

一定間隔で削除された音声信号の特徴

白 滝 順*・石 原 学**

Characteristic of Intermittently Eliminated Speech Waves

Jun SHIRATAKI* and Manabu ISHIHARA**

Abstract

The intelligibility of intermittently eliminated Japanese speech wave is discussed in this paper. In our experiment, a segment of a Japanese speech wave is intermittently removed by varying the erasing duration and the period. Here, ratio of the erasing duration to period are varied from 10 : 90 to 90 : 10 every 10 percent. Listening tests for evaluating intermittently eliminated speech wave have been carried out by about 1,000 males aged from 21 to 24 years. Experimental results show that the intelligibility falls from 0.1 ms for 1 ms. There is the condition that improvement of intelligibility is from 1 ms.

1. はじめに

情報の中で音声は人間にとって最も重要で簡単な通信の手段であり、ヒューマンインタフェースの上からも有効な手法である。そのため、音声信号の持つ特徴については現在までに各種の研究が行われてきている¹⁾。例えば、日本語単音節の構造を研究するために語頭を切りとった場合の特徴についての研究²⁾や、時間引き法による音声信号の時間軸圧縮³⁾などについて報告がされている。さらに、音声信号の一部を間欠的に削除した場合の、削除部分と明りょう度の関係について研究が行われている⁴⁾。このときの削除部分の実験は、削除周期の短い場合と削除周期の長い場合の特徴についての研究であった。さらに削除周期を 0.1 ms から 20 ms の範囲についての明りょう度特性についての研究が行われている⁵⁾。これらの研究では、音声信号を部分的に切りとる、例えば、語頭を切りとることなどで、音声信号の個別の特徴を見いだすのではなく、音声信号の全体のどの部分を削除するのかを特に指定しないで、一定周期で削除した場合の音声信号の特徴、すなわち日本語音声信号のどれもが保ち得る単音節明りょう度特性を明らかにするための実験であった。

本論文では、音声信号の削除周期を一定として、さらにその周期の中を 10% から 90% まで削除した場合の日本語音声信号の特徴について単音節明りょう度及び文章了解度の実験を基に報告する。

2. 実 験

本実験システムの構成を図 1 に示す。本実験では、音声信号を DAT から入力し、バンドパスフィルタ(300-5000 Hz, 48 dB/oct)を通した後、波形処理できる範囲まで増幅する。その後、アナログスイッチを利用して、音声信号を一定の時間間隔で削除する。この、削除された音声信号を被験者は、オーディオアンプ(テクニクス社, SU-C01, SE-C01)を通してヘッドホン(パイオニア社, SE-M70)にて聞きとり、解答表に記入する方法で行った。被験者は 21 歳から 24 歳までの聴覚正常な男子 5 名(延べ約 1,000 名)である。そのときのオーディオアンプの出力レベルは、被験者の各個人が一番聞きとり易い値としている。また、室内は静かな 3 m 四方の部屋を利用した。

音声波形の削除周期の関係を図 2 に示す。削除周期は、削除区間と音声区間から構成されており、削除周期と削除区間の比を % で表示した。今回の実験では 10% から 90% まで分割した。削除した音声信号の一部は、特別な波形処理を施さずに聞きとることとした。

1993 年 9 月 24 日受理

* 電気工学科

** 電気工学科非常勤講師

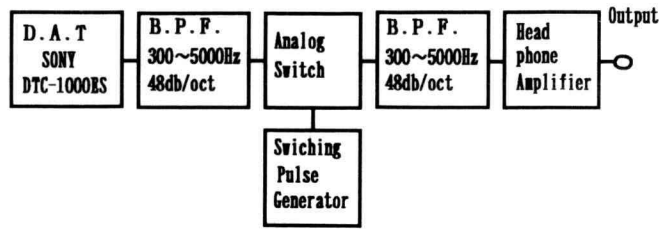


図1. 実験装置のブロックダイアグラム

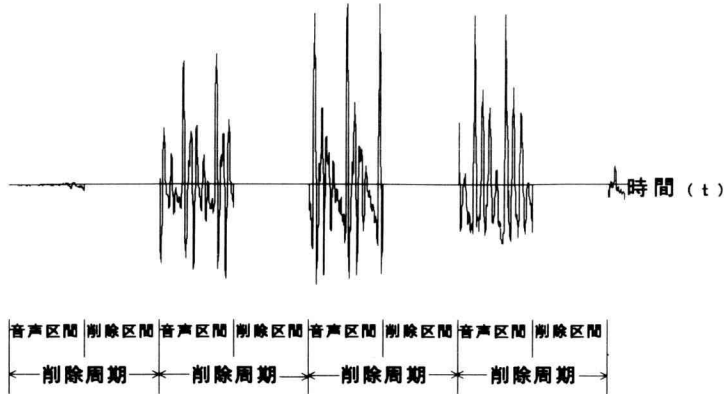


図2. 音声波形の概要と削除部分の関係

2.1 単音節明りょう度の実験と結果⁵⁾

2. に記した実験条件に設定し、単音節音声について実験を行った。実験に用いた音声資料⁶⁾は、NHK アナウンサーによる100単音節の発声テープを利用し、明りょう度の測定を行った。一単音節が正解すると1%として、100単音節すべて正解のときに10%である。

音声信号を削除する時間間隔は、0.1, 0.143, 0.2, 0.5, 1.0, 2.0, 3.33, 5.0, 10.0 および 20 ms の都合10通りについて行った。

実験結果を図3に示す。図からも分かるように、削除周期が0.2 ms よりも小さい場合は、削除区間が90%の場合でも明りょう度が80%以上得られている。しかし、削除周期が大きくなるに従って、明りょう度が低下していくことがわかる。

例えば、削除区間が90%の場合には、削除周期が2 ms 程度まで明りょう度は急激に低下し、それ以後はほぼ一定になっていく。

また、削除区間が80%よりも小さくなっていくと、1 から 2 ms を境にして、明りょう度の低下が鈍化し、または逆に明りょう度が上昇する傾向が見られた。

1 ms から 2 ms は周波数に換算すると、1 kHz から 500 Hz に相当する。この周波数は、音声情報で重要な意味を持つフォルマント周波数の近傍である。また、人が聞こえる周波数としては、良好に聞こえる周波数帯域でもある。これらの時間軸上で削除することにより、高周波成分を始めとするノイズ成分が発生したために、明りょう度の低下をまねくものと考えられる。さらに、この周波数を境として明りょう度が上昇するのは、削除周期が長くなった結果、発生するノイズの周波数が低くなり、そのノイズによる聞き取り難さが低下した結果と考えられる。

以上の実験結果より、削除区間が大きくなるに従い明りょう度は低下するものの、1 から 2 ms 程度の削除区間で明りょう度の低下が鈍化する。さらに、削除区間によっては明りょう度が上昇することがわかった。

2.2 文章了解度の実験と結果

2. に記した実験条件に設定し、文章了解度について実験を行った。

実験に用いた音声資料としては、NHK アナウンサー(男性)による NHK ニュースを DAT に録音して

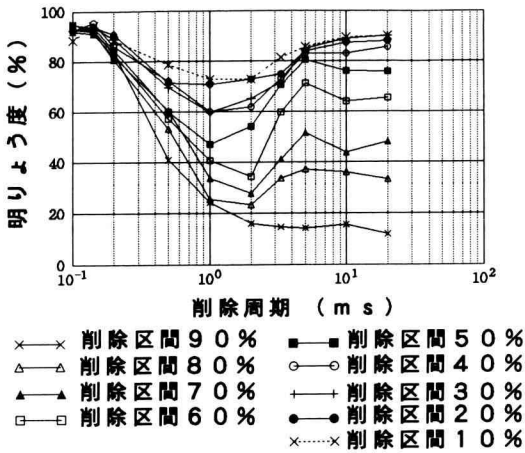


図3. 音声信号の削除時間と明りょう度の関係

利用した。ニュースの内容は天気予報や実験日の近くで何度も報道されるような頻度の高いものをのぞく、ローカルニュース的なものを利用した。ニュースの長さは、時間的には3分~5分程度とし、含まれる単音節数は350単音節~550単音節程度で、意味の途切れない部分で区切った。また、上記の設定のニュースを一文節から三文節程度ごとに5秒程度の時間をあけ、書き写す時間を設定した。また、聞き直しは認めず、すべて一回しか聞くことはできない。被験者は21歳から35歳までの聴覚正常な男性延べ約500名である。このときの文章例を図4に示す。図中の“/”は、書き取りの時間に設けた、一時休止部分である。

了解度の測定では、NHKニュースをいかに正確に聞きとることが出来たかで評価した。聞きとりの評価として、被験者は聞きとった文章を解答表に記入する。

ぐたいきには、/しよくあんがちゅうたいしゃを・
 するほか、/こうこうにしよくあんの/ばんふれっ・
 たいしたばあいには/しよくあんをつうじて/しゆ・
 よう/よびかけるようにしています。/そしてしよ・
 /ちゅうたいしゃには、/しよくぎょうについての・
 /じっちのこうしゅうをおこのない、/しゅうしよ・
 たんとうしゃが/しよくばをまわって/そうだんに・
 かなしどうと/えんじよをおこなうことにしていま・
 うでは/とうきょうのしんじゅと/ながのしよく・
 にえらんで/こんげつからじゅぎょうを/すたーと・
 /ぜんこくのしよくあんに/ひろげていくけいかく・

図4. 文章データの一例

文章中で1単音節ごとに正解すると+1点を与え、間違えると0点とし、文章が全部正解で100点とした。また、聞きとれないものも0点とした。以下に、計算式を記述する。

$$\text{文章了解度} = \frac{\text{正解の単音節数}}{\text{全文章の単音節数}} \times 100(\%)$$

実験結果を図5に示す。本実験での削除周期は、変化の様子を見るために、3.3, 5.0, 10.0 および 20.0 ms の4通りについて表示し、削除期間については10, 30, 50 および 70% について表示している。

実験より、削除期間の大きな90%や70%の了解度の低下が著しいのがわかる。これは、被験者が文章を聞き取る場合に、自分の理解できる文章にしようとした結果、正しく聞き取った文章についても変化が加えられたり、次の文章を聞き取る時に先入観となって、文章が正しく聞き取れなかったものと考えられる。これらは、書きとった文章と試験用テープの比較を行った結果わかった。しかし、その聞き間違いの文章や単語については、必ずしもすべての被験者が同一の間違いをおこすとも限らない。すなわち、被験者の各個人での特徴と考えることができる。

また、音声区間が50%よりも大きい場合には、削除周期に関係なく文章了解度は90%以上得られている。音声区間が50%よりも大きければ、通常の会話等での理解が十分に出来ると考えられる。このことは、被験者が特別な聞き取り試験のための訓練を受けていないこと、さらに聞き返し等を認めていない状態で90%の文章了解度が得られていることによる。

さらに、単音節明りょう度が60%得られている場合の文章了解度が90%程度得られることは、一般の電話回線を利用した了解度とほぼ一致する¹¹⁾。このことから、電話での了解度を基準とした通話音質が同程度と考えられる。また、実際に意味を理解するだけであれば、音声区間を50%よりも小さくすることができる。このことは、試験用テープと書きとった文章を比較した結果、意味を理解していると判断できた。

3. 応用の一例⁷⁾

音声信号の符号化はデジタル技術の発達とともに、現在までに各種の研究が行われてきている⁸⁾。例えば、音声の生成モデルに基づいた符号化などが研究されている⁹⁾。また、時間間引き法による音声信号の時間

圧縮³⁾などについて報告がされている。本論文ではデジタルの時間圧縮に応用する例を示す、即ち、音声信号の一部を間欠的に削除した場合でも明りょう度が保存されていることを利用して、削除した音声信号をデジタルデータとして保存するときに圧縮データとして保存しようとするものである。本方式(以下、I-S圧縮方式と呼ぶ)を利用した圧縮法では、音声データが元のデータの約40%程度に圧縮できる。

3.1 実験

本実験では、音声信号を音声データベースからコンピュータに入力し、圧縮を行った。

本圧縮法の手順について概略を述べる。

- ① 音声データをA/D変換器を利用してコンピュータに入力する。
- ② コンピュータに入力された音声データを、削除周期を0.5msとして、削除周期を設定する。削除周期とは、音声信号をある一定の周期に設定し、さらにその削除周期の中を分割していくものである。削除周期を0.5msに設定するのは、単音節明りょう度が60%以上得られていることにより設定した。

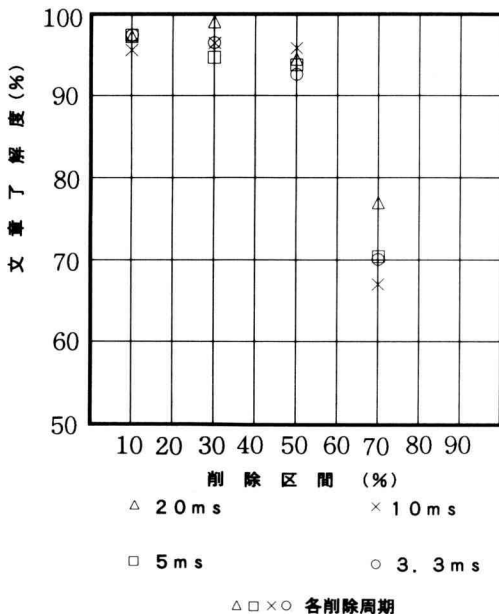


図5. 文章了解度特性

表1. I-S方式と他の圧縮法の比較

	バイト数	圧縮率
圧縮前	約1692k	0%
ファイル圧縮法	約1223k	約27.2%
I-S法	約733k	約56.6%

- ③ この削除周期中の削除区間を1:1(音声信号:削除部分)として、削除部分は削除し音声信号部分は削除周期の50%になる。

3.2 結果及び考察

実験結果を表1に示す。音声信号の削除周期を利用したI-S圧縮方式と、ファイルデータを圧縮する方法(ここでは、ファイル圧縮法と呼ぶ)¹⁰⁾で比較したものである。今回のファイルデータ圧縮法は、一般的なものとしてスライド辞書法とハフマン符号化を組み合わせた方式を利用した。

その結果、I-S圧縮法はファイル圧縮法と比較して、十分圧縮が出来ることがわかる。しかし、このファイル圧縮法はもともとファイルデータの圧縮法であるために、必ずしも有効な比較ではないが、音声データをファイルとして扱い、ソフトウェアにおいて簡単に圧縮が行え、また、簡単に再生が行える。この点を考えると、電子メールなどでの伝送に有効性を見いだすことができる。

表からも分かるように、元の音声データより約40%程度に圧縮することができる。

4. おわりに

本論文では、音声信号を一定の周期で削除した場合の明りょう度および文章了解度を実験について基に明らかにした。

1) 明りょう度の実験からは、削除周期が0.2msよりも小さい場合は、削除区間が90%の場合でも明りょう度が80%以上得られている。しかし、削除区間が80%よりも小さくなっていくと、1msから2msを境にして、明りょう度の低下が鈍化し、または逆に明りょう度が上昇する傾向が見られた。

2) 文章了解度実験からは、音声区間が50%よりも大きいければ、文章了解度が90%以上得られることが

明らかになった。削除時間と削除周期の組み合わせを利用すれば、削除区間を大きくとることも可能である。

これらの実験結果から、音質が低下するものの、何を話しているのか(言語の明瞭性)、誰が話しているかを理解できる範囲を示し、実用になる限界を表していると考えられる。さらに、音声信号の削除部分をつねに子音部分や母音部分・語頭や語尾等と特定しないことにより、削除した場合の平均的な明りょう度および文章了解度が得られていると考えている。

これらのことは、削除した音声信号をどの程度聞きとれるかという、聴覚上の有効な資料になるものと考えられる。

また、削除した音声信号の利用としてファイルとして伝送するときの可能性について述べた。

最後に、被験者になった各位に感謝いたします。

また、本論文の3章での実験には日本音響学会の研究用連続音声データベースを利用した。

文 献

- 1) 例えば、三浦監修：“新版 聴覚と音声”，電子通信学会編（1982）。
- 2) 桑原，境：“連続音声中の切り出し母音および音節の音韻知覚”，音響学会誌，28(5)，pp. 225-234（1972）。
- 3) 加藤，藤原，佐藤：“時間間引きによる音声信号の時間軸圧縮に関する基礎検討”，情報処理平元年前期全大，3K-8，pp. 680-681（1989）。
- 4) 石原，白滝，家入，志方：“音声信号の一部を削除した場合の明りょう度に関する実験的研究”，機械学論 C，Vol. 54，No. 529，pp. 2410-2413（1990）。
- 5) 石原，白滝，家入：“音声信号の削除時間に対する明瞭度特性”，電学論 C，Vol. 103，No. 7，pp. 577-578（平 5. 7）。
- 6) 日本音響学会編：“明瞭度試験法の基準”および“明瞭度試験用テープ”，日本音響コンサルタント協会（昭和 54）。
- 7) 白滝，石原，家入：“音声の削除区間を利用した音声データの圧縮法(I-S 方式の提案)”，平 5 年電気全大，No. 1724（平 5. 3）。
- 8) 例えば，中田和男：“音声の高能率符号化”，森北出版（1986）。
- 9) 例えば，北脇，板倉，齊藤：“PARCOR 形音声分析合成系における最適符号構成”，信学論，J61-A，2，pp. 119-127（1978）。
- 10) 植松，宮下：“可逆圧縮アルゴリズムを理解する”，インターフェース，pp. 88-123（1992. 8）。
- 11) 齊藤，中田：“音声情報処理の基礎”，コロナ社（1986）。