

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 1 日現在

機関番号：17104

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26330197

研究課題名(和文) 連続聴効果を積極的に利用した断続音声の知覚的修復

研究課題名(英文) Perceptual restoration of intermittent speech relying on illusory continuity

研究代表者

水町 光徳 (Mizumachi, Mitsunori)

九州工業大学・大学院工学研究院・准教授

研究者番号：90380740

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：携帯電話の普及に伴い、パケットロスによる通話音声の断続が、深刻な問題となっている。低遅延かつ低演算量での断続音声修復を実現するために、聴覚の錯覚現象である連続聴効果に基づく知覚的修復法について検討した。音声の消失区間に雑音を挿入することにより、我々は断続音を滑らかに知覚できる。提案法は、この心理学的知見を工学的に展開したものである。

本研究では、主観評価実験により、知覚的修復法の雑音環境下での有効性を確認した。また、挿入音として重畳音声を利用した場合の妥当性について、挿入重畳音声の話者依存性についても検証した。

研究成果の概要(英文)：According to a rapid increase of packet-based speech communication, packet loss is one of the most serious problems in speech communication. A perceptual restoration of intermittent speech has been proposed based on an auditory illusion, that is, the phonemic restoration effect. It is, however, difficult to directly employ the phonetic restoration for packet loss recovery due to its harsh gap-filling signal. In this study, utterances through a smart phone were simulated with a head and torso simulator, and noisy environments were reproduced in a soundproof room. The validity of the proposed method was subjectively confirmed under noisy conditions. Furthermore, a human speech-like noise was substituted for the gap-filling signal. The speech-like noises were prepared by repeatedly overlapping short-term human speech signals. It was confirmed that the proposed gap-filling signal succeeded in reducing its noisiness.

研究分野：音情報処理

キーワード：断続音声修復 パケットロス 知覚的修復 連続聴効果

1. 研究開始当初の背景

携帯電話の急速な普及により、パケットロスによる通話音声の断続が深刻な問題となっている。パケットロスの修復に関しては、単一パケットあるいは数十ミリ秒程度のパケットロス隠蔽法として、音声波形置換法[1, 2]や ITU G.723.1 音声コーデック[3]のような音声の特徴に着目した線形予測に基づく手法が提案されている。

近年、新たな問題として、パケット通信量の爆発的な増加に伴って数百ミリ秒に及ぶパケットのバーストロスが頻発している。物理的観点では、バーストロスにより消失した音声成分を厳密に修復することは難しい。ましてリアルタイムでのバーストロス修復は、極めて困難な課題である。

2. 研究の目的

本研究では、パケットロスによる断続音声の修復を低遅延かつ低演算量で実現するために、知覚的修復の有用性について検討する。知覚的修復の基本原理は、我々の聴知覚能力である連続聴効果(音韻修復現象)[4, 5]である。一般に、連続聴効果を生起させるためには、音声の消失部分に、目的信号対雑音のエネルギー比が -10 dB 程度のホワイトノイズを挿入する必要がある[6]。これは、ホワイトノイズによるマスキング効果[7]により、音声が存在するのか消失しているのかを知覚できないようにするためである。連続聴効果の生起条件[8]を満たす雑音を用いると、断続音声は滑らかに修復できても、挿入したホワイトノイズが非常に耳障りとなる。

研究代表者では、これまでに、挿入雑音による不快感を低減することにより、工学的に有用な断続音声の知覚的修復を実現するための基礎研究を行った[9, 10]。断続音声の修復に用いる挿入信号として、音声の短期的かつ長期的な周波数的特徴を考慮し、ホワイトノイズに代わる挿入音について調査し、その有用性を検証した。更に、音声の動的特徴を考慮し、断続音声修復のためのより妥当な挿入音について検討した。音声の周波数的特徴と時間的特徴を断続音修復のための挿入音へ適用することにより、より滑らかな修復音声の実現と挿入音の不快感の低減が実現できることを確認した。

本研究では、より現実的な観点で、

- (1) 雑音環境下での断続音声の知覚的修復
- (2) 重畳音声を挿入音とする知覚的修復について検討する。

3. 研究の方法

(1) 雑音環境下での断続音声の知覚的修復

これまででは、静寂な環境において、パケットロスによる断続音声の知覚的修復法の妥当性を検証してきた。しかし、パケットロス



図1: HATSへの携帯電話の取り付け方法

は、人混みなどの雑音環境下で発生する。そこで、背景雑音が存在する場合に、背景雑音が「音声の滑らかさ」と「挿入音のうるささ」に与える影響を検証する。

本検討項目では、電話での通話を想定し、背景雑音や通話音声の聞こえ方を再現するために、Head and Torso Simulator (HATS)を用いて音源から受聴点(両耳)までのインパルス応答を測定した。インパルス応答の測定は、ETSI EG 202 396-1[11]に準拠し、図1に示すように、HATS右耳にスマートフォンを装着した。HATSは防音室の中央付近に設置し、HATS中心より 1.4 mの距離に背景雑音源として4つのスピーカを設置した。そして、各スピーカとダミーヘッドの両耳位置に存在するマイクロホンまでのインパルス応答を測定した。

背景雑音は、「建築と環境のサウンドライブラリ」[12]と「電子協騒音データベース」[13]より選定した。背景雑音は、同一データの異なる部分を切り出した信号に4つのスピーカとマイクロホン間のインパルス応答をそれぞれ畳み込むことにより、実環境を模擬した。実際に使用した騒音データは、以下の通りである。駅騒音は、駅の改札の音、電車発車の警告音が含まれた音、BGMが含まれた音を選定し、左前と左後のインパルス応答を駅の改札、右前のインパルス応答を警告音、右後のインパルス応答をBGM音にそれぞれ畳み込んだ。人混み雑音は、足音が含まれている音、会話音声が含まれている音、交通音が含まれている音、自転車の警鐘音が含まれる音を選定し、左前のインパルス応答を足音が含まれている音、左後のインパルス応答を会話音声が含まれている音、右前のインパルス応答を交通音が含まれている音、右後のインパルス応答を自転車の警鐘音が含まれる音にそれぞれ畳み込んだ。交通騒音は、自動車の走行音が頻繁にある音と走行音が少ない音を選定し、左前と右前のインパルス応答を走行音が頻繁にある音、左後と右後のインパルス応答を走行音が少ない音に畳み込んだ。車内騒音は、特徴のある音が得られなかったため、方向性については考慮していない。

(2) 重畳音声を挿入音とする知覚的修復

断続音声の知覚的修復では、挿入音の時間・周波数特性を音声の性質に近づけることにより、修復音声の滑らかさと挿入音のうるささが低減できることがわかった。そこで、音声そのものを挿入音として使用することも考えられるが、予備検討の結果、連続聴効果を生起させるためには非常に大きなレベルで音声を挿入する必要があり、なおかつ発話内容によって修復音声の滑らかさの評価が大きく異なることがわかった。これらの課題を解決するために、断続音声の知覚的修復のために、音声を多数回加算して作成する重畳音声[14]を用いる。

重畳音声は、重畳回数によりその特徴が異なる。小林らの報告[14]によると、重畳回数が2~10回では音声の重ね合わせと知覚でき、16~256回では音声の特徴と定常的な雑音の特徴とを合わせ持ち、512~4096回では定常的な雑音と知覚される。そこで、断続音声の知覚的修復に適した重畳回数を検討するために、重畳回数の異なる重畳音声を作成する。

本手法では、複数話者が発話した単語音声を用意し、それを連結したものを500 msのブロックごとに分割する。本研究では、150 ms程度のバーストロスを含む断続音声を用意しているため、500 msのブロックから150 msの音声信号を切り出す。この際、切り出し開始位置はランダムとする。この150 msの音声ブロックを複数回重畳する。

4. 研究成果

(1) 雑音環境下での断続音声の知覚的修復

提案法の雑音環境下での有効性を検証するために、現実的な音声対背景雑音エネルギー比(Speech-to-Noise Ratio; SNR)と音声対挿入音エネルギー比(Speech-to-Insertion Ratio; SIR)で聴取実験を実施した。予備実験の結果、SNRは、3 dB、6 dB、9 dBとし、SIRは、-6 dB、-3 dB、0 dBと設定した。

聴取実験の参加者は、正常な聴力を有する大学生及び大学院生(男性8名、女性2名)である。実験参加者は、それぞれの修復音声ヘッドホンで聴取し、修復音声の滑らかさ、修復音声のうるささ、総合評価の3項目について、7段階MOS評定を行った。

主観評価実験の結果を図2に示す。図2では、修復音声の滑らかさ(上段)と総合評価(下段)は、値が大きいほど望ましく、修復音声のうるささ(中段)に関しては、値が小さいほど望ましい。修復音声の滑らかさ(上段)は、60 msのバーストの場合、SIRに依存せず、バースト長が大きいほどSIRに比例して評価値が小さくなった。修復音声のうるささ(中段)は、SNRに比例してMOSが大きくなったが、その差は小さい。総合評価(下段)に関しては、バースト長によって、SNRとMOSの関係が異なることがわかった。

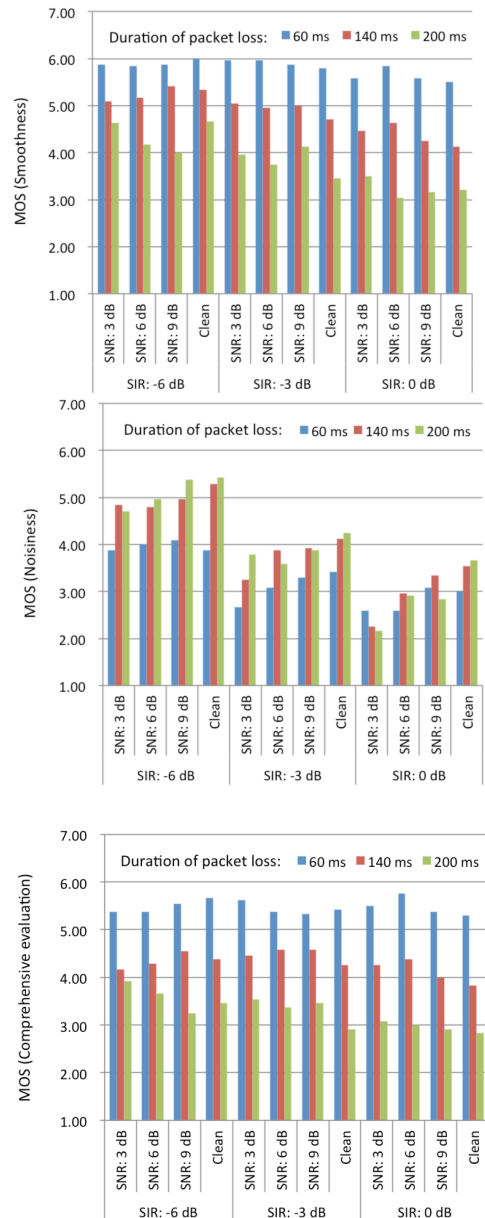


図2: 提案法の雑音環境下での主観評価結果 (上段: 修復音声の滑らかさ、中段: 修復音声のうるささ、下段: 総合評価)

(2) 重畳音声を挿入音とする知覚的修復

断続音声の知覚的修復における挿入音による差異を検証するために、重畳音声として用いる提案法と調波複合音と環境雑音の混合音を挿入音とする従来法との比較評価実験を行った。なお、重畳音声は、被修復音声の話者の音声のみで作成した話者依存条件と、不特定話者の混合音声の2条件とした。

図3に実験結果を示す。修復音声の滑らかさに関しては、図3上段より、被修復音声の発話話者音声よりも、それ以外の話者による音声で作成した重畳音声の方が高い評価を得ている。また、SNRが大きい場合は提案法

と従来法の差異は見られないが、SNR が 3 dB の条件では、わずかに提案法が優位であることが確認できる。挿入音のうるささに関しては、評価値は小さい方が望ましい。図 3 中段より、提案法は、従来法より有意に優れていることを確認できる。なお、重畳音声は、被修復音声の発話話者の音声を用いて作成することが望ましい。また、これらの関係は、SNR に依存しない。総合評価に関しては、図 3 下段より、背景雑音が存在しない静寂な環境下では、従来法が提案法よりも高い評価を得ているが、背景雑音が存在する環境下では重畳音声挿入音がわずかに優位であることが確認できる。

以上の結果より、重畳音声の作成にあたっては、被修復音声の発話話者の音声を利用することが望ましい。従って、通話開始後に、受信した音声から挿入音を作成しながら、断続音声の修復を行うことが望ましいことが示唆された。

<参考文献>

- [1] D. J. Goodman *et al.*, IEEE Trans. Acoust. Speech and Signal Process., vol. 34, Issue 6, pp. 1440-1448 (1986).
- [2] O. J. Wasem *et al.*, IEEE Trans. Acoust. Speech and Signal Process., vol. 36, Issue 3, pp. 342-348 (1988).
- [3] ITU Recommendations G.723.1, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s,” (1996).
- [4] G. A. Miller and J. C. R. Licklider, J. Acoust. Soc. Am., 22, 167-173 (1950).
- [5] R. M. Warren, Science, 167, 392-393 (1970).
- [6] 柏野 牧夫, 日本音響学会誌, 61(5), 263-268, 2005.
- [7] 柏野 牧夫, 音のイリュージョン, 岩波書店, 2010.
- [8] M. Kashino, Acoustical Science and Technology, 27(6), 318-321, 2006.
- [9] M. Mizumachi *et al.*, Proc. ICSV19, Paper ID: 208, July 2012.
- [10] M. Mizumachi *et al.*, Proc. ICSV20, Paper ID: 173, July 2013.
- [11] ETSI EG 202 396-1, 2008.
- [12] 社団法人日本建築学会, 建築と環境のサウンドライブラリ, 技報堂出版, 2004.
- [13] 社団法人電子情報技術産業協会音声入出力方式標準化委員会, 電子協騒音データベース, サンライズミュージック.
- [14] 小林ら, “ヒューマンスピーチライク雑音における音声的特徴の分析,” 信学技報, SP95-105, 1995.

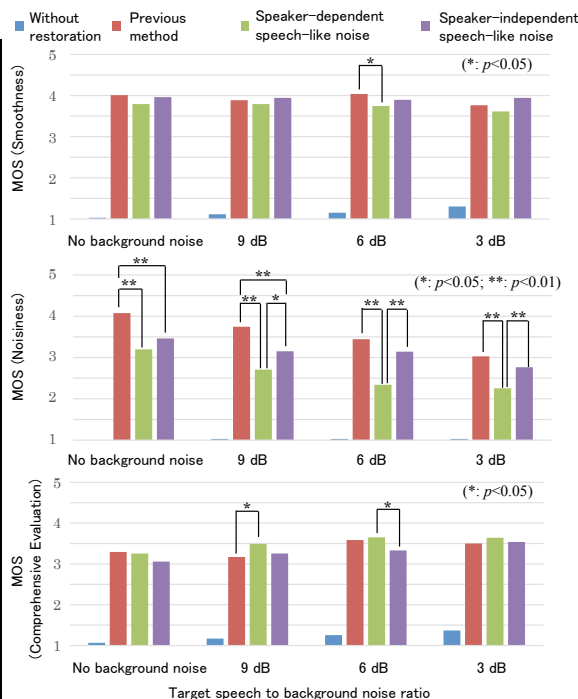


図 3: 重畳音声を挿入音として用いた提案法の主観評価結果 (上段: 修復音声の滑らかさ, 中段: 修復音声のうるささ, 下段: 総合評価)

5. 主な発表論文等

[学会発表] (計 3 件)

- ① M. Mizumachi, S. Imanaga, and T. Horiuchi, “Perceptual restoration of intermittent speech using human speech-like noise,” Proc. 23rd International Congress on Sound and Vibration (ICSV23), Paper ID: 1050, July 10-14, 2016, Athens, Greece.
- ② M. Mizumachi, S. Motomura, T. Takakura, and T. Horiuchi, “Perceptual restoration of intermittent speech under noisy environments,” Proc. 22nd International Congress on Sound and Vibration (ICSV22), Paper ID: 1050, July 12-16, 2015, Florence, Italy.
- ③ M. Mizumachi, S. Motomura, T. Takakura, and T. Horiuchi, “Perceptual restoration of intermittent speech relying on illusory continuity,” Proc. 7th Forum Acousticum, Paper No. PJ-3-5, Sep. 7-12, 2014, Krakow, Poland.

〔図書〕(計 1 件)

① 水町 光徳, “音環境理解,” 音響キーワードブック, 日本音響学会編, pp. 45-46, コロナ社, March 2016.

〔その他〕

ホームページでの情報発信

<http://www.sp.ecs.kyutech.ac.jp/~mizumach/>

6. 研究組織

(1) 研究代表者

水町 光徳 (MIZUMACHI, Mitsunori)

九州工業大学・大学院工学研究院・准教授

研究者番号： 90380740

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし