成して、最終的な合成音声をつくり、変換された話速の音声を得ている。

この形の変換では、話速を一定倍率で変えるため、話速を遅くすると、必然的に音声区間が延びることになるが、無音区間以外は、音声の有音区間だけを変換しているだけであり、100 秒の長さの音声を 1.2 倍に引き延ばしても 120 秒と 20 秒延びるわけでなく、ニュースの話声などでは、無音区間に変化を加えない場合は、9.1 秒程度延びるだけである。この伸びも解消する方法が考えられ、シミュレーションが進められている。

このようにして話速を変換してやれば、識別速度の低下を補ってやることができ、ご老体にも速い話速の話が楽に聞き取れることになる。
補聴器などへの利用も背景音と音声の分離ができれば、有望である。
また、この方法によれば、外国語に対しても同様な効果があるので、語学の学習や外国語ニュースの聴取にも応用できることになる。音声区間の話速をそのままにして、無音区間だけを引き延ばしても、直前のセンテンスを理解する時間が増加することになり、利用の道もあると考えられる。

こういう方法が可能になってきたのも、ディジタル信号処理のためのＩＣや手法の開発が進み、演算処理の高速化が進んだことが原動力になっていると考えられる。　今後、ますます、いろいろな方面で音声信号へのディジタル処理

技術が盛んになり、便利な機器が出来て来ることと楽しみにしている。

（財団法人空港環境整備協会・航空環境研究センター）