

これからの音声 ～技術の実用的利活用～

庄境誠¹

ハンズビジー／アイズビジー状況下での音声 HMI, 大規模情報の音声検索などのこれまでの音声技術の実用化を振り返り, これからの音声技術の実用的利活用に関して私見を述べる. 持続可能で健全な音声経済の実現を目指すために, 日常におけるアンコンシャス型音声に音声技術を適用することを提案する. アンビエント情報社会での音声ライフログの実用化における具体的な技術課題として, 実環境耐性, プライバシー保護, 要約などを示唆する. また, 住まいにおけるセンサとしてのマイクの設置場所について考察する. 最後に, 住まいに設置されたマイクで収集された大規模音声ストリームの3階層情報処理フレームワークについて論じる.

Speech for Tomorrow - Practical Use and Applications of the Technologies -

Makoto Shozakai¹

A couple of existing examples of applying speech technologies such as speech HMI under hands-busy/eyes-busy situation and large information retrieval by speech is reviewed. Several personal comments in terms of practical use and applications of speech for tomorrow are discussed. In order to realize sustainable and healthy speech economy, applying the speech technologies to unconscious-type speech in daily life is proposed. Robustness in real field, privacy protection, summarization and so forth as concrete technical issues for practical application of speech life log in an ambient information society are suggested. Furthermore, positions of places where a microphone as a sensor is embedded in a house are considered. Finally, three layer information processing framework for large volume speech stream collected by a microphone installed in a house is discussed.

1. はじめに

一昨年の第5回 SLP デベロッパーズフォーラムでは, ハイテクが普及する上でどうしても避けられない落とし穴として, マーケティングの世界でよく知られている「キャズム(Chasm)」¹⁾ について紹介し, 音声 HMI 技術も他のハイテクと同様にこの「キャズム」を免れることが出来ず, それを乗り越えてさらなる実用化・普及の局面に到達するための課題とその解決策について, 事業面及び技術面から議論した²⁾.

昨年の第6回 SLP デベロッパーズフォーラムでは, その議論を継承し, 音声 HMI 技術・製品の品質の計測, 評価, 保証が「キャズム」を乗り越えるためのキーポイントであることを指摘し, その技術確立の必要性, 課題と解決方法などを論じた³⁾.

ここ10年での音声 HMI の技術的な進歩は目覚ましい. その実用化・普及の代表的なケースは, 自動車環境であろう. なぜ, 自動車環境において, 音声の利用が進んでいるのかというと, 運転行動中はハンズビジー／アイズビジーであり, 他の HMI 手段に乏しいからである. 昨年の4月には, 車の運転中, 携帯電話や音楽再生機器などの操作を音声 HMI を通じて行っている運転手は, 手で操作する運転手に比べて道路から目を離すわき見が極端に少なく, Mind Distraction を大幅に回避できるため, 近い将来すべての自動車メーカーが音声 HMI を車に搭載するようになるという調査結果が発表された⁴⁾. 但し, 行き過ぎた音声 HMI は, 逆に運転行動中の Mind Distraction のリスクを増大させる. そこで, その点に関する国際標準化の活動が進行している⁵⁾. ハイテクとしての音声 HMI は自動車環境において必須機能となり, 既にキャズムを越えたと言える. また, 自動車産業はそもそも, 自動車の品質の計測, 評価, 保証を究極なまでに追求し, それによりブランドを確立し, 現在の産業繁栄や隆盛を誇っており, 音声 HMI 技術・製品に関しても, 品質の計測, 評価, 保証の面に意が払われ, 意欲的な取り組みが進められてきた. 本年の第7回 SLP デベロッパーズフォーラムでも, 北米自動車市場で有名な Ford Sync システムにおけるチャレンジと成功要因が紹介されており, 大変興味深い⁶⁾.

一方, Google Voice Search に代表される, 音声を入力手段とするクラウド型情報検索無料サービスも着実に拡がり, 利用が進んでいる. スマートフォンには, クラウド音声認識サービスにアクセス可能なアプリケーションが標準搭載され, 広く利用されている現状を考えると, キャズムを越えたと言えるのではなかろうか. しかし, ベストエフォートを基本的なコンセプトとする IT 産業において, 品質の計測, 評価, 保証の意識は自動車産業に比べると薄い. 例えば, Google Voice Search の性能に関する保証はコミットされていない. 従って, 本当の意味で, クラウド型情報検索無料サービスがキャズムを越えた結論付けるのは早計かもしれない. 本 SLP デベロッパーズフ

[†] 旭化成株式会社 新事業本部 情報技術研究所
Information Technology Laboratory, New Business Development, Asahi Kasei Corporation

オーラムでも、音声検索実用化の現状と課題が議論されている⁷⁾。

自動車環境の音声 HMI や、Google Voice Search などのクラウド型情報検索サービスは、それぞれの目的に応じてユーザが音声認識システムに対して意図的に発声した音声を認識処理する機能である。この音声をコンシャス型音声と呼ぶことにする。SF 映画や TV ドラマのエンターテインメントの影響が大きいのかかもしれないが、長年の音声技術の研究開発により、コンシャス型音声の利用シーンに関する音声技術の利活用は進んだ。しかし、一日の中で、コンシャス型音声を発声する頻度は多くない。

そういう観点も無縁ではないだろうが、自動車環境の音声 HMI 製品は、(新規製品の受託開発費用を別にすると、) 安価なライセンスフィーで売買されており、例え、世界自動車販売台数(参考: 2010 年度 7200 万台)の全てに音声 HMI 製品が搭載されたとしても、その世界市場規模は高々数百億円でしかない。また、Google Voice Search は、検索連動型広告ビジネスモデルという全く異質の経済モデルに支えられ、無料で提供されている。

音声技術そのものの経済を音声経済と呼ぶとすると、コンシャス型音声に係る音声経済はこれまでの音声技術の研究開発投資を回収するにはあまりにも小さい。持続可能で健全な音声経済を形成するためには、さらに何桁も大きな産業規模、市場規模を追求せざるを得ない。

そもそも人間は、程度の差こそあれ、他人とコミュニケーションをするために、音声という手段を獲得した生き物である。人間は、家庭生活、会社生活、余暇生活、医療生活、介護生活などの中で、自然に音声を発する。音声技術が登場する前から、(人びとは音声技術や音声機能を意識することなく、) 日常的に喋る。これをアンコンシャス型音声と呼ぶことにする。また、日常生活には、人間の音声ばかりではなく、雑音や音楽などの音響信号も多く存在する。人間の音声を収集するために設置したマイクは、いやおうなく雑音や音楽など音響信号も収集するので、雑音、音楽などもアンコンシャス型音声に含めるべきかもしれない。(音声は、「おんせい」ではなく、「おと・こえ」と読むべきかもしれない。) 実世界に存在する「おと・こえ」のデータを用いた分類・検索技術として、音声と非音声の判別、音声ログデータからの音声検索、音響イベントのセンシングなどの音声・音響処理の新しい方向性については、既に論じられている⁸⁾。

振り返ってみれば、アンコンシャス型音声を対象とする音声技術の利活用はまだまだ進んでいないことに気がつく。これからは、アンコンシャス型音声に対して、これまで長年に渡って蓄積してきた音声技術を実用的に利活用して、さらに大きな規模の音声経済を創出するという観点が重要である。そこで、本 SLP デベロッパーズフォーラムでは、アンコンシャス型音声を対象として、「これからの音声」のあり方とそのため技術の実用的な利活用に関する私見をいくつか述べさせていただきます。

2. アンコンシャス型音声の利用シーン

昔から、人びとは家族、友人、会社の仲間、顧客などと会話をするアンコンシャス型音声を発する時間が圧倒的に長い。そういう意味で、音声は社会環境であり、社会システムとも言える。従って、住まい、くらし、医療・医療、ヘルスケア、業務の現場で利活用可能な音声技術も、社会環境であり、社会システムであらねばならない。

2.1 これまで

アンコンシャス型音声を対象とした音声技術の代表的な実用化例は、書き起こしの分野である。例えば、以下の試みがなされている。

- 1) 証跡: コールセンターの通話音声や対面販売音声の文字化、検索の容易化⁹⁾
- 2) 会議録: 衆議院や地方議会の新会議録作成システム¹⁰⁾¹¹⁾
- 3) 字幕: TV 番組音声への付与¹²⁾
- 4) 文字化: 音声認識技術と Wiki によるタグ付け¹³⁾、音声ドキュメント検索¹⁴⁾、講演音声の書き起こし¹⁵⁾

筆者も自分の講演音声の書き起こしを自らやってみたことがあるが、会話調の音声は、文法的なエラー、言い淀み、語尾の脱落、仮名のあいまい発声がありにも多く、何度も音声を再生させなければ書き起こしができなかった。会議録の自動作成に対する需要は幅広く存在すると思うが、使い勝手の良いシステムを実現するためには、まだまだ解決しなければならない技術課題が多い。

その他のユニークな実用化として、ハンズビジーな業務での音声 HMI の活用がある。物流現場でのピッキング作業の効率化や作業ミスの低減を目的として、インタラクティブな音声 HMI システムを提供し、業績を伸ばしている会社がある¹⁶⁾。ハンズビジーな物流、工場、医療、介護などの業務現場での音声 HMI の活用は、今後有望と思われる。これらのハンズビジー現場に音声 HMI を導入する場合、最初はコンシャス型音声を強いることになるが、その利便性が現場で受け入れられれば、いずれアンコンシャス型音声に転換する。そうなれば、音声技術はまさに社会環境であり、社会システムに進化したといえるだろう。

また、自動音声翻訳は、アンコンシャス型音声に対する音声技術の究極の姿といえる¹⁷⁾。

2.2 これから

一方、クラウドコンピューティング、センサーネットワークなどの ICT は、マイクロデバイスなどのエレクトロニクス技術と組み合わせられ、情報エレクトロニクス基盤

として、人びとの「いのち」と「くらし」を常時見守るアンビエント情報社会を形成しようとしている¹⁸⁾。これは、ネットワークロボットという概念とも相通じる¹⁹⁾。ネットワークロボットの中で、ビジブル型ロボットは、ヒューマノイド、ペット、ぬいぐるみなど姿形のあるロボットで、人間との親しみやすいコミュニケーションやインタラクションを目指している。バーチャル型ロボットは、ネットワーク上のサイバー空間やバーチャル空間の中で活動し、携帯電話、パソコン、液晶 TV のディスプレイ上のエージェントやアバターなどが人間とコミュニケーションする。アンコンシャス型ロボットは、住宅や建物などに埋め込まれたカメラ、赤外線センサなどの環境センサ、衣服や装身具に埋め込まれたウェアラブルセンサとこれらを制御する CPU が一体化したロボット機能がさりげなく人びとを見守り、「いのち」と「くらし」を支援する。最近では、ライフセンシングやライフモニタリングという言葉も使われる。

これらの世界においては、センサと情報科学を融合した「スマートセンシング」が大きな技術・市場分野に成長すると目される。その中で、センサとしてのマイクやセンサ信号としての音声の有意義な活用として、音声ライフログの分野に目が向けられてもよいのではなかろうか。例えば、以下のような利用シーンが考えられる。

[1] 防犯

窓ガラスの割れる音、不審者に吠える犬の鳴き声、留守中や深夜に聞こえる足音・話し声を検出すると、赤色灯を点灯したり、警備会社に通報したり、音・映像ログとして記録するということが考えられる。

[2] 見守り

浴室やトイレでの転倒を音で検知して家族に通知したり、乳児の泣き声を検知して保護者へ通知したり、睡眠時のいびき／無呼吸を検知して本人に注意喚起したり、本人が後で振り返りを行えるように保存するということが考えられる。また、音声は、健康・ヘルスケア情報でもある。小児喘息音、風邪による咳の頻度、花粉症発症などの体調のモニタリングにも音声は有用な情報源である。

また、今後は家族を見守るという従来の考え方から、家族に迷惑をかけないために自分で自分を見守るという考え方にシフトして、自分の音声を含めた生活音を音声ライフログとして記録し、必要な時に検索するという新しい文化が定着する可能性もある。

[3] 医療・介護

これからのチーム医療・チーム介護の現場で、医療コミュニケーション・介護コミュニケーションの証跡として音声ライフログを残すということは大変重要になる。

[4] 生涯電子日記

ビデオカメラで音・映像を連続的に撮り続け、気に入ったところのみを保存する利用シーンが考えられる。例えば、赤ちゃんが初めての「まんま」としゃべった時の声や初めてつかまり立ちやよちよち歩きをした時のお母さんの喜びの声を日記として記録したり、家族の団欒をアルバムとして残したりするという新しいライフスタイルがより普及する。また、家族同様のペットの面白動画を蓄積したり、動画サイトに披露したりするペット愛好家が増えるだろう。そのような利用シーンで音声技術が活かされる場面が登場するだろう。

[5] いつでも検索

センサとしてのマイクが常時住まいの音環境をモニタリングし、(通常はマイク入力音から不要情報を破棄しながら)情報検索のキーワードやコマンドをアンコンシャス型音声からアンビエントに検知し、クラウド型音声検索サービスに音声を送信して、Web 検索等で情報を表示したり、音声合成で読み上げたりして、検索情報を提示したりするかもしれない。

2.3 マイク位置

ところで、これらの利用シーンにおいて、音声ライフログの利用が普及する場合、住まいの中で、センサとしてのマイクはどこに設置されるのだろうか？

タイプ A : 天井, 照明, モニターカメラ

部屋のレイアウトに影響されず、一度設置されると安定して部屋の中の音声を収集できるメリットがある。天井にマトリックス的にマイクを設置すれば、人間の移動に追従可能なマイクシステムを構築することもできる。天井付けの照明やモニターカメラにマイクを設置すると、電源供給が簡便であるというメリットがある。一方で、部屋の中で人間は様々な姿勢を取る。一般住宅の 2.4m 高の天井の場合、立っている姿勢の場合は、口とマイクとの距離は、60-90cm くらいであるが、椅子やソファに座った姿勢ではその距離が 120cm くらい、床に寝た姿勢ではその距離が 220cm にもなる。従って、姿勢によっては、必ずしも好適なマイク設置場所とはいえない。

タイプ B : 壁, 家具

壁も天井と同様に、マイクをマトリックス的に収納しやすい有力な設置場所である。ただし、部屋のレイアウトによっては、家具などでマイクが隠れることが多い。それにより、収音ゲインが下がり、SNR が下がるデメリットも考えられる。一方、ベッドにマイクが設置される場合は、比較的顔に近い位置にマイクを設置することが可能

であろう。

タイプ C: 浴室, トイレ

浴室やトイレは、住まいの中でもプライベートな空間である。そのため、事故が起こった時に手遅れになることも多い。過去に、浴室内の音声認識の研究開発を行ったことがあるが、残響による誤認識の低減に相当な工夫が必要であった²⁰⁾²¹⁾。トイレは、浴室に比べて残響の影響は少ないであろうが、比較的狭い空間であり、同様な困難を伴うであろう。

タイプ D: 家電, ロボット

今後、Skype TV の普及に伴い、マイクが TV に標準装着されることになる。また、エアコンの新製品には、マイクが装着され、テレビや掃除機がけなどの生活音を検知して、生活シーンに合わせてきめ細かくエアコンを制御することで省エネ性能や快適性に配慮した製品が登場している²²⁾。また、最近、卓上型ボイスコントローラにマイクを装着し、スタートボタンを押すか、拍手やノックなどのハンドクラップ（3回手をたたく等）で音声認識の起動を行ない、音声コマンドでエアコンを制御する商品も発表された²³⁾。リモコンや卓上型ボイスコントローラは、比較的口元に近い状態で使用することが多く、高めの SNR が確保し易いため好適といえる。一方、TV にマイクを装着した場合は、遠隔マイクの問題が不可避である。ハンズビジーな台所での炊事を考えると、冷蔵庫や電子レンジにマイクを装着することもありそうである。

さらには、マイクがついたペットロボットも実用化されている²⁴⁾²⁵⁾。

タイプ E: 携帯電話, スマートフォン

携帯電話やスマートフォンには、既にマイクがついており、遠隔マイクによる低 SNR や部屋の残響の課題からも開放され、利用しやすいデバイスである。しかし、在宅時に常に携帯電話やスマートフォンを身につけるわけではなく、在宅時には充電していることも多い。バッテリー残量の問題から常時携帯は困難であり、音声ライフログ利用に向いているとは必ずしもいえない。

以上のようなマイクの位置を考えた場合、今後、以下の技術課題を解決する必要がある。

- ・残響や遠隔マイクによる低 SNR, 高歪みに頑健な音声認識技術（または残響や非定常な背景雑音の除去技術）の登場が待たれる。
- ・同時に、様々な設置箇所のマイクの入力音声の品質を均質化する技術も必要になりそうである。
- ・どこかに設置されたマイクひとつで音声ライフログをカバーすることはできそう

もない。天井や壁などの環境に埋設されたマイク、家電やリモコンに装着されたマイク、携帯電話やスマートフォンのマイクで取得された音声を連携させ、最終的な音声ライフログとして記録し、活用する技術も求められる。

2.4 音声ライフログのプライバシー

ところで、住まいにマイクを設置するというと、プライバシー問題の懸念の声が多い。筆者くらいの年齢の人間が集まると、遠隔に残した親のことが心配なので、ICT で親を見守るシステムが合理的な費用で可能なら利用したいという話がよく出る。その一方で、それは見守る側の論理であって、見守られる側の親はたとえわが子であっても一方的に見守られるのはプライバシーの侵害であり嫌だという意見もある。見守りは監視、盗聴と同じであるという人もいる。

そのようなネガティブな面と音ライフログの利用によるポジティブな面を天秤にかけた上で、利用するかどうかは個人の判断に委ねられているが、いずれにしても、今後は「自分で自分を見守る」という概念が基本となるように思う。自分の音声は自分の持ち物であり、自分で管理するとか、自分の音声ライフログを他人に発信、公開するのは自分の判断で行うという考え方である。この考え方は、既に、ブログや SNS の世界では常識となっている。

2.5 音声ストリーム新技術

しかし、毎日の音声ストリームは膨大であり、とても自分で管理しきれない。そこに、情報科学技術としての音声技術が必要になる。今後は、以下のような音声ストリーム新技術の高度化が新たに求められるのではないだろうか？

1) 音声の持ち主を特定する技術

住まい環境に設置されたマイクから收音された音声の持ち主を特定する技術が必要になる。生活者の人数が少数の住まいにおいては、家族の誰がしゃべったか、家族以外の人がしゃべったかを特定することは比較的容易である。音声は本来パーソナルな情報源であり、音声の持ち主が特定できれば、その持ち主に特化した音声技術を利用すればよい。不特定多数の音声を対象とする不特定話者音声技術は不要である。不特定話者大語彙連続音声認識を実現するために、膨大な計算資源を利用するクラウド型音声認識の潮流が顕在化しているが、個人にとって、不特定話者音声認識システムは冗長である。音声の持ち主を特定できれば、その持ち主にフィットした特定話者音声認識システムを駆動すればよい。

2) 音声ストリームのプライバシーを保護する技術

音声変換技術を利用して、音声ストリームに含まれる個人の音声を他の音声に変換してプライバシーを保護する研究が進められている²⁶⁾。

3) 他人の音声をマスクする技術

住まい環境に設置されたマイクが出力する音声ストリームには、自分の音声だけでなく、他人の音声も収録されている場合がある。自分の音声を公開する前に、他人の音声をマスクしたり、削除したりする技術の需要もあるかもしれない²⁷⁾。

4) 音声ストリームを要約する技術

潜在的で未だ深く認識されていない技術的課題は、住まい環境に埋め込まれたマイクが常時出力する大量の音声ストリーム出力をどのように要約するかという情報処理技術である。膨大な音声ストリームは、そのままでは扱いきれないので、何らかのスイッチレスな要約技術が必要である。背景雑音があまりない場合、一人の人間のみが喋っている場合、背景で音楽が流れていない場合、発音が明瞭な場合、音声の発声位置がマイクに近い場合は、現行の音声認識技術でもかなり文字化が可能である。一方、それらの逆の場合は、文字化の精度は大きく低下する。そのため、音声ストリームの要約には、いくつかのレベルが考えられる。

レベル1：一切文字化せず、音声区間検出だけをする。

文字化するのは諦めて、音声が発生していると思われる区間のみを検出し、その区間の音声を音声ライフログとして保存する。

レベル2：文字化できるところだけを文字化する。

現行の音声認識技術で文字化できるところのみを文字化する。しかし、文字化が可能な場合を正確に判定する技術は依然として確立されていない。

レベル3：全てを文字化する。

音声認識技術を用いて、音声ストリームの全区間を文字化する。しかし、多くの場合、正確な文字化は難しい。

5) マルチセンシング技術

他のセンサとの併用によるマルチセンシング技術も音声ライフログには有望であるが、紙面の都合からさらなる議論は省略する。

3. 3階層情報処理フレームワーク

上記のように、住まい環境の音声ストリームをハンドリングして、音声ライフログとして扱えるようにするためには、まだ未確立な技術要素が多い。以下では、音声ストリームを効率的にハンドリングする情報処理フレームワークについて論じる。音声ストリームを情報処理する構成要素としては、センサ（マイク）、ローカル端末、クラウドサーバの3階層が考えられる。

センサ（マイク）：収集した音声ストリームに含まれる雑音を除去し、音声アクティビティを検出し、情報価値を粗く判断して、粗い要約をする。過去には、天井付けのマイクで音声検出を試みた研究がある²⁸⁾。後述のクラウドサーバでの音声認識処理を支援する技術として、複数マイクを用いた、背景雑音除去技術や有意な音声アクティビティを検出する技術は今後大変重要となるであろう。

ローカル端末：センサ（マイク）で情報価値があると粗く判断された音声ストリームに対して、モデルを用いて情報価値をより正確に判断し、より精度の高い要約を行い、必要に応じてクラウドサーバに送信する。センサ（マイク）が複数存在する場合には、その統合処理もローカル端末が行う。

クラウドサーバ：情報価値の高い音声ストリームを蓄積し、クラスタリングし、外れデータを除去し、モデリングし、評価し、活動ログを付与する。ローカル端末にモデルを送信し、ローカル端末からの要求に応じて、統計情報を見える化し、音声ストリームの検索を可能とする。

その際、新たな研究課題としては、以下が挙げられる。

①ラベルなし音声ストリームからの統計モデル自動作成技術

音声は人間が聞けば内容をラベル化できるため、従来の音声認識技術においては、音声データの書き起こしを手で行って、ラベルを付与してきた。しかし、リアルタイムで大量に生成される音声ストリームにラベルを付与することは実際上困難である。ラベルがなければ、統計モデルを作成することも困難である。従って、音声ストリームに対し、雑音除去などの信号処理を施した後、音声アクティビティを抽出し、クラスタリングし、外れデータを除外し、クラスタ毎に統計モデルを作成し、パターン認識性能を評価する自動化技術が必要となる。

②情報価値付与技術

音声ストリームは常時生成されるため、それら全てをクラウドサーバに蓄積すると、

直ぐにサーバは溢れる。従って、全ての音声ストリームをクラウドサーバに蓄積することは事実上困難である。正常な住まい空間で生成される音声ストリームには情報価値がなく、異常な住まい空間で生成される音声ストリームには情報価値があると考えられるが、その判断基準があいまいである。

普段の安心・安全な状況では、異常は正常よりも圧倒的に発生頻度が低いので、異常をローカル端末で検出し、その音声ストリームを蓄積し、一方で、発生頻度の高い正常の音声ストリームを破棄すれば、クラウドサーバが溢れることは回避できる。また、時間変化とともに正常／異常空間は絶えず変化するため、正常／異常空間を常に学習し、正常と異常の判断基準を適応させることが必要となる。その判断基準に応じて、音声ストリームに情報価値を付与する技術は大変重要である。

③音声ストリーム要約技術

情報価値のある音声ストリームを全て生波形で蓄積するのは冗長である。情報価値に基づいてデータを圧縮したいが、その規範がない。過去の情報価値のある音声ストリームを要約し、検索し、あるいは音声ライフログとして再現する方法が不明である。過去と同じ音声ストリームデータは、過去と同一と記録し、過去と類似している音声ストリームは過去との差分を記録し、過去と異なる音声ストリームは生波形を記録するような、新たな音声ストリーム要約技術が確立される必要がある。ローカル端末での検索要求に応じて、過去の音声ストリームを効率的に検索できる3階層アーキテクチャの整備が必要であろう。

4. おわりに

アンコンシャス型音声を対象として、これまで長年に渡って蓄積してきた音声技術を実用的に利活用して、持続可能で健全な音声経済を創出するという観点で、「これからの音声」に期待する私見をいくつか述べさせていただいた。但し、書きなぐった感が否めず、議論も未熟で乱暴な部分が多いと自覚している。アンコンシャス型音声の音声経済に関する考察もまだできておらず、音声合成技術に関する議論もない。これを機会誤解に対するご指摘やご示唆をいただけたら、大変幸いに思う。

尚、本稿をまとめるに当たり、有意義な議論を重ねていただいた、旭化成株式会社新事業本部情報技術研究所の中川竜太氏、加藤智之氏に深く感謝する。

参考文献

1) Geoffrey A. Moore, "Crossing the Chasm," 1991, 1999, (日本語版) キャズム～ハイテクをブレイクさせる「超」マーケティング理論～, 川又政治訳, (株) 翔泳社

- 2) 庄境誠, "第5回 SLP デベロッパーズフォーラムイベント企画の趣旨について～コーディネータの立場から～," SLP78-7, 2009.
- 3) 庄境誠, "音声 HMI 技術・製品の品質の計測, 評価, 保証," SLP83-6, 2010.
- 4) http://www.usfl.com/Daily/News/10/04/0414_028.asp
- 5) 武田一哉, "車内情報インタフェースに関する国際標準化の動向," SLP88-8, 2011.
- 6) Bart Baeyens, 村上久幸, "Speech Recognition in the Car: Challenges and Success Factors – The Ford Sync Case," SLP88-7, 2011.
- 7) 大淵康成, 神田直之, "音声検索実用化の現状と課題," SLP88-5, 2011.
- 8) 大淵康成 "実世界に存在する音声・音響を対象とした認識技術," SLP78-11, 2009.
- 9) 寺西博人, "音声認識ソリューション開発について" SLP68-6, 2007.
- 10) 秋田祐哉, 三村正人, 河原達也, "会議録作成支援のための国会審議の音声認識システム," 電子情報通信学会論文誌, Vol. J93-D, No.9, pp.1736-1744, 2010.
- 11) 藤田泰彦, "音声認識実用化事例の紹介とその課題," SLP-078, No.6, 2010.
- 12) 今井亨, 奥貫裕, 小林彰夫, "音声認識によるリアルタイム字幕放送の進展," SLP88-4, 2011.
- 13) 後藤真孝, 緒方淳, 江渡浩一郎, "PodCastle: ユーザ貢献により性能が向上する音声情報検索システム," 人工知能学会論文誌, Vol.25, No.1, pp.104-113, 2010.
- 14) 中川聖一, Welly Naptali, 岩見圭祐, "音声認識・検索のための未知語の扱い," SLP87-6, 2011.
- 15) 篠崎隆宏, 久保田雄, ディクソン・ポール, 古井貞照, "識別学習モデルと教師なし CV 適応を用いた CSJ 講演音声認識," 音講論, 2010年3月, 2010.
- 16) <http://www.voccollect.jp/>
- 17) 西田明弘, "分散型音声認識, 翻訳サービスの現状と展望," SLP83-5, 2010.
- 18) 村田正幸, "アンビエント情報社会の実現に向けた取組み," 電子情報通信学会誌 Vol.93, No.3, pp.233-238, 2010.
- 19) 萩田紀裕, "ネットワークロボット概論," 電子情報通信学会誌, Vol.91, No.5, pp.346-352, 2008.
- 20) 馬場朗, 日比谷新平, 奈木野豪秀, 小笹詩織, 庄境誠, 竹原清隆, "浴室向け音声コントローラの開発," SLP73-4, 2008.
- 21) 奈木野豪秀, 小笹詩織, 馬場朗, 日比谷新平, 竹原清隆, 庄境誠, "浴室向け音声コントローラに用いる音響モデルの開発," SLP73-5, 2008.
- 22) <http://www.hitachi.co.jp/New/cnews/month/2010/10/1019.html>
- 23) <http://www.toshiba.co.jp/tha/about/press/110912.htm>
- 24) <http://www.nec.co.jp/products/robot/>
- 25) http://www.aist.go.jp/aist_j/topics/to2002/to20020315/to20020315.html
- 26) 中川聖一, 山本一公, 土屋雅稔, "音声に含まれるプライバシー情報の保護," 人工知能学会誌 Vol.24, No.2, pp. 190-195, 2009.
- 27) 厚田 和也, 浜中 雅俊, 李 昇姫, "安全見守りシステムにおけるプライバシー保護手法," SP2010-51, pp.37-40, 2010.
- 28) Obuchi, Y., et al. "International Voice Command Detection for Completely Hands Free Speech Interface", Proc, INTERSPEECH 2008 119-122, 2008.