

高速無線キャンパス LAN 上の PHSoverIP 通信における音声通話品質の検討
 Investigation of quality of voice communication of PHS over IP communication system using
 O-2 Wireless LAN campus network

姚 萌† 若原 俊彦† 松本 充司† 清水 隆雄‡ 中谷 信樹‡
 Meng Yao Toshihiko Wakahara Mitsuji Matsumoto Takao Shimizu Nobuki Nakatani

1. まえがき

早稲田大学キャンパス内に構築された 60GHz 高速無線キャンパスネットワークを用いて VoIP 技術の一種である PHS over IP を利用した内線通話が実現されている。PHS over IP は、ネットワークの遅延時間やパケット損失などの伝送品質によって音質が劣化し、音声途切れたり雑音が重畳したり、通話品質は低下する。

PHS over IP などのパケット音声伝送による音声品質の評価に当たっては、従来のオーディオ試験パラメータ (周波数帯域, 歪, ダイナミックレンジなど) だけでは評価できず、声の変化を伴う音声データ圧縮方式の「音声伝達品質評価手法」が必要とされている。

電話の通話品質の主観評価法として MOS (Mean Opinion Score) がある。これは、複数の人間に音声データを聞かせ、「非常に良い」から「悪い」までの 5 段階で評価する方法である。MOS 評価では、被験者の数を多くしないと得られた結果のばらつきが大きくなり、被験者の選び方も重要である。さらに、試験を行うため雑音の影響を受けない無響室などの環境を準備する必要がある。

そこで、音質を簡易に客観評価するため、日本語音声発生ソフトと音声認識機能のソフトを利用して PHS over IP の伝送実験を行い、音質評価を行った。具体的には、キャンパスネットワークのトラフィック変動による PHS over IP の日本語音声の認識特性を評価した。

2. システム構成および主要諸元

このシステムは PHS 信号が 60GHz 無線キャンパスネットワークを介してどれだけ正確に伝わるかを評価するシステムである。送信テキストデータ (話す内容) と受信テキストデータ (聞き取った内容) を比較し、誤った言葉の数をチェックすることで、音声品質を定量的に測定する。図 1 に音声品質評価システムの構成図を示す。

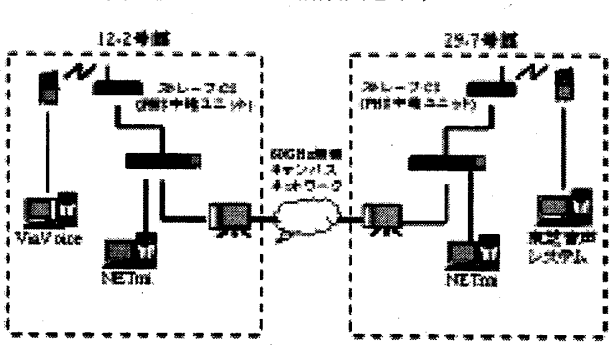


図 1. PHSoverIP 音声品質評価システム構成

† 早稲田大学 国際情報通信研究科

‡ (株)NTT データ

1) 送信側の構成

送信テキストデータの読み上げには東芝製の「おしゃべりテキスト」ソフトを使用し、PC のサウンドカードの出力から音声信号を取り出した。この音声信号を PHS 電話のマイクを介して PHS 中継ユニットにより無線キャンパスネットワークで伝送する。

2) 受信側の構成

無線ネットワークを通ったデジタル音声圧縮データは、PHS 携帯のイヤホンからパソコンのサウンドカードの入力に入り、音声認識しながら認識結果をテキストに変換する。音声認識には IBM 製の音声認識ソフト「ViaVoice」を使用した。音声認識には言語予測処理が必要になるため、CPU の負担も大きいので Pentium III 800MHz の高性能パソコンを利用した。バージョンは「IBM(R) ViaVoice(TM) for windows Release9.1」である。

ここで音声認識処理を向上させるため、「認識ソフト」側のイントネーションや音質を音声認識ソフト「ViaVoice」で学習させ、Enroll 作業を行った。この Enroll ソフトは、最初は認識率が良くなかったが、学習があるレベルに達すると ViaVoice の認識行動が把握できるようになり、音声認識学習効果がでてきた。

今回、実験に使用した PHS システムは日本ビクター (株) の PHS デジタルコードレスシステム「INTAM E S S A」(インタメッサ) である。表 1 にシステムの主要諸元を示す。

表 1. PHS システム主要諸元

項目	内容
交換方式	個別着信・ダイヤルイン方式 (PHS 中継ユニット)
通話路方式	CSMA/CD (IEEE802.3LAN 接続)
無線制御方式	R C R S T D - 2 8 準拠
音声符号化方式	A D P C M (32kbps)
送信出力	1 0 m W
同時通話チャンネル	3 チャンネル / 1 C S あたり

3. 評価結果

今回の品質評価では日本語音素に着目し、日本語音声テキストファイルを全文ひらがなに変換させ、「ひらがな」の出現頻度により、日本語の音素の伝送特性を評価した。また、「NetMi」というネットワークのスループットを測定するツールを利用し、ネットワークに負荷をかけ、ネットワークのトラフィック変動による伝送特性を評価した。

3.1 日本語音素の認識率

読み上げテキストファイルと聞き取ったテキストファイルを比較し、ネットワーク負荷 0% と 90% 時の平均認識率を求めた。図 2 に各日本語音素の平均認識率を示す。

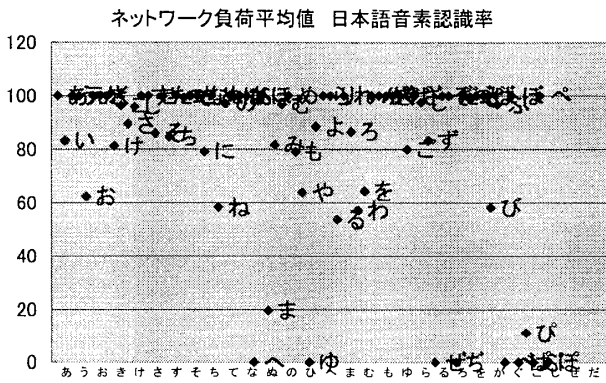


図2 日本語音素平均認識率

日本語音素は全部で71個用い、認識率95%以上は42個、90%~80%は10個、80%以下は11個、0% (使用しなかった音素)は8個であった。ここで得られた音声認識率が95%以上の42個の音素を音声品質評価の実験に用いることとした。

3.2 日本語音素の誤認識率

3.1と同様、読み上げテキストファイルとネットワークを介して受信した音声テキストファイルを比較し、ネットワーク負荷0%と90%時の平均誤認識率を求めた。この誤認識率には、受信テキスト側で正しく認識した音素以外の、他の音から誤認識された音素と、他の音素に誤って識別された音素の双方が含まれている。音素の誤認識率が高いと、その音素は他の音素から誤認識されやすい音素であるといえる。

図3に日本語音素に平均誤認識率を示す。

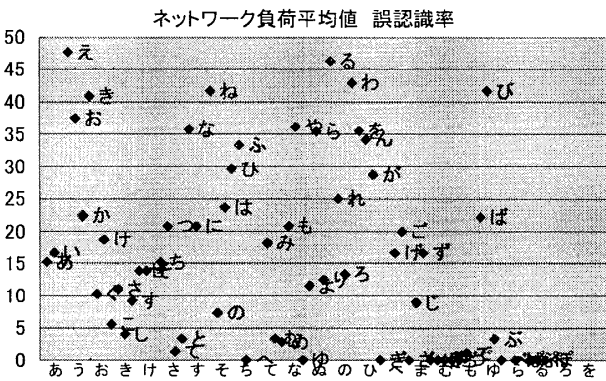


図3 日本語音素平均誤認識率

図3から「ひらがな」音素の誤認識は大体0%~50%の間であることが分かった。また、この図から誤認識率10%以上の音素を削除して10%以下の音素だけを音声品質評価の実験に利用することとした。

また、「ひらがな」音素の単語を連続発音して各音素の誤認識しやすい音素を求めた。誤認識される音素は以下の通りである。

- 1) 類似の音素になる
は → ば, を → う, が → だ,
ど → ぞ, ば → わ
- 2) 別の音素が付加される
ぐ → どう, げ → じえ,
ど → どう, お → おう
- 3) 類似の発音の音素と判別される

び → べい, ぼ → べいお

3.3 FFT分析

FFT分析にあたっては、誤認識した音素のFFT波形を専用ソフトで取り出し、3D-FFT波形として分析した。一例として、「は」の音を連続して発音すると「ば」になる。これは前音の残存する音と新たに発せられた音が重なり、「はは」のスペクトラムが「ば」に似てくることによる。図4に「ば」のFFT波形を示し、図5に「は」の連続発音のFFT波形の例を示す。図5の「は」単音波形と図4の連続発音波形を比較すると、1823.32Hz~4316.64Hz帯域の波形は前音の残存する音と重なり、「ば」音素になってしまうことがわかった。

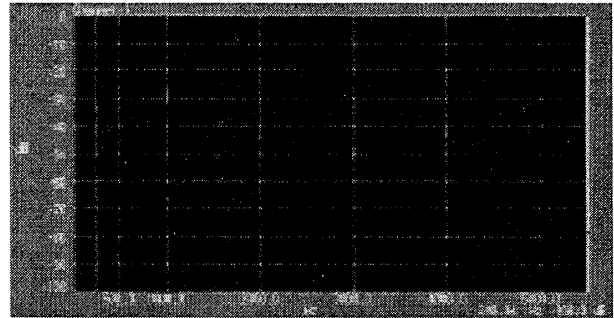


図4 “ば”のFFT波形

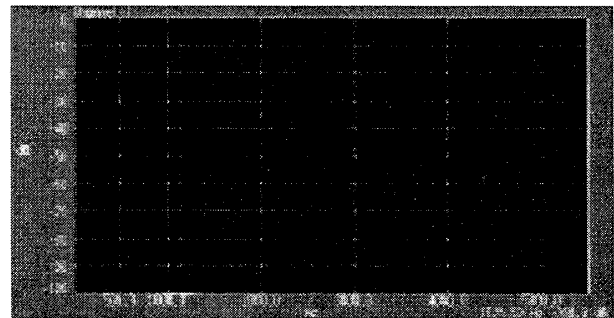


図5 “は”のFFT波形

4. まとめ

本稿では、PHSoverIPによる日本語の音素の認識特性について評価した。今後は、各単音節のFFT波形を取り、各音素間が誤認識されやすいパターンを求める予定である。また、PHSのμ-lawの符号化による音節と、純音節の波形の相関を取り、間違いを起こす原因を追究する予定である。なお、音声認識ソフトと音声認識ソフトには相性の問題があり、安定度と認識率を高くするため他のソフトウェアの適用を検討する。

参考文献

- 1) 川崎, 関, 清水隆雄, “音声認識機能を利用した携帯電話やVoIPなどのデジタル音声通話品質評価手法”, 信学技報 OFS2000-72, pp.13-20(2001-03)
- 2) 清水隆雄, 姚萌, 松本充司, 若原俊彦, 中谷 信樹, “60GHz高速キャンパスネットワークのPHSoverIP通信の一検討-VoIP音声品質測定手法”, 信学技報 OIS2002-9(2002-5)