

発音パラメータに基づく超低速音声通信に関する研究

A study on very low bit-rate speech communication scheme employing speech parameter transmission

研究代表者 群馬大学 工学部 情報工学科 助教授 嶋本 薫

Associate Professor, Dept. of Computer Science, Faculty of Engineering, Gunma University, Shigeru SHIMAMOTO

Lately, personal mobile communication system have grown widely, rapidly in the world. However, it is difficult to utilize broad band wireless channels due to lack of frequency resource. Therefore, the speech coding techniques which can transmit voice data efficiency with low bit-rate channel, is required in mobile communication fields now. In this study, I propose very low bit-rate speech transmission scheme. I had some experiments in search of the amount of transmission data and the quality of voice with alteration of speech-analysis parameters. I also study speech analysis parameters to achieve best speech quality.

1 研究目的

現在、自動車電話、携帯電話等、移動体通信が急激に発展してきており、移動体音声通信を支える音声符号化技術について、多くの企業および教育機関で盛んに研究されている。移動体通信を行なう場合、割り当てられる周波数帯域や電力、伝送誤り等の要因により高速回線を確保することは難しいため、移動体音声通信のための音声符号化技術は高効率で伝送データ量が少ないことが求められる。既に音声の転送に関して ADPCM、VSELP など様々な圧縮方式が開発されているが、音声波形に基づく波形符号化をベースにしたものが殆んどであり、いずれも音声信号を周波数軸でサンプリングし符号化する点では基本的には同一の手法であり、振幅データおよびその予測値、または周波数データおよびその予測値を伝送する従来の音声通信方式では、削減できる伝送データ量には限界があった。

本研究の目的は限られた伝送回線を用いて効率的に、かつ少ない伝送データ量で音声通信を行なう超低速音声通信方式を構築することにある。提案する音声通信方式は音声の送信側、受信側の両システム間において、同様の音声資料(以下音声テンプレート)を保持し、音声テンプレートから入力音声に近い音声を探し、その音素の位置情報を転送する。受信側では送受信システムで同様の音声テンプレートを保持するため、受けとった位置情報に基づき該当する部分を再生する。転送する情報として音声テンプレートの位置情報を用いることにより、伝送データ量の削減を実現させる。

提案方式は音声認識と違って、言葉として認識する必要がなく、音として音声を扱うため、従来とは異なるアプローチでの音声解釈が必要だが、逆に実用化を考えた場合、通常の音声認識よりも簡易に実現できる可能性がある。

また通信容量的に従来の10分の1~100分の1程度にデータ量が圧縮可能と考えられる。本研究では、音としての人間の音声認識、および音声合成を中心に、それに付随するパラメータ抽出法の確立、通信プロトコルの構築等を行なう。

2 研究経過

音声通信は人間の喋る音声を、人間が聞くという形態が基本であり、伝える情報量としては言葉として記述可能なものが殆んどで、また言葉でなくても偽声語により表現可能な音も数多く存在することから、一般の音声通信においては、音声を発音記号のような音に分解し、その記号を伝送することで可能となると考えられる。提案する超低速音声通信方式は、マイクロホンなどで観測される音圧情報を変調し伝送することで通信を行なうのではなく、音圧情報から得られる音声としての特徴を示すパラメータを抽出し伝送する。更に、送信側で発音記号のような記号に加え、人物、音量、音程を抽出、伝送し、受信側ではこれらパラメータに基づき音声合成することによりより忠実な音声通信を実現可能とする。

提案する超低速音声通信方式は、送信側と受信側のシステムにおいて同様の音声テンプレートを保持し、送信側では音声分析、受信側では音声合成に使用する。音声テンプレートは会話における大部分を示す母音と一般の会話音声を録音することにより作成し、更に、必要に応じて使用中に新たな音声の追加登録を行なって作成する。本方式のシステム概念図を図1に示す。まず、入力した音声波形から表1の諸元に基づく分析により音声の特徴量を求め、入力音声と音声テンプレートの特徴量を要素とするマッチングを行なうことにより、入力音声が最もマッチする音声テン

プレート部分を抽出する。マッチングに用いる距離尺度は、音声波形をフーリエ変換することにより得る周波数スペクトルをフィルタにより簡略化したベクトルにおけるベクトル距離を用いており、分析音声での総和を類似度とする。最終的な分析結果は入力音声に対してすべての音声テンプレートとの類似度を算出した後に最も類似度の高い(値の小さい)音声テンプレートを採用し、結果としてその音声テンプレートの開始時刻(以下 オフセット)を出力する。この際、音響的性質の異なる音声を採用され出力されてしまう問題を回避するため音声テンプレートの類似度に対してしきい値を設定する。類似度が小さい(値の大きい)場合は現在の音声テンプレートでは入力音声が発現できないと判断し、入力音声波形をそのまま伝送することにより出力音声の劣化を抑制しており、新たに音声テンプレートとして追加登録する。

送信側では、以上の方法で求めた分析データを図4に示す伝送フレームと呼ばれる伝送形式に基づき、データの伝送を行なう。一方、受信側では受信情報を基に音声データを音声テンプレートから取得し、音声合成を行ない、出力する。

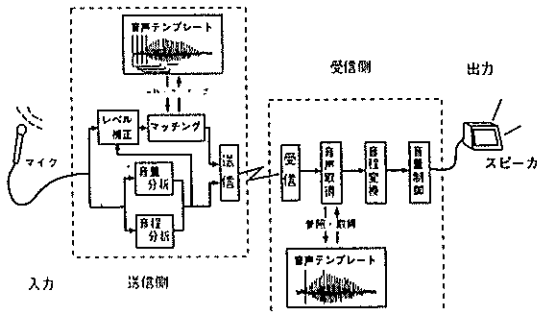


図1: 超低速音声通信方式の概念図

音声データ	
サンプリング周波数	22050 Hz
レゾリューション	16 bits
音声方式	リニア PCM
チャンネル	モノラル
音声分析諸元	
分析区間	50 ms ~ 200 ms
分析フレーム	約 46 ms
分析ステップ時間	1 ms ~ 20 ms
FFT精度	1024 値

表1: 分析諸元

次に、本音声通信方式で用いる音声分析方式の概要を示す。音声の分析には表1の諸元に基いて数値化した音声の特徴量を用いたマッチング手法を利用し、入力音声は分析フレームに分割され、分析フレームに対して高速フー

リエ変換(FFT)を行ない、更に複数の帯域フィルタを用いて数値化、特徴量化を行なう。また、時間経過に伴う音声の変化をとらえるため、分析フレームはオーバーラップするよう設定し、オーバーラップする時間パラメータを分析ステップ時間と呼ぶ。入力音声と音声テンプレートの特徴量に対して、分析フレームごとに特徴量のベクトル距離を求め、分析区間全体におけるベクトル距離の総和を入力音声と音声テンプレートの類似度とする。ある音声に対し、類似度を求めた様子を図2に示す。図2において、横軸は分析フレーム数を、縦軸は類似度を示す。図中に示したAでは入力音声と音声テンプレートの差が小さく、類似していることを示し、一方、Bでは入力音声と音声テンプレートの差が大きく、類似していないことを示している。

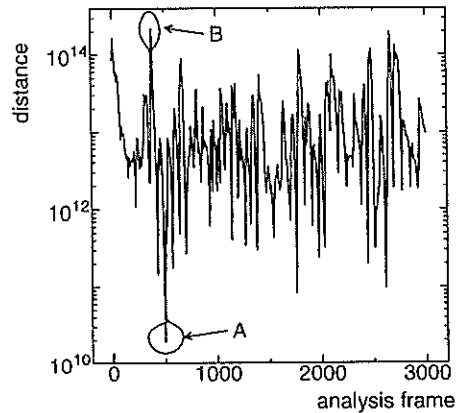


図2: 音声の類似度

分析区間において以上のようにして得られるデータを音声の特徴量として用い、マッチングする時の概念を図3に示す。図3上部は入力音声を、中部は音声テンプレートを、下部に分析区間を連続的に変化させた際の類似度の様子を示している。すべての分析位置に対して類似度を求め、音声テンプレート中で最も入力音声の特徴量と類似度の高い(値が小さい)分析開始時刻(オフセット)を分析結果として出力する。

また、受信側での合成音声の著しい劣化を考慮し、分析結果として採用する類似度に対し、しきい値を設定する。分析により得られた類似度がしきい値を越える場合は人力の音声波形をそのまま伝送し、同時に新規音声テンプレートとして追加登録することにより合成音声品質の著しい劣化の抑制を実現している。

本音声通信方式では、音声进行分析することにより得られた伝送データは、図4に示す伝送フレームと呼ぶデータ形式に基づいて送付を行なう。伝送フレームには制御フレームと情報フレームの2つの形式が定義され、データの伝送を行なう。

制御フレームは、回線接続、回線切断といった回線制御

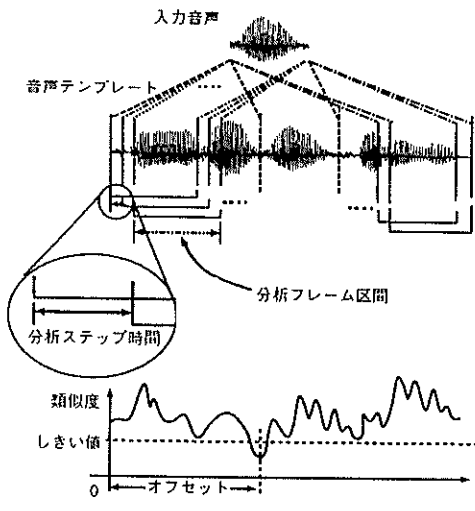


図3: マッチングの図

のための制御信号を送信するほか、送信側における話者の変更などにより、話者の声質の変化が生じた際にその情報の伝送を行ない、受信側において音声合成に使用する音声テンプレートの切替を指示する。音声テンプレートは性別、年齢に応じて多数の種類を用意し、受信側は声質の変化情報を受けると音声合成のための音声テンプレートの変更を行なう。一方情報フレームは、音声の分析、マッチングによって得た分析データの伝送を行なう。音声のマッチングにより入力音声に近い音声が現在の音声テンプレートから得られないと判断した場合についても、情報フレームを用い、入力音声波形をそのまま伝送する。

図5、5、5は伝送フレームを用いたシーケンスの例を示している。縦軸は時間経過(時間軸)を、局間に示された矢印はデータの伝送を示し、破線は制御フレームの伝送を、実線は情報フレームの伝送を示す。本音声通信方式のデータ伝送は双方向で行なうが、ここでは一方に注目して説明を行なう。

図5では、まず制御フレームによって回線接続が行なわれ、音声合成として既登録話者Aの音声テンプレートの使用を指示し、受信側は既登録話者Aの音声テンプレートを用意する。回線接続の確立を行なった後、送信側では入力音声の分析結果を情報フレームに基づき、順次伝送する。既登録話者Aの音声テンプレートを用意する。図5では、時刻tにおいて、送信側(X局)の話者が既登録話者Bに変化したため、送信側は制御フレームを用いて受信側に音声テンプレート変更を指示し、受信側は音声テンプレートの変更を行なう。

本音声通信方式を用いた時のデータの伝送に要する伝送回線速度は、新規の音声波形データの伝送が全く生じない場合には情報フレーム(全28ビット)を1秒間に約5~20回伝送するため、この時要する伝送回線速度は140bps(bits per second)~560bpsであり、また、1分間に1つ程度の新

規音声波形データの伝送が生じた場合でも、平均1k bps程度の伝送回線速度となる。

制御フレーム

フレーム識別子	制御コード	テンプレートコード
4 bits	4 bits	8 bits

情報フレーム

フレーム識別子	音響コード	音響コード	オフセット
4 bits	4 bits	4 bits	16 bits

または

フレーム識別子	波形データレイズ	音声波形データ
4 bits	4 bits	4 bits

4 bits 4 bits 伝送フレーム形式

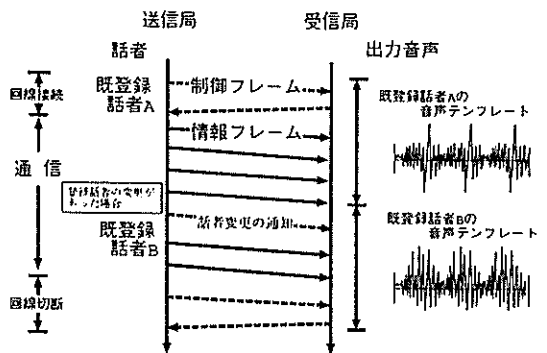


図5: シーケンス例

3 研究成果

本研究では、本音声通信方式を用いた際に要する伝送回線速度と再現される音声の品質との関係を明らかにするため、以下の環境で音声品質主観評価実験を行なった。

表1に示す諸元の音声を用い、分析のパラメータとして

分析区間	: 50ms, 100ms, 150ms, 200ms
分析ステップ時間	: 1ms, 5ms, 10ms, 15ms, 20ms
しきい値	: 低、中、高

を設定し、それぞれの合成音声波形と8000Hz、16000Hz、22050Hzの音声(計:62音声)をランダムな順序でDAT(Digital Audio Tape)に録音し、各々の条件での分析による合成音声波形に対して被験者10名の主観評価を行なった。評価法はオピニオン評価法を用い、音声に対して非常に良い、良い、まあ良い、悪い、非常に悪いの5段階で評価し、それぞれに4~0点を与え、投票率を加味した加重平均値を算出する。この平均値を平均オピニオン値(MOS: Mean Opinion Score)という。

図6に新規音声データを50%の割合で伝送するようしきい値を低く設定した場合の結果を示す。しきい値を低く設定した場合、伝送データの多くは新規に登録された音声波形(入力)となり、受信音声データを用いて出力音声合成される。図6の結果より、しきい値を低く設定し、多くの新規音声データが伝送される場合は、分析区間を100ms、

分析ステップ時間を 10ms ~ 15ms に設定することにより良好な出力音声を得られることがわかる。

図 3 は新規音声データを 1% 未満の割合で伝送するようしきい値を十分高く設定した場合の結果である。しきい値を高く設定した場合、伝送データはほぼすべてが既存の音声テンプレートのオフセット値であり、システム内の音声テンプレートデータを基に出力音声合成される。図 11 の結果より、しきい値を高く設定し、音声合成に既存の音声テンプレートが用いられる場合は、分析区間を 100ms ~ 150ms、分析ステップ時間を 5ms ~ 10ms に設定することにより良好な出力音声を得られることがわかる。

図 6、3 の結果から、要求される伝送回線速度において最も良い合成音声を得られる分析パラメータの組合せがあり、その場合の音声品質を予想することができる。

図 8 に伝送速度と評価値 (MOS) の関係を示す。図 8 では、音声の分析において適切な音声分析パラメータを設定することにより、伝送速度を低ビットレートに抑えても音声合成において音声品質の劣化を低減できることがわかる。

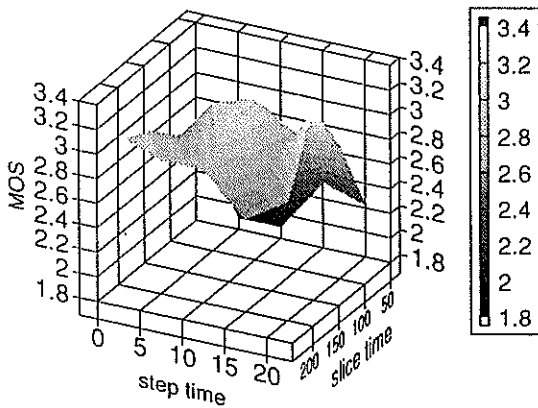


図 6: 分析区間による音声品質の変化 (新規音声の伝送が 50% の場合)

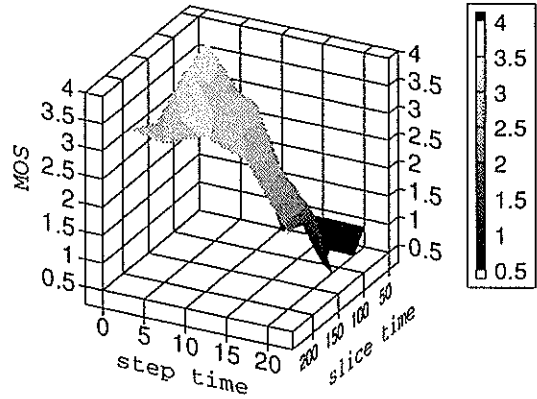


図 7: 分析区間による音声品質の変化 (新規音声の伝送が 1% 未満の場合)

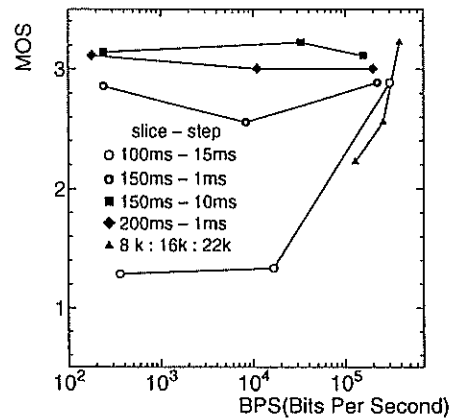


図 8: 平均伝送速度と評価値

4 今後の課題と発展

今後は音声テンプレートの圧縮、統合化に関するものと、前処理として、音圧レベルの抑制、伸長、音程の制御等が上げられる。複数人のテンプレートを効率良く、統合することで、将来的には 1つのテンプレートで全ての音声カバー可能なテンプレートの作成を目指して研究を行なう予定である。

5 研究発表リスト

- 1 発音パラメータを用いた超低速音声通信方式の考察、電子情報通信学会総合大会、金沢、9月、1996

- 2 超低速音声通信に関する研究～パーソナル通信を目指して～、電子情報通信学会、無線通信システム研究会技術報告、RCS97-9, pp.15-20, 5月、1997
- 3 音声テンプレートを用いた超低速音声通信方式に関する考察、電子情報通信学会総合大会、東京、9月、1997