

話速変換機能付きハンズフリーインタホン

Hands-Free Intercom with Speech Rate Converter

吉田 恵一* ・ 鷲 哲平* ・ 福島 実*

Kei-ichi Yoshida

Teppei Washi

Minoru Fukushima

早口の受話音声をやっくりとした口調にする話速変換技術において、従来の音声区間の信号を伸長する処理に加え、騒音区間の信号を圧縮する独自の音声／非音声判別制御アルゴリズムの開発により、音声／騒音レベル比（S/N比）0 dBの周囲騒音環境下での聞き取りやすさを実現した。

さらに、ハンズフリー通話処理において、音声スイッチの状態に応じて話速変換メモリーデータの消去を行う音声スイッチ連動型話速変換処理法の開発により、通話中の話速変換によるエコー、不自然な音の途切れ、ハウリングの発生を抑制し、自然な交互通話を可能にする話速変換機能を搭載したハンズフリーインタホンを実現した。

A hands-free intercom with speech rate conversion function for slowing rapid speech has been developed, which enables to hear the voice clearly under noisy circumstances with 0 dB voice/noise level ratio (S/N ratio). This has been achieved by the combination of the conventional process that lengthens the periods of voice signals and newly developed algorithm for identifying voice/noise signal to compresses noise signals.

In addition to it, this intercom features natural bidirectional communication by suppressing echoes, disconcerting sound loss and acoustic feedback caused by the speech rate conversion. This has been achieved by a voice switch-linked speech rate conversion method that erases data in speech rate conversion memory in response to the status of the voice switch.

1. ま え が き

国内の60歳以上人口が30%を超えるという高齢化社会の到来を向かえた現在、インタホン等の通話機器においては話者の音声をやり大きくかつよりわかりやすく聞きたいというニーズが顕在化している。また、宅配業者等とのインタホン越しの通話においては、業者は決まりきった定型文しか話さず非常に早口でわかりにくいいため、これをわかりやすくしたいという要望が多くある。

近年UD（ユニバーサルデザイン）対応機能として、話速変換機能を付加した通話端末が携帯電話や会議システム等にみられる¹⁾。これは、デジタル信号処理により話者の声を「ゆっくり」にして聞き取りやすくする技術である。

しかし、単純に時間領域で音声信号の伸長処理を行うと声質が変り、本人とは識別できず他人の声のようになってしまうという問題がある。そのため、ピッチと呼ばれる声帯の振動周期（声の高さ）情報をリアルタイムで抽出し、これを用いて伸長処理を行うことによって、声質を変えない話速変換を可能としている（図1）。

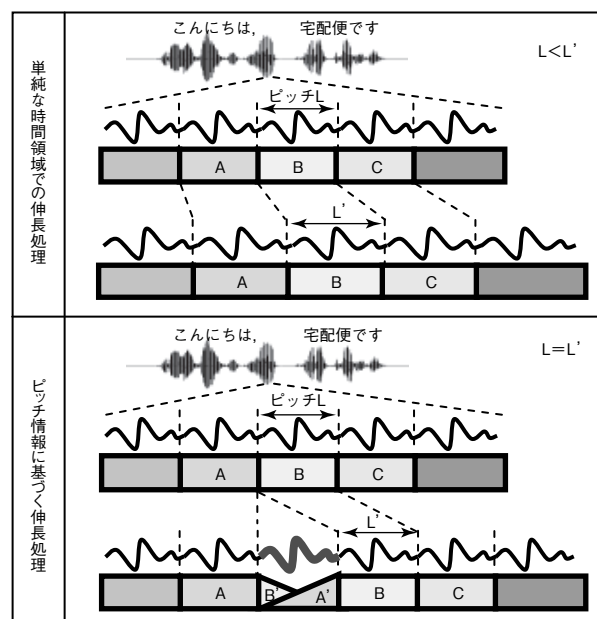


図1 話速変換処理の概要

* 情報機器事業本部 情報機器R & Dセンター Research & Development Center, Information Equipment & Wiring Products Manufacturing Business Unit

一方、当社が得意とするハンズフリーインタホンにこの技術を導入するに際しては、劣悪な周囲騒音環境下における使用の問題や、通話中の話速変換によるエコー、不自然な音途切れ、ハウリングの発生等が考えられる。

そこで筆者らは、独自の話速変換アルゴリズムと音声スイッチ連動型処理技術の開発により、自然な交互通話が可能なハンズフリーインタホンを実現したので報告する。

2. 開発課題

2.1 騒音環境下での話速変換処理

話速変換処理を行うと、遅延分の音声信号データがメモリーにだんだん蓄積されて最後はメモリー容量不足を起してしまう。これを防ぐため、一般的にはできるだけ会話中の無音区間を圧縮してメモリー容量不足が起きないように処理が行われる²⁾(図2)。

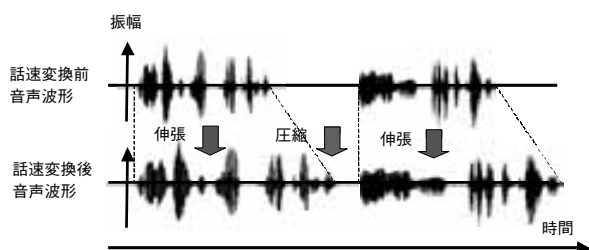


図2 話速変換伸長圧縮処理概要

インタホンの設置環境は、宅内端末はリビング、寝室、浴室などの各部屋、宅外端末は外壁面やサインポスト、集合住宅における共用玄関部など多様である。また、その周囲騒音も、宅内ではTVの視聴音、機器の報知音、換気扇の音などが、宅外では道路騒音、工事騒音、風切り音などが存在し、多様である。これらの騒音の音圧は設置環境や時間帯などによりさまざまであるが、最大で80 dBA以上となる場合もある。

このような環境下においては、会話中の無音区間圧縮が行われず、話速変換に使えるメモリー容量がすぐに不足してしまうため、話速変換処理ができなくなるという問題がある。たとえば、サンプリングレート8 kHz、メモリーサイズ8000ワード、話速変換速度1.5倍で話速変換処理を行った場合、2～3秒でこのような状態になる。したがって、周囲騒音の大きな環境における話速変換性能の確保が課題となる。

2.2 話速変換ハンズフリー通話処理

筆者らが考案した通話アルゴリズム^{3), 4)}と話速変換処理を組み合わせるハンズフリーインタホンに搭載した通話システム構成図を図3に示す。受話路に話速変換処理を挿入している。

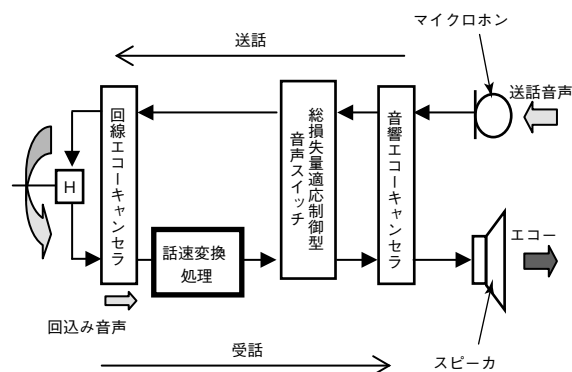


図3 話速変換処理通話システム構成図

図3の構成でマイクホンから音声を入力する場合を考えると、回線エコーキャンセラで抑制しきれない回込み信号に対して話速変換処理が行われるため、送話音声と回込み音声で時間差が発生し、送話と受話を切り替えるためのスイッチである音声スイッチが受話側に切り替わることで話者側にエコーが発生する。たとえば、話者が「こんにちは」と発話した場合に、エコーキャンセラで抑圧しきれなかった回込み音声の話速変換処理され、送話音声よりも時間的に長くなることで、音声スイッチが受話状態となった「ちは」の部分が話者へのエコーとなる(図4)。したがって、話速変換機能付きハンズフリー通話の実現のためには、このエコーの発生を防ぐことが必要となる。

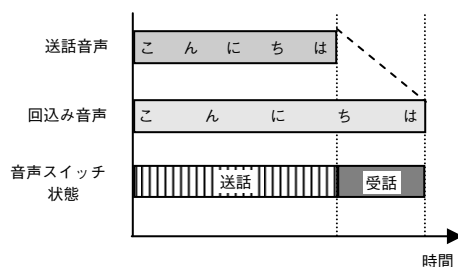


図4 話速変換によるエコー発生例

3. 話速変換アルゴリズム開発

3.1 騒音環境対応話速変換アルゴリズム

前述の課題を解決するため、図5に示す音声／非音声判別処理部、伸張圧縮判定処理部、騒音レベル判定処理部、伸長圧縮処理部およびメモリーで構成した話速変換アルゴリズムを提案する。

伸長圧縮処理部は、伸長圧縮判定処理部の出力である伸長圧縮判定フラグの状態、および騒音レベル判定処理部の出力である伸長圧縮停止フラグの状態に基づき、入力in(n)に対して伸長処理、圧縮処理、あるいはスルー処理(なにもしない)のいずれかを行い、その結果をメモリー経由で出力する。

以下に、主要各部の詳細について述べる。

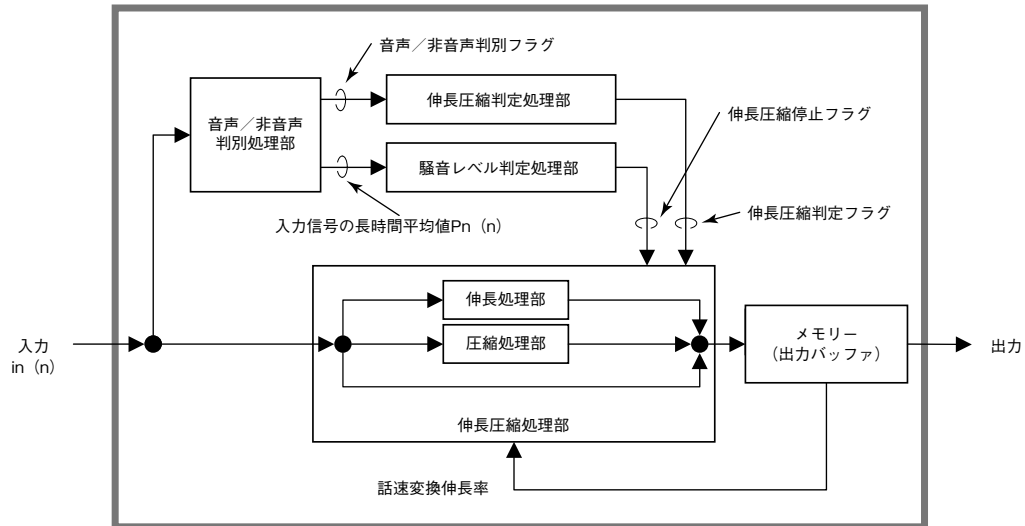


図5 話速変換アルゴリズム

3.1.1 音声/非音声判別処理

音声/非音声判別は、離散サンプル時刻 n における入力信号 $in(n)$ の短時間平均値 $Ps(n)$ と長時間平均値 $Pn(n)$ の比を用いて行う。

$Ps(n)$ は以下の式で求められる。

$$Ps(n) = \begin{cases} (1-a) \cdot Ps(n-1) + a \cdot |in(n)| & (1) \\ \quad \quad \quad (if \ \gamma \cdot |in(n)| < Ps(n-1)) \\ |in(n)| & (2) \end{cases} \quad (2)$$

ここで、 $a (< 1)$ は Ps に対する入力信号の重みを決める係数、 $\gamma (< 1)$ は急峻な信号の立上がりには追従するための係数である。

時刻 n における $Pn(n)$ は以下の式で求められる。

$$Pn(n) = (1-\beta) \cdot Pn(n-1) + \beta \cdot |in(n)| \quad (3)$$

$\beta (< 1)$ は $Ps(n)$ に対する入力信号の重みを決める係数であり、 $a > \beta$ とすることで、 $Ps(n)$ と比較して $Pn(n)$ の変動をなだらかにしている。

$Ps(n)$ は入力信号の変動に対してただちに追従することで、音声信号のパワーを表しており、 $Pn(n)$ は入力信号に対して緩やかに追従することで、定常的な雑音のパワーを表している。

3.1.2 伸長圧縮判定処理

まず、伸長圧縮判定処理から話速変換処理までの概要を図6に示す。音声/非音声判別処理の結果に基づき、音声区間では入力信号に対して伸長処理を、非音声区間の信号に対しては圧縮処理を行ったうえで話速変換処理を行うことでメモリーを有効活用する。

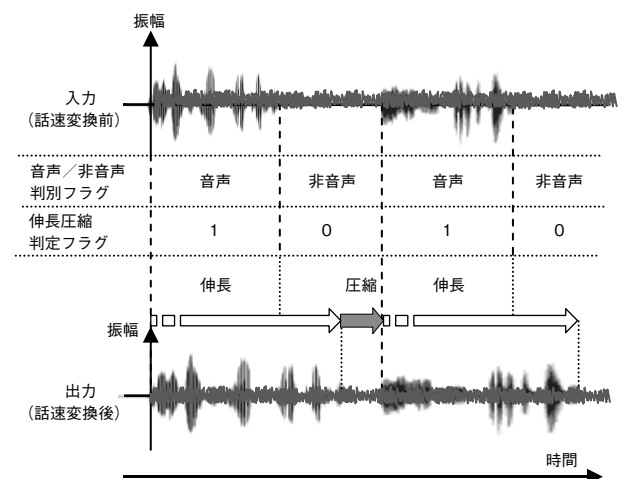


図6 伸長圧縮処理

伸長圧縮判定処理のフローチャートを図7に示す。入力信号に対して音声/非音声判別フラグが音声と示した場合には、図5の伸長圧縮処理部で伸長処理を行う。また、フラグが非音声と示した場合には、非音声区間が一定時間続いた後に圧縮処理を行う。

以上の処理を行うことで、通話中の短い非音声区間が圧縮されて出力音声が無音になることなく、また音声/非音声の一時的な誤判別により音声圧縮されて通話が途切れる問題も回避できる。

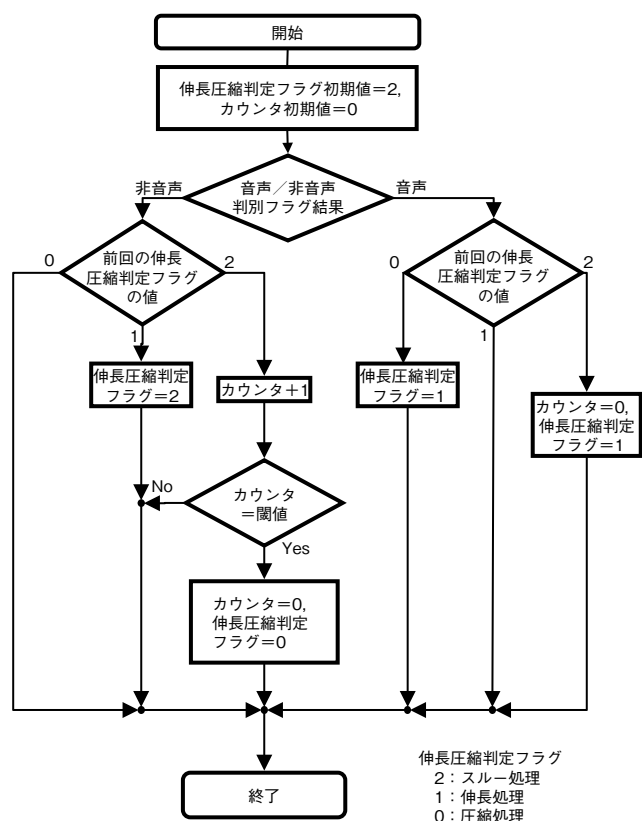


図7 伸長圧縮判定処理フローチャート

3.1.3 騒音レベル判定処理

騒音レベル判定処理の概要を図8に示す。この処理は、式(3)で定義した入力信号の長時間平均値 P_n から騒音レベルを3段階に分類し、それに応じて話速変換処理を行う。

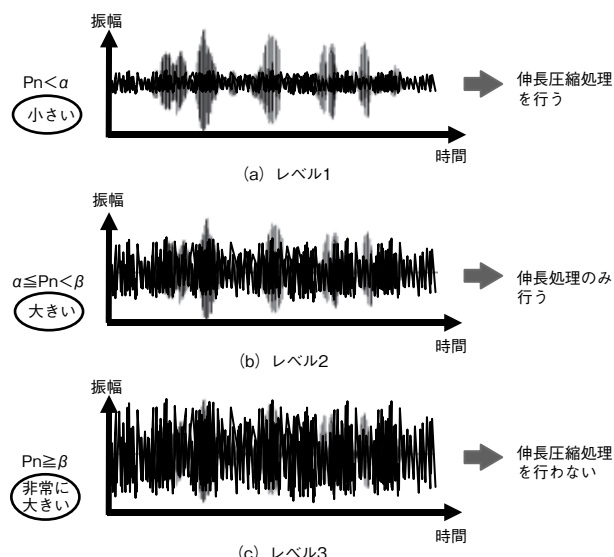


図8 騒音レベル判定処理の概要

まず、騒音レベルが小さい場合（レベル1）には音声区間を伸長し、非音声区間を圧縮する通常の話速変換処理を行う。次に、騒音レベルが大きい場合（レベル2）には音声区間の信号に対して伸長処理は行うが、非音声区間の圧

縮処理は行わない。これは、騒音が大きく音声と同等の音圧の場合に、音声区間が非音声区間と誤判別され、圧縮されることで音声途切れることを防ぐためである。そして、騒音レベルが非常に大きい場合（レベル3）には伸長処理、圧縮処理とも行わない。これは、騒音が音声よりもかなり大きい場合、音声のピッチ検出ができずに信号を伸長処理したときに音声不自然に変換されることを防ぐためである。

以上で述べた音声/非音声判別処理、伸長圧縮判定処理、騒音レベル判定処理により、音声と同等レベルの騒音が重畳している場合でも話速変換処理を行うことができ、また騒音が音声より大きい場合でも音声信号の圧縮による音声途切れ等がないアルゴリズムを実現している。

3.2 話速変換アルゴリズム処理ソフト評価

話速変換アルゴリズムをDSP (Digital Signal Processor) に実装し、評価を行う。その評価環境を図9に示す。

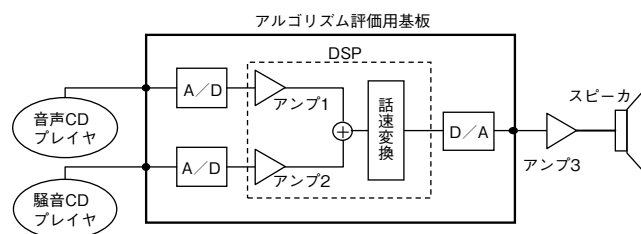


図9 話速変換アルゴリズムの評価環境

評価条件を以下に示す。

(1) 測定条件

音声信号：音声明瞭度評価用音源

(10 文章, 1 分間, 無音区間あり)

騒音信号：白色雑音 (帯域幅 20 kHz)

音声/騒音レベル比: +10 dB, 0 dB, -10 dB

話速変換メモリーサイズ: 4800 ~ 8000 ワード

伸長処理倍率: 1.0 ~ 1.5 倍

(2) 判定条件

(a) 音声/騒音レベル比が 0 dB 以上で音声低速に変換されること。

(b) すべての音声の内容が正しく聞き取れ、異質な音 (人工音) がしないこと。

評価結果を表1に示す。

表1 評価結果

音声/騒音レベル比	10 dB	0 dB	-10 dB
判定条件 (a)	○	○	—
判定条件 (b)	○	○	○

音声/騒音レベル比が 0 dB 以上の環境下ですべての音

声信号を低速に変換することを確認している。また音声／騒音比 -10 dB 以上の環境下でも誤動作せず、すべての音声信号に対して通話性が確保されることも確認している。

4. 話速変換ハンズフリー通話処理法

話速変換ハンズフリー通話時に発生するエコーの問題を解決する方式の検討を行う。

4.1 従来音声スイッチ連動型話速変換通話方式

従来、回込み音声の信号が話速変換されることで音声スイッチが受話状態となりエコーが発生することから、音声スイッチが送話状態のときには話速変換処理を停止する方式が考案されている。

この処理を追加すると、図 10 に示すように送話音声が入力されている間は音声スイッチが送話状態となるため、回込み音声の話速変換処理されなくエコーの発生を防止できる。

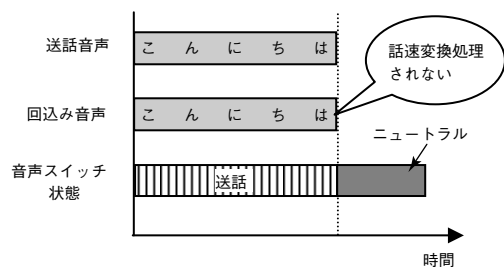


図 10 音声スイッチ連動によるエコー防止の例

しかし、この方式では、音声スイッチが送話状態のときに話速変換処理が停止した場合、図 11 に示すように、伸張処理後データがメモリー内に蓄積された状態で送話音声が入力されると、メモリー内に残っているデータが時間差をおいて受話音声としてスピーカから出力されるという現象が発生する。

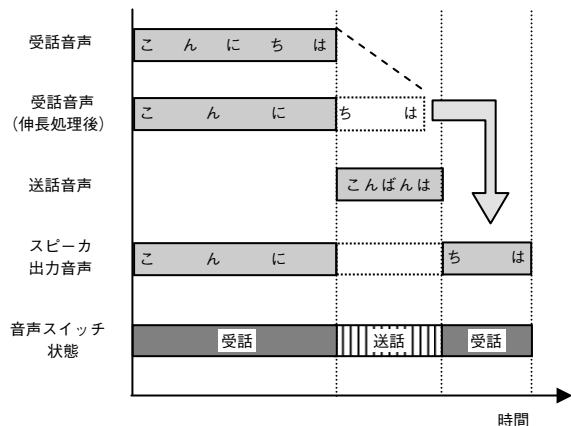


図 11 不自然な話速変換の例

4.2 メモリーデータ消去型音声スイッチ連動話速変換通話方式

そこで、送話時にエコーが発生せず、音声スイッチの切り替えによってメモリー内のデータが遅延して出力されることのないメモリーデータ消去型音声スイッチ連動話速変換通話方式を考案した。

その方式のブロック図を図 12 に示す。この図において、話速変換処理部内の音声伸長圧縮処理部とメモリーは音声スイッチにより制御され、音声伸長圧縮処理部は音声スイッチが送話状態のときに伸長処理を停止し、メモリーは音声スイッチが送話状態に切り替わるとバッファ内に蓄積されたデータを消去する。

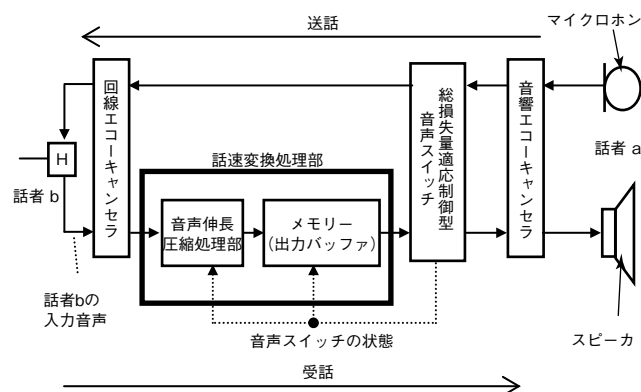


図 12 メモリーデータ消去型音声スイッチ連動話速変換通話方式のブロック図

この方式における、音声スイッチの状態に基づく処理を図 13 に示す。音声スイッチが受話状態のときには、話者 b からの入力音声は伸長処理され、データ B はデータ B および B' となってメモリー（出力バッファ）にセットされる（図 13 (a)）。音声スイッチが送話状態に切り替わると、伸長処理が停止してメモリー内のデータ B および B' は消去され、入力データはメモリー（出力バッファ）に蓄積さ

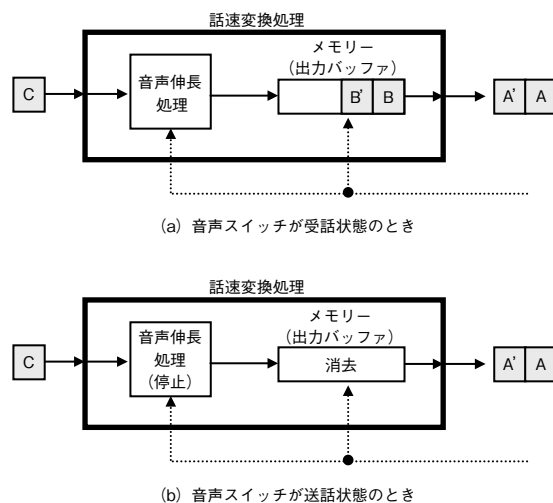


図 13 メモリーデータ消去型音声スイッチ連動話速変換通話方式の動作例

れずにそのまま出力される（図 13 (b)）。

以上の処理により、図 11 に示した状況が発生した場合においても、メモリー内のデータが消去されるため、時間を置いて音声出力されるような不自然なことが発生しない（図 14）。

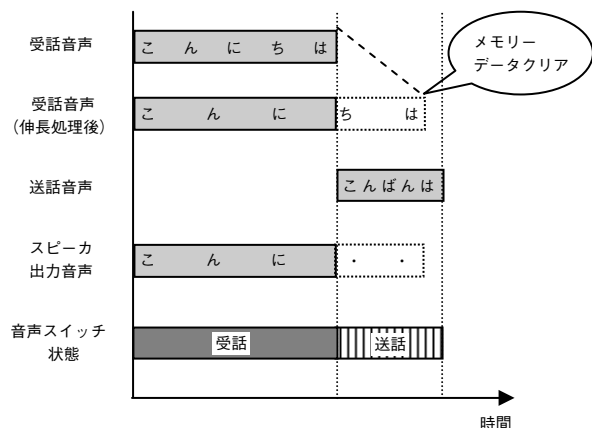


図 14 音声スイッチ連動によるエコー防止

4.3 実機における通話性能検証

開発した話速変換処理機能を図 15 に示す集合住宅系セキュリティインタホン「Smart MONION」および「MONION-R」に搭載している。

いずれも話速変換通話時においてハンズフリー通話機器として十分な音量と音質を有しており、不快なエコーや音の途切れ、ハウリングを生じることがなく、80 dBA 以上の周囲騒音環境下においても従来品と同等の良好な通話性能を示している。



図 15 話速変換通話機能を搭載した製品例

5. あ と が き

早口の受話音声をゆっくりとした口調にする話速変換技術において、従来の音声区間の信号を伸長する処理に加え、騒音区間の信号を圧縮する独自の音声／非音声判別制御アルゴリズムの開発により、音声／騒音レベル比（S/N 比）0 dB の周囲騒音環境下での聞き取りやすさを実現した。

さらに、ハンズフリー通話処理において、音声スイッチの状態に応じて話速変換メモリーデータの消去を行う音声スイッチ連動型話速変換処理法の開発により、通話中の話

速変換によるエコー、不自然な音の途切れ、ハウリングの発生を抑制し、自然な交互通話を可能にする話速変換機能を搭載したハンズフリーインタホンを実現した。

本開発の成果は、当社集合住宅用セキュリティインタホンに搭載され、通話速度を聞き取りやすく変換する「ゆったりトーク^{*1)}」機能として高齢者から子供まで誰もが快適に使える操作性を実現している。

●注

* 1) ゆったりトーク：パナソニック（株）の登録商標

*参考文献

- 1) 岩崎 直哉, 井上 健生, 大倉 計美: 高齢者による電話向け話速変換音声の評価実験, 映像情報メディア学会技術報告, Vol. 27, No. 37, p. 9-12 (2003)
- 2) 都木 徹, 武石 浩幸: 話速変換機能を搭載したラジオ, テレビの開発, 電子情報通信学会技術研究報告, Vol. 106, No. 20, p. 31-36 (2006)
- 3) 吉田 恵一, 福島 実, 竹山 博昭: 双方向通話型の浴室用ハンズフリーインタホン, 松下電工技報, Vol. 52, No. 4, p. 32-36 (2004)
- 4) 福島 実, 吉田 恵一: 双方向同時通話型ハンズフリーインタホン, 松下電工技報, Vol. 55, No. 1, p. 61-67 (2007)

◆執筆者紹介



吉田 恵一

情報機器 R & D センター



鷺 哲平

情報機器 R & D センター



福島 実

情報機器 R & D センター
電気通信主任技術者