

次世代オーディオ ネットワークマルチメディアと音

1996年11月20日

早稲田大学理工学総合研究センター 山崎 芳男

キャンパスに学生戻る

最近学生が大学によく来るようになった。講義が急に魅力的になったわけではない。大学では自由にインターネットが使えるからである。

学生と接していると、特に生活を切りつめても買うものを見ていると、世相がよく分かる。60年代はテレビ、面白い番組があると水が引くように学生たちは帰ってしまった。VTRは学生の生活をテレビ番組から開放した。70年代後半になるとテレビ番組自体にあまり興味を示さなくなり、真っ先に電話を自分の部屋に引くようになった。80年代、ファックス付き留守番電話は学生の必需品となり、最近ではPHSや携帯電話を持つ学生も増えてきた。

パソコン購入のきっかけは音楽

この間、文科、理科を問わず学生たちが変わることなく親しんでいるのが音楽である。カラオケにも興味は示すが、むしろもっと積極的にバンドを作ったりオーケストラに入ったりと音楽は生活の一部となっている。

90年代に入り学生の手の届く価格でMIDIやCD-ROMを備えたパソコンが市場に出はじめると、彼らはこぞって買い求めた。パソコンはステレオ装置を買わなくともCDを楽しめる上、MIDIを使って作曲や友人との合奏が簡単に実現出来る、夢のような音環境を彼らに提供した。

パソコンは一台でレポート作成、電話、ファックス、インターネットさらに音楽や動画を楽しむことができる学生に欠かせない道具となった。

ハードはもう買わない

毎月のように高性能のパソコンが登場しているが、最新型のパソコンを購入する学生は意外なことほとんどいない。彼らの使用目的には先輩や友人の中古あるいは一世代前の安価なハードウェアで十分なことをよく知っているのである。

物のない時代に育った我々世代はどうしても最新

製品を購入してしまう。そしてすぐ古くなり、また新しい買い物をしてしまう。その上厄介なことに物を捨てられない。彼らは驚くほど物の所有に執着しない。

最新のパソコンを購入し、自宅でインターネットを通じて全ての情報を得ているパソコンマニアも中にはいるが、多くの学生は自宅では電子メールを日常の道具として使う程度で、インターネットは専ら学校で使っている。現在の課金システム、特に電話代は彼らには受け入れられていないのである。

自分で持つより安心

大学に行けば端末はどこにでもある。いや今や端末は町のあちこちにある。自分のハードウェアを持つ必要はない。ノートパソコンなど持ち歩かなくても手近な端末はたちまち自分専用のコンピュータに早変わりする。

ハードウェアばかりではない。考えてみれば膨大な、それも常にバージョンアップされるOSやソフトウェアを個人で所有する必要など全くないのである。

さらに彼らにとってはレポート、実験結果、場合によっては日記や私信でさえ自分の手元に置くよりはネットワークに委ねた方が安心なのである。

確かに将来どこに住むか分からないという特殊な状況に置かれているからかもしれないが、これは明らかに新しいスリムな生活形態の誕生である。

ネットワーク上に情報を登録しておくことにより、どこにいても利用できるばかりか自分の意志によってそれを他人に開放することも可能である。

近い将来CDやDVDと同じ物を内容のデジタル情報を複製するなどという愚挙はなくなるであろう。同じデジタルデータを何百万枚も複製するのは資源の無駄以外のなにものでもないからである。ネットワークを介してアクセスするたびに課金すれば済むことである。

この分では早晚自動車も、場合によっては家も個人所有しない、会社もネットワーク上といった時代が来るかもしれない。学校や大学の授業、学会、新聞や

雑誌、全国規模の放送局なども近い将来大きく形を変えることだろう。

コンテンツは自然に出来る

彼らは決して情報を受け取るばかりではない。自分の周りの情報をごく自然に発信している。いや発信というよりはまわりの出来事を使われるかどうかはともかく、取り合えずネットワークに登録しておくのである。この積み重ねがネットワークの価値を高めることを本能的に知っているに違いない。

よく言われることであるがネットワーク上では大企業も個人も平等である。むしろ決済や著作権といった制約と無縁な無数の個の結集が途方もない力を生む現場を見せられている。恐らくすでに多くの中学生いや小学生の力も結集されているのだろう。

世の中で現在作り出されている膨大なマスメディアに加えて全人類の私信にいたるまでの天文学的な情報を収容することなど不可能だという指摘もあるが、現在の技術動向からみれば記憶容量は十分確保出来る。

奥の深い音

百聞は一見にしかずと言うがマルチメディアにおいて音は映像やコンピュータデータに比べて従属的なものと捕らえられがちであるが、実は音は映像や文字情報に優るとも劣らぬ重要な情報伝達手段なのである。

普段意識しないが、目で据えられる範囲は前方に限られ、それ程広いものではない。したがって人間は元来後ろからの情報摂取は多くは聴覚、気配というものに頼っている。実際聴覚は後ろに対して驚くほど鋭い識別能力を持っている。

人間はたった2つの耳で様々な方向から来る音を聞き分けている。信じられないことであり、また1本のマイクロホンでは原理的に不可能であるが、人間は片耳でも慣ればかなり正確な音源位置の特定が可能である。これは永い経験でさまざまな方向や距離から来る音の特徴を正確に記憶しているからである。

バーチャルリアリティという言葉が頻りに耳にするようになったが、音の分野では古くは蓄音機、ステレオと音環境の忠実な伝送を志向していた。電話の通話品質は決して高いものではないが、電話はそれをVRと意識するかいなかは別として多くの人々に自然に空間移動の道具として利用されていた。

最近新幹線内や自動車運転中の電話が社会問題化しているが、これは音だけによるコミュニケーション

がいとも簡単に人間を現実空間とはまったく次元の異なる世界に引き込んでいる証である。

衣食足りて...ではないが、マルチメディアには快適な音環境が不可欠である。

方言や祭りの復活 過密過疎の解消も

言語は声という音を伴う人間だけの持つコミュニケーション手段である。言語はそもそも地球上のそれぞれの地域の自然環境あるいは民族等の環境に最も都合のいいように出来上がったもので、その数は無数といってよいほどある。

かつては日本にも生活に密着したたくさんの方言や村祭りがあったが、放送や新聞が普及するとともに急速に勢いを失いつつある。地球規模でみればたとえばナイジェリアには現在500以上の言語があるそうであるが、皮肉なことに学校教育の普及に伴い今世紀末には100程度に減ってしまうであろうといわれている。情報化や教育は重要であろうがさびしい話である。

マルチメディアを名乗る以上方言や様々な言語が再び盛んに使われるようになり、態率追求のあまり大都市に集中していた人々が故郷あるいは好みの土地で生活出来るようになる環境を提供しなくてはならない。

もちろんマルチメディアなど興味もないという人たちのところにまで土足でずかずか上がり込み、情報を垂れ流すようなことがあってはならない。

英語かエスペラントか

情報化社会に乗り遅れないように、幼児期から英語を教育すべきであるという声も聞く。またインターネットのお陰で英語に対する拒絶反応がなくなったなどと歓迎する向きもあるようだが、いかがなものだろうか。

英語が今以上にネットワークの共通語になったり、あるいはエスペラントのようなネットワークに適した共通語が誕生するかもしれない。もちろんネットワークへの参加を通じて英語や中国語が堪能になるのは結構なことであるが、多くの人々と直接語り合うにはやはり母国語、それも感情の機微が伝わる肉声が一番である。

母国語の肉声を自動翻訳

音については高態率符号化技術と通信速度の向上によりかなりの質の音をインターネットを通じて供給できるレベルに達している。インターネットを利用

したラジオ放送も始まっている。しかるに地球で使われる全ての言語でのサービスは到底不可能である字幕や合成音声尾を使った自動翻訳に頼るしか道はないだろう。

自動翻訳ソフトウェアは発展途上にあるが、既にホームページの翻訳程度であれば多少誤訳が混じるものの役立っている。

一方、不特定話者に対する音声認識ははまだ研究段階にある。ある程度の誤りを許容すれば、現時点でも自動翻訳ソフトと組み分わせることにより補助的役割を果たす道具としては十分機能するはずである。

母国語の肉声を聞きながら、最も理解しやすい言語で文字表示（必要ならば合成音提示）させることにより、発声の様子で感情を把握したうえ内容のある程度理解することができる。

様々な言語や方言に対応した手軽な音声認識ソフトウェアの要求が音声認識の研究を飛躍的に発展させる可能性も大である。

“文明人”の感性に頼るな

最近 MPEG (Moving Picture Encoding Expert Group) など人間の聴覚や視覚特性を積極的に利用した音や映像の高効率符号化方法が数多く提案され、既にビデオ CD や DVD V 等の符号化に使われ始めている。これらの規格・標準化作業は当然いわゆる先進国主導で行われている。

眼鏡や拡声器が当たり前の生活をして、知らず知らずのうちに聴覚や視覚が衰えてしまった、「文明人」の目や耳を使って高効率符号化を評価したり、方式を決定してしまつてよいものであろうか。地球上には目や耳の優れた人々がたくさんいるはずである。新しいシステムこそこれらの人の力を結集して人間が本来持っていた優れた感性を満足する仕様にしなければならない。

マルチメディアが人々の英知と感性を結集して、優れた聴覚・視覚能力をも十分満足する符号化方法などが実現できればと期待している。

自転車・鉱石ラジオの偉大さを

自転車、鉱石ラジオは偉大であった。鉱石ラジオは方鉛鉱や黄鉄鋼に針を立てて作る検波ダイオード、(後にはゲルマニウムダイオードが使われた)を簡単な同調回路(コイルとコンデンサ)と組み合わせただけのわずか数個の部品で構成されるラジオである。これにアンテナ(遠隔地の場合にはアースも)をつなぎ、ヘッドホン(当時はレシーバと称した)で放送を聞く。

放送局からの電力だけで放送を楽しむことができるすばらしいシステムである。

同じように自転車も、道路がある程度整備されているという条件もとではあるが、人間が到底一日で移動出来ない距離を自分の力だけでいとも簡単に移動することができるすばらしい道具である。

現在のマルチメディアに欠けているものはこの鉱石ラジオ・自転車的な道具としての素晴らしさではないか。

新しい信号処理

デジタルオーディオ パラメーター一般に標準化周波数が帯域を決め、量子化ビット数がダイナミックレンジを決定すると据えられがちであり、標準化周波数と量子化ビット数は密接な関係をもっており個々に論じられるものではない。

量子化雑音は入力信号の振幅が大きい場合には入力信号と無相関な量子化ステップ幅 Δ に一様に分布する電力 $\Delta^2/12$ の周波数成分の平坦な白色性の雑音となる。一方、入力信号の振幅が小さい場合には量子化雑音はその総電力は $\Delta^2/12$ と変わらないものの、入力信号と相関の強い聴感的に不快な歪となってしまう。

ところが量子化ステップ幅 Δ あるいはその整数倍に一様分布するディザと呼ばれる確率変数を量子化に先立ち入力信号に加算し、量子化器出力から同じディザをデジタル減算すると、量子化雑音は振幅の如何にかかわらず無く大振幅入力のとときと同様に入力信号と無相関な電力 $\Delta^2/12$ の白色性の雑音とすることができる。⁽¹⁾

従つて標準化周波数を高くすれば信号帯域内の量子化雑音は減少し、原理的には1ビット量子化で100のダイナミックレンジを確保することも可能であるが、標準化周波数は3GHzにも達してしまい実現不可能である。

高速1bit信号処理 情報理論の先駆者であるシャノンは1940年代後半に信号の平均情報量が伝送容量よりも小さければ無歪伝送可能な符号化方法が存在することを明らかにしている。これは標準化周波数と量子化ビット数が個々にではなく、むしろその積である情報量が重要な意味をもっていることを示している。しかし現実のシステムの構築に当つて標準化周波数と量子化ビット数の決定は一筋縄でいくものではないし、現行の標準化周波数や量子化ビット数が必ずしも人間の聴覚特性に馴染むものではないことは明らかである。

著者らは十数年来、標本化周波数と量子化ビット数の組み合わせの選択が困難であるのならその片方の極限である高速標本化 1bit 量子化で信号の情報量と人間の聴覚特性に見合った符号化方式が可能ならばと様々な検討を加えてきた。その結果、量子化雑音のスペクトル分布を制御することにより高速 1bit 符号化が可能であることを明らかにした。⁽²⁾

現行の 48kHz 標本化 16bit 量子化と等しい伝送容量である 768kHz 標本化 1bit 量子化で 100kHz 程度までの信号の伝送が可能ハードウェアシステムを実現し音響計測やスタジオ録音等実用に供している。

ところで前述のように聴覚特性を利用した各種の高効率符号化が実用に供され始めているが、その出発点が現行の CD, DAT と同じ方式におかれているのには疑問がある。ここで述べた高速 1bit 符号化方式などより広帯域の情報源符号化を行えば、伝送容量を節約したうえ現行のシステムよりも質の高い符号化も可能なはずである。

AD/DA 変換にハードウェアは不要 現在パーソナルコンピュータの入出力への導入や光無線伝送など検討している。本方式の導入によりパソコンから押の入出力用の AD/DA 変換器など能動素は切不要となる。

また、音場の能動制御では処理による遅延が問題となるが本方式では遅延のほとんどない高速処理が期待できる。1 ビットであるので乗算が単純なゲート回路で実現できるので LMS (Least Mean Square) 等適応におけるアナログ信号との乗算が著しく簡素化できるという特徴を持っている。

一般調和解析 コンピュータの処理能力の飛躍的な向上によりウィナーが 1950 年代に提唱した調和解析の原点ともいえる一般調和解析⁽³⁾ですら、時間はかかるもののパーソナルコンピュータで実行することが可能となった。⁽⁴⁾ 一般調和解析は観測窓内で原波形から残差成分が最小となる正弦波を抽出し、残差成分に対して同様の操作を繰り返すという単純明快な解析手法である。一般調和解析は定常的でない僅かな周波数変動に対しても正確な周波数成分の抽出が可能であり、観測区間を越えて信号の予測が可能であるなどの特徴を持ち、人間の聴覚特性によくなじむ解析手法である。東山、平田らはこれを音響信号の解析に適用した。我々は周波数成分の変化が正確に据えられることに着目した共分散処理による信号分離、ピッチ変換、時間軸圧伸や高速 1bit 信号処理への応用など試みてい

る。現在のところ膨大な演算時間を必要とするものの信号処理の飛躍を感じさせるものがある。

隣の家にも地球の裏側へも 10 円

現在のインターネットのもう一つの問題点は冒頭にも指摘した、家庭、あるいは移動体を通じての参加に対して大きな障壁となっている通信容量と運用コストであろう。

戦後間もないころ、多くの農村において極端な電話網の不足を補う手段として電話を共同利用したり、農事作業等の連絡、さらにある種のエンターテイメントにも有線放送が大活躍していた。当時の有線放送のように必要な人が自分の費用負担で隣の家あるいは最寄りの接続点までケーブルを引くかアンテナを用意すれば、運用費用など無料でネットワークに参加出来るシステムが望まれている。

実際日本の電話も以前は同一交換機内無料、ついこの間までは市内電話についてはいくら話しても 7 円、その後もしくは 10 円であった。現在のネットワークシステムはどこに接続しようがコストが変わるわけではない。地球の裏側だろうが隣の家だろうが無料、あるいは 10 程度のシステムは社会資本として実現してもいいのではないか。

社会システムがついていけるか

マルチメディアは専門家がいるわけではなく、人々の係わりかたも多種多様で、既に一人歩きを始め誰も行く先を知らぬ大きな動きとなっている。もはや法律を含めた今までの管理システムなど機能しようはずもない。情報に関する限り国境もなかった時代の地球と同じ状態に直面しているわけである。

文 献

- (1) 山崎芳男, “広帯域音響信号の量子化への大振幅ディザの適用,” 音響学会誌, Vol. 39, No. 7, 452-462 (1983).
- (2) 山崎芳男, 大田弘毅, 西川明成, 野間政利, 飯塚秀, “広帯域音響信号の高速標本化 1 ビット処理,” 信学技報, Vol. EA93-102 (1994.3).
- (3) N. Winer, “The Fourier Integral and Certain of Its Applications,” Dover Publication Inc. (1958).
- (4) 牛山 聡, 東山三樹夫, 飯塚昌弘, 平田能睦, “一般調和解析による波形分析,” 信学技報, EA93-103, 39-44 (1994.3).
- (5) 山崎芳男, “次世代デジタルオーディオに望むもの,” JAS コンファレンス予稿集掲載予定, 1996.11.8.
- (6) 山崎芳男, “マルチメディアと音,” 通産ジャーナル掲載予定, 1996.12.