2007－SLP－68 (5)
2007／10／19

音声技術の社会化

木村 晋太†

†株式会社アニモ 〒231-0015 横浜市中区尾上町 2-27
E-mail: †skimura@animo.co.jp

あらまし 我々は音声技術が社会で広く認知される技術となることを願って、従来の「音声技術の実用化」という言葉に代えて、「音声技術の社会化」という言葉を使っている。音声技術の社会化では、音声技術を新しい市場にいち早く提供するために、1) 音声本来の機能の正しい理解、2) 組込機器やインターネットでの音声技術の実装技術、3) 顧客の新しい価値観の把握の３つの観点が重要であると考える。音声本来の機能を知るには、発音や聴覚に障害のある方むけの開発を行うことが有効である。また、実装技術については、組込機器やインターネット向けの最新の IT 技術への対応が重要である。さらに、音声に関する理解だけでなくＣＳＲやＵＤといった顧客の新しい価値観に対する理解が必要である。本稿では、これらの３つの観点での、当社の取り組みを紹介する。
キーワード 音声技術、実用化

Socialization of Speech Technologies

Shinta KIMURA†

†ANIMO LIMITED, 2-27 Onoe-cho, Naka-ku, Yokohama city, 231-0015, Japan
E-mail: † skimura@animo.co.jp

Abstract Wishing that speech technologies are recognized widely in the society, we are using the phrase "socialization of speech technologies" instead of "practical use of speech technologies". We think the following three viewpoints are important in the socialization of speech technologies: 1) understanding of original functions of speech, 2) implementation using the latest IT technologies and 3) understanding of customer's new sense of values. To understand the original functions of speech, it is effective to develop the systems for persons who have difficulties in their pronunciation or auditory sense. Moreover, the latest implementation technologies for embedding speech technologies to small equipments and Internet are very much required. In addition, we have to understand customer's new sense of values such as corporate social responsibility (CSR), and universal design (UD). In this report, we introduce our activities in these viewpoints.
Keywords speech technology, practical use

１．はじめに

音声技術を事業として成り立たせることは非常に難しい。世の中でその事業が一人前の事業として認められるには、年間１００億円の売り上げあるいは５００人の従業員が必要ではないか。これを達成している音声事業は、世界的にも少ない。

当社は、富士通株式会社のベンチャー制度を利用し、音声技術の専門会社として作られた会社であり、音声技術を事業として成功させる使命を負っている。

当社では、音声技術が事業として成り立つには、まず音声技術の有用性が社会で認知されることが必要であると考えている。そのため音声技術の製品化や実用化という従来の表現にかえて、「音声技術の社会化」という表現を使っている。

音声技術の社会化では、音声技術そのものの開発のほかに、以下の３つの観点が重要ではないかと考えている

１）音声の本来の機能の正しい理解

そもそも音声とはどういう機能を持つメディアであろうか。我々の日々の生活では、音声コミュニケーションはなくてはならないものであるが、あまりにも当たり前のもので空気のような存在である。そのため、一般ユーザや音声技術者は、その機能を正しく理解していないことが多く、漠然と音声を使えば楽になると言ったりする。果たしていつもそうであろうか？

音声コミュニケーションは、メールやチャットなどのようなテキストによるコミュニケーションと同じではない。音声コュニケーションでは、パラ言語情報や非言語情報がたくみ使われ、コミュニケーションを円滑かつ正確にしていることは周知の事実である。パラ言語情報や非言語情報がないコミュニケーションは音声コミュニケーションではないと言ってよいであろう。最近の携帯メール世界では、顔文字や絵文字を使って、パラ言語情報を伝えることで、コミュニケーションの円滑さを実現しているようである。

従来の音声認識や音声合成という技術は、音声からテキスト情報を取り出したり、テキスト情報から音声を作りだしたりする技術である。ここでは、パラ言語情報や非言語情報がまったく欠落しており、これらを使って音声によるコミュニケーションを実現しようとしても、本来の音声コミュニケーションとはならないのは自明である。

音声を使ったアプリケーションやサービスを考える際に、パラ言語情報や非言語情報を如何に利用しているかが重要なポイントであると思われる。

音声コミュニケーションの機能を考察するときに、音声の発音あるいは聴覚に障害のある方を研究することは大いに役に立つ。発音ができないとどのように困るか、聴こえないとどのように困るかという点に、音声コミュニケーションの機能の本質が存在する。

２）音声技術の実装技術

音声技術をすばやく市場に投入するためには、実装技術が重要である。組み込み向けマイクロプロセッサー（ＭＰＵ）の高性能化、インターネットの高速化など、最近のＩＴプラットホーム技術の進展は目覚しい。音声技術は、これらのＩＴプラットホーム技術の上で利用するものであり、これらへの、素早い対応が重要である。また、さまざまなターゲットに対応するためにスケーラブルな構成を取れることが重要である。

３）顧客の新しい課題と価値観の把握

企業ユーザの従来の関心事は、いかに売り上げを増やし、利益を上げるかであった。しかしながら、最近これは大きく変わってきている。最近では利益を上げるだけの企業は、社会での存在価値を問われている。企業ユーザの価値観は、ＣＳＲ（企業の社会的責任）やＵＤ（ユニバーサルデザイン）という観点にも拡大してきている。これらのＣＳＲやＵＤの分野では、音声技術が活躍できる場面が多くなってきている。

以下、これら３つの観点での当社の活動を紹介する。

２．障害のある方や高齢の方むけの音声技術

音声コミュニケーションの本質を知るために、音声の発声は聞き取りに障害のあるかた向けのシステムの開発が有用であることが多い。当社では、このような観点で、この分野に取り組んでいる。

２．１　スピーチトレーナー

スピーチトレーナーは、聴覚障害児向けの発音訓練システムである。聴覚障害児は自分の声が聴こえないために、フィードバック制御がきかず、発音をうまく制御することができない。たとえば、ピッチ周波数を画面に軌跡として表示することにより、視覚的にフィードバックを与えることによって、適切なピッチで発音できるようになる。訓練を繰り返せば、視覚的なフィードバックがなくても適切な発音が可能となる。図１は、スピーチト

レーナーでピッチ（声の高さ）を訓練する画面の例である。

健常者の場合でも、音声がどのようなものかというのは、音声を聴いただけでは説明しにくい。音声を数値化、可視化することで、音声を的確に捕らえることが可能となる。音声の診断、教育といった場面では、音声の数値化、可視化は重要な技術である。

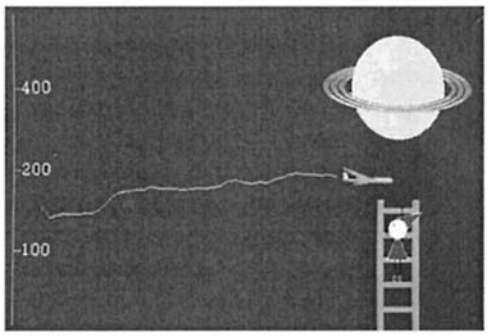

図１　スピーチトレーナでのピッチ訓練画面

２．２　花鼓

脳溢血や脳梗塞の後遺症として、失語症という病気がある。失語症は、概念とことばのリンクがなくなり、物の名前が発音できなくなる病気である。

花鼓は、全体構造法という理論に基づき、マルチモーダルな刺激を使ってことばを訓練する失語症のリハビリシステムである。聴覚、視覚、触覚、運動というマルチモーダルな刺激により、概念とことばのリンクを再形成する。また、音声については、分節素的情報よりも韻律的情報の訓練を重要視している。これは人間の赤ん坊がことばを習得するプロセスと似ていると言われており、新しい言語の学習への応用も期待されている。

図２　花鼓の利用シーン

２．３　ポケットリズム

ポケットリズムは、ボタンを押すと一秒程度の一定間隔の振動を発生する小さな箱である（図３）。

吃音の方は、歌を歌うときにはまったく症状が出ないことが知られている。外部からのトリガがあれば吃音にならないのである。ポケットリズムの振動をきっかけにしてスムーズに発音を開始することができる。またパーキンソン病の方は、この振動をトリガにして、スムーズに歩行を開始することができる。

図３　ポケットリズム

２．４　音声会話エイド

喉頭摘出された方やＡＬＳ患者の方で発声が困難な方のための音声合成を搭載したＰＤＡである。日常会話の例文集をあらかじめ用意してある（図４）。個人用に例文集を編集することが可能である。例文集の文をペンで選択することにより、音声を合成することができる。

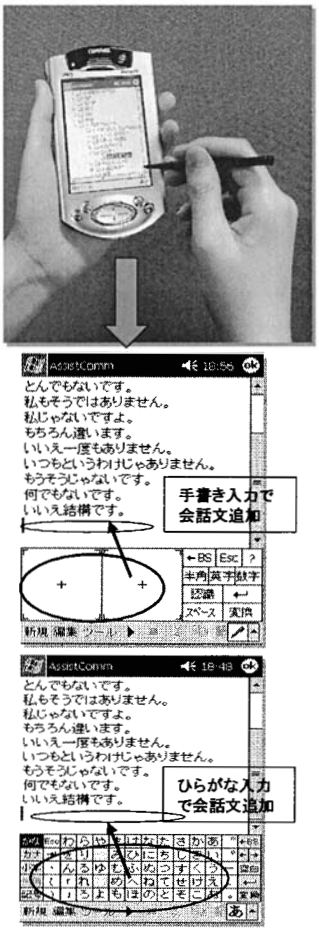

図３　音声会話エイドの画面

２．５　話速変換

高齢の方では、テレビやラジオのニュースの話すスピードが速すぎて聴き取れないことがおこる。視覚に障害のある方では、情報は耳からしか入ってこないために、神経が聴覚に集中し、健常者に適切な通常の話速では、遅いと感じることが多い。

この技術の健常者への応用としては、語学学習のコンテンツにおいて初期の段階では、発声の速度が速すぎると感じることが多く、話速変換により遅くすることが行われる。また、ある程度学習が進んだ状態では、高速の外国語の音声で耳を慣れさせ、通常の速度で聴くときに聴き取りやすくするという速聴という手法がある。

３．音声技術の実装技術

開発された音声技術をいち早く市場に投入するには、高度な実装技術が必要とされる。ここでは、機器組み込み向けの技術とインターネットサービスのための技術について述べる。

３．１　組み込み向け技術

音声技術を、携帯電話、カーナビ、ＩＣレコーダ、電子辞書等の専用機器に組み込んで提供することが、よく行われる。

この場合、元のソースコードを専用機器向けにポーティングするが、その場合によく問題になるのが次の４点である。以下のような配慮が必要となる。

a)　処理量

組込用途ではリアルタイム処理が要求されることが多いが、搭載されているＭＰＵやＤＳＰの処理性能（ＭＩＰＳ値）で、リアルタイム処理が可能かどうかが問題となる。

処理量が足りない場合には、処理の細かさを荒くするなどして処理量の削減を行うが、当初の性能からの低下を伴うことがあるので、シミュレーションによる性能評価を実施することになる。

b)　メモリ使用量

最近では、携帯電話やカーナビでは、かなり大きなメモリが利用可能となってきているが、ＩＣレコーダや電子辞書では、メモリの使用が大きく制限されている。

また、メモリ管理については、システム関数を使用せず、上位アプリから渡されるメモリ領域を使って、音声処理内部でメモリを管理する。

メモリ量削減のためには、他の処理とのメモリの共用化、処理単位の上限を小さくすることが行われる。

c)　プラットホーム依存性

プラットホーム関連で問題となるのが、ハードウェアによる浮動小数点演算の有無、システム関数・エンディアン・構造体のパッキング・ワードバウンダリなどの違い、ファイルへのアクセスである。

方式設計時から浮動小数点演算をできるだけ使わない方式とする、システム関数を使わない、エンディアンが関係するようなトリッキーな処理は行わない、構造体の要素はワード単位にする、ファイルアクセスはシステム関数を直接用いずに独自のファイルアクセス関数を用いるなどの配慮が必要である。

d）コーディング

処理系によっては、ＣコンパイラだけでＣ＋＋コンパイラが提供されない場合があるので、基本はＣ言語でのコーディングとする。アプリ側でＣ＋＋のＡＰＩを要求される場合があるため、Ｃ＋＋のＡＰＩラッパーを用意する。

３．２　インターネット向け技術

１）音声Ｗｅｂサービス

ネットワークの高速化に伴い、音声処理をサーバーに集約し、ネット越しに音声処理サービスを利用することが行われるようになってきた。従来はイントラネット内でのサービスに限られていたが、最近では、インターネットでのサービスも利用可能となっている。

当社が提供するＳｐｅｅｃｈＦａｃｔｏｒｙでは、音声処理サービスを独自のＷｅｂアプリで利用できるほか、他のＷｅｂアプリから音声処理機能のみを利用できる音声ＷｅｂＡＰＩを提供している。現状、多国語音声合成（日本語、英語、中国語、韓国語、ポルトガル語）と翻訳の機能を提供している。今後、信号分析、音響診断、雑音除去など、処理メニューを増やしていく。

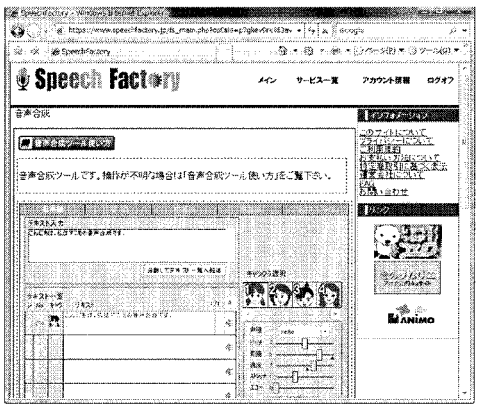

図５　SpeechFactory（音声合成）の画面

２）音声ＳＮＳ

最近、ＳＮＳ（Ｓｏｃｉａｌ　Ｎｅｔｗｏｒｋｉｎｇ　Ｓｅｒｖｉｃｅ）がインターネット上の新しいコミュニケーション手段として注目を集めている。当社では、音・音声が使えるＳＮＳ「コエノワ」をサービス開始した。従来のＳＮＳでは、テキスト、静止画、動画が利用できるが、コエノワでは、音声の貼り付けが可能となっている。コエノワは、音・音声を使った日記やコミュニティでどのようなコミュニケーションが成り立つか、どのような音声処理を提供することが有効かを研究する実験サイトである。

コエノワでは、マイクから録音した音声を記事に貼り付けられるほか、テキストから音声合成した音声も貼り付けられる。音声合成の声質は１０種類のキャラクタを用意している。また貼り付けた音声に対して信号処理を施すことも可能となっている。音声合成や信号処理は、ＳＮＳからＳｐｅｅｃｈＦａｃｔｏｒｙのＷｅｂＡＰＩを呼び出すことで実現している。

図６　音声 SNS「コエノワ」

４．顧客の新しい課題と価値観の把握

４．１　コールセンター向け通話録音

当社では、コールセンター向けの通話録音としてＶｏｉｃｅＴｒａｃｋｉｎｇを提供している。

コールセンターは、企業とお客様の接点として、大変重要な役割を担っている。コールセンターのオペレータの対話はその企業の印象そのものであり、お客様から寄せられる情報は、その企業にとって非常に重要なものである。

コールセンターの品質を向上させるためには、まずオペレータの対話スキルの可視化・数値化が重要であり、そのためには通話の録音は必須である。

また、お客様から寄せられる情報は、オペレータが聴く場合と、その製品の開発担当者が聴く場合で、まったく意味が違ってくる場合もありえる。お客様の生の声を担当部署に聴かせる意味は大きく、そのためにも通話録音は重要である。

図７　通話録音の利用シーン

４．２　コールセンター向け音声認証

コールセンターでお客様の本人確認を行うためには、知識の確認による方法がよく行われる。お客様番号と暗証番号を確認する方法が一般的である。実際この方法では、事故が多いという。そこで、お母様の旧姓、お子さんの通っている小学校の名称など、お客様しか知らないであろう個人情報を確認する方法を取ることになる。ただ、この方法だと自分は本人なのに、疑われているのではないかという印象があり、コールセンターのサービスの印象として良くない。そこで、あらかじめ登録していただいたパスフレーズ、お客様番号やお名前の発声などを使って音声認証を行うことが行われている。音声認証で、本人との一致度が高い場合に、従来の個人情報の確認のプロセスを簡略化することで、サービスの印象を向上させることができ、また利便性も向上する。またセキュリティのレベルを確保することが可能である。さらに、セキュリティ確保のために乱数表を使うこともあるが、高齢者の方や視覚に障害がある方に乱数表などを使っていただくのはハードルが高い。音声認証で行うことは、ＵＤの観点でも優れているといえる。

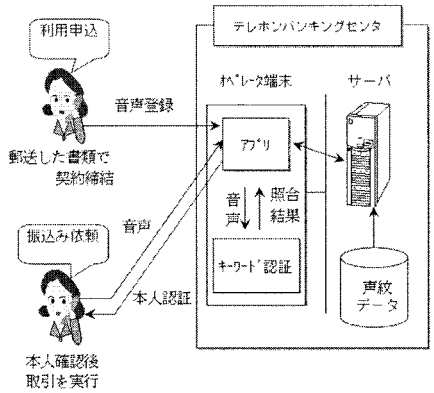

図８　テレホンバンキングでの音声認証の利用

５．さいごに

本報告では、当社で推進している音声の社会化の中での実際の取り組みについて紹介した。今後も、最新の音声技術、音声本来の特徴、最新のＩＴ技術、お客様の直面する問題を正しく理解し、お客様に最新の音声技術を適切なソリューションとして、いち早く提供していきたい。