Vol.2019-MUS-123 No.31
Vol.2019-SLP-127 No.31
2019/6/22

定常雑音環境下での単語音声先頭部の話速変換による単語了解度の向上

村尾天地 [1,a)]　村上泰樹 [1]

概要：
高騒音環境において音声を聞き取りやすくする方法として、音声の拡声が用いられる。しかし、音声の音量を適切な値にすることは困難である。そこで本研究の目的は音声を増幅するのではなく、話速を変化させることによって音声の聞き取りを向上させることである。これまでの研究で、話速を変化させて聞き取りを向上させる試みが行われていたが、定常雑音環境下では効果がなかった。この実験で使用された音声は、話速を一様に伸長されていたものが使用されていた。また、定常雑音環境下で効果がないという以外に、聞き取り率が著しく悪かったのは、単語音声の最初の第一モーラであることも報告された。そこで本研究では、聞き取り率が悪かった単語音声先頭部を伸長させ、聞き取りの向上を試みた。その結果、定常雑音環境下でも聞き取りの向上があった。このことから、音声の聞き取りには単語の先頭部分の話速が重要であることが示唆される。

Word intelligibility improvement by speech rate conversion of the beginning of waveform under stationary noise

1. はじめに

高騒音環境では、音声を聞き取りやすくする方法として、音声の拡声が用いられる。音声を拡声する際に、近くにいる人に聞き取りやすいように音量を設定すると、遠くにいる人は聞き取りづらくなる。逆に遠くにいる人に聞き取りやすいように音量を設定すると、近くにいる人はうるさく感じる。このように、人は音声を聞き取る位置によって聞き取りやすい音の大きさが異なるため、聞き取る人の位置によって音声の音量を適切な大きさにすることは困難である。そこで本研究では、音声の振幅を増減するのではなく話速を変化させることに着目し、音声の聞き取りを向上させることを目的とする。

これまでの研究[1][2]で、話速を変化させて聞き取りを向上させる試みが行われていた。しかし、定常雑音環境下では話速を変化させても聞き取り率が向上しないことが示された。また、聞き取り率が著しく悪い箇所は、音声の第一モーラであることが報告されている[2]。これらの実験で使用された音声は、話速を一様に伸長されていたものが使用されており、モーラごとの話速は同じである。したがって、音声先頭部の話速を優先的に遅くすることで、第一モーラが聞き取りやすくなり、聞き取り率が向上するのではないかと推測した。

そこで本研究では、第一モーラに注目するにあたって、音声先頭部を優先的に伸長した単語を使用することで定常雑音環境下での聞き取り率向上を試みた。

2. 手法

本研究では音声の先頭部の話速を変化させるために、音声分析合成系である WORLD[3][4]を用いて話速を変化させた。WORLD とは音声合成をする際に基本周波数、スペクトル包絡、非周期指標を独立に操作を行えるシステムである。音声波形を伸長させる際には、話速のみを遅くさせることが可能である。ここで、本研究では式(1)を用いて Speech rate f を制御した。

$$f = \frac{-A1 + A2}{2} * \tanh\big(B * (T - C)\big) + (A1 + A2) \ldots (1)$$

ただし、変化させる音声先頭部の話速を $A1$、変化させる音声終端部の話速を $A2$ とする。また、話速を変化させるときの変化率を B [1/s]、音声全体の時間を T [s]、話速を変化させる時間を C [s] とする。

1 大島商船高等専門学校
National Institute of Technology, Oshima College
a) d811@oshima-k.ac.jp

ⓒ 2019 Information Processing Society of Japan

Vol.2019-MUS-123 No.31
Vol.2019-SLP-127 No.31
2019/6/22

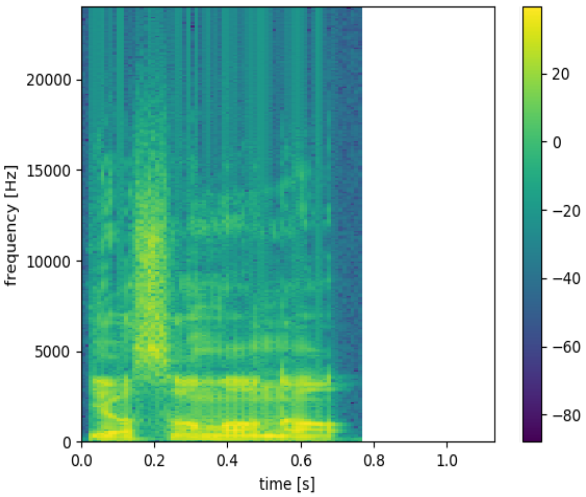

図 1. 原音声のパワースペクトル

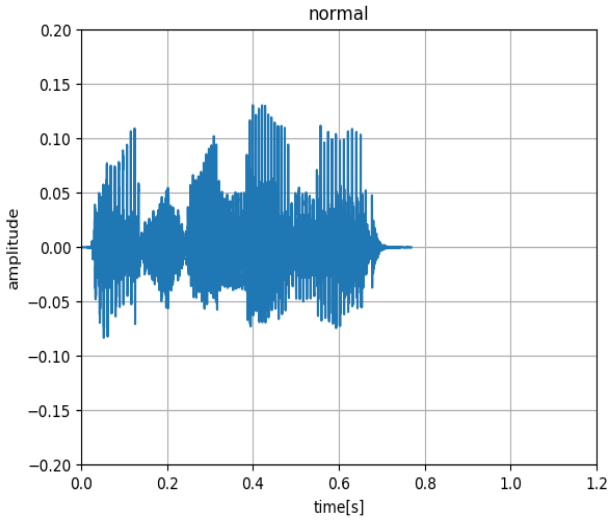

図 2. 原音声の音声波形

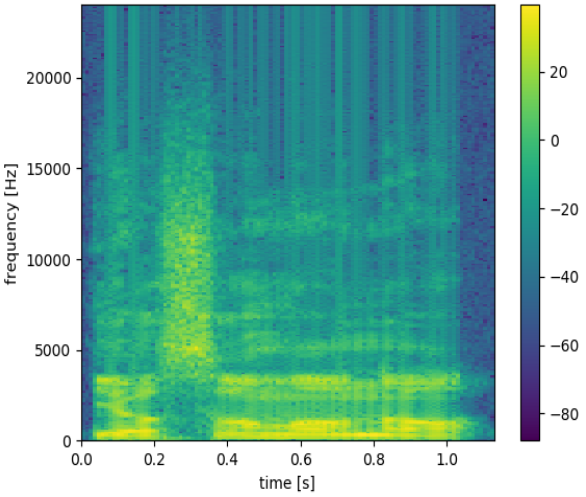

図 3. 一様に伸長した際のパワースペクトル

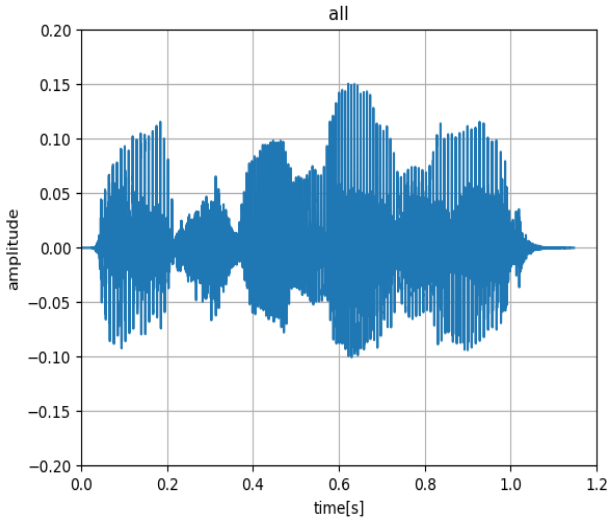

図 4. 一様に伸長した際の音声波形

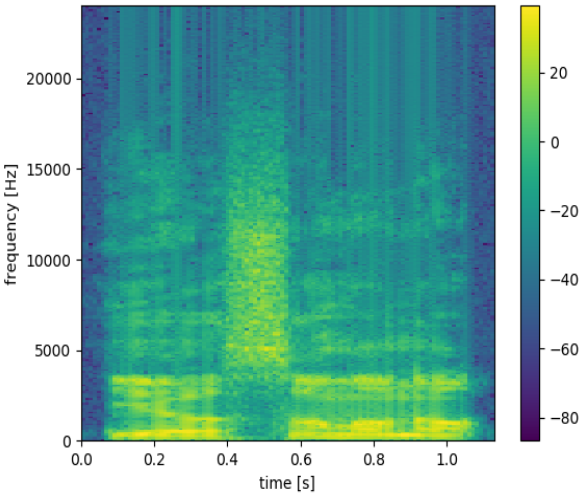

図 5. 先頭部を優先的に伸長した際のパワースペクトル

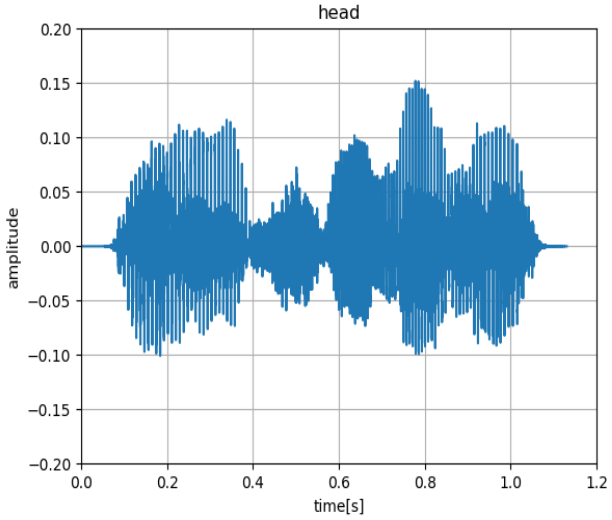

図 6. 先頭部を優先的に伸長した際の音声波形

ⓒ 2019 Information Processing Society of Japan

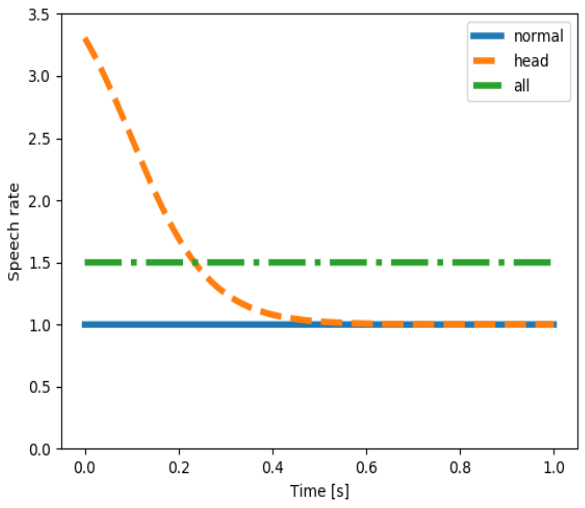

図 7. 話速を変化させた際の各音声の Speech rate

WORLD と式(1)用いて話速を変化させた音声のスペクトラムと波形を図 1 ~ 図 6 に示す。図 1, 3, 5 は各音声のパワースペクトルを表している。図 2, 4, 6 は各音声の音声波形を表している。この図は実際に実験で使用した音声データの 1 つである 4 モーラ単語音声「ヨソモノ」を用いて作成した。図 7 は話速を変化させた際の Speech rate を表している。実験では話速を 3 つの条件で変化させた。条件 normal は原音声、条件 all は一様に伸長させた音声、条件 head は先頭部を優先的に伸長させた音声を表している。条件 normal は原音声である。式(1)で使用した値は A1 =1、A2 =1、c =0.1 である。条件 all では式(1)を使用して一様に話速を変化させた。変化させるにあたって使用した値は A1 =1.5、A2 =1.5、c =0.1 である。条件 head では式(1)を使用して先頭部を伸長させた。変化させるにあたって使用した値は A1 = 4、A2 = 1、c = 0.1 である。

実験で用いた原音声には、単語親密度が統合されている 4 モーラの日本語音声データセット[5]から話者が男性の音声を使用した。本実験では、最も親密度が高い単語音声を使用し実験を行った。実験を行う際に使用した音声は、疑似音声雑音と音声を加算した刺激を使用した。音声の聞き取りを 50%程度にするために SNR は-6 dB に設定した。加算した雑音には日本語音声データセットに付随している疑似音声雑音を使用した[5]。

実験には、19 歳から 35 歳まで平均 21 歳の聴力の異常をうったえない日本語を母語とする男性 11 名が参加した。今回は、条件 normal、all、head の 3 種類を各 30 単語ずつ、計 90 単語を用いて実験を行った。実験内容は聞こえた単語を書き取る試験を行った。

実験は、簡易防音室で行った。本実験前に、音圧レベルを各実験参加者たちに聞き取りやすいレベルに調節してもらい実験を行った。

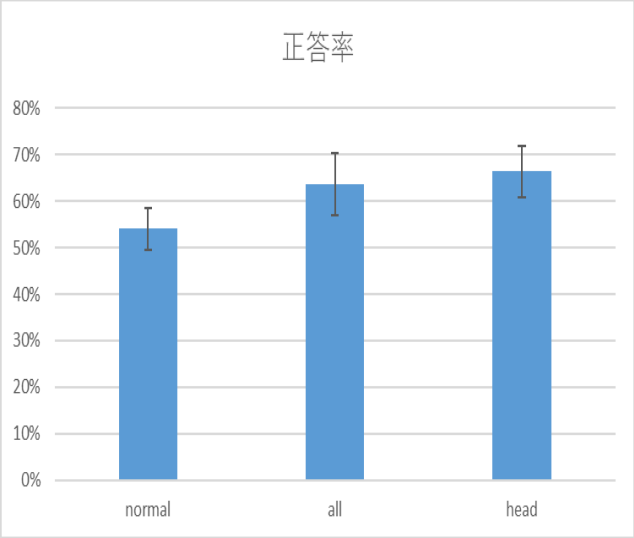

図 8. 各話速ごとの単語了解度

表 1 原音声と変化させた各話速ごとの相違

条件	t 検定
all	ns
head	*

ns : 非有意 *: $p < 0.05$

3. 実験結果

図 8 に各話速ごとの単語了解度を示す。話速ごとの単語了解度は、原音声は 54%一様に伸長した音声は 64%音声の先頭部を優先的に伸長させた音声は 66%であった。

原音声と変化させた各話速ごとの単語了解度で t 検定を行い有意差を調べた。検定の結果を表 1 に示す。話速を一様に伸長させた音声の場合、有意差は認められなかった($t(10) = 1.63, p > 0.05$)。 先頭を優先的に伸長させた音声の場合、統計的に有意な差が認められた($t(10) = 2.27, p < 0.05$)。

4. 考察

これまでの研究[1][2]では、定常雑音環境下で音声の話速を、一様に伸長した場合に聞き取り率が向上することは示されなかった。同様に、本研究では図 8 と表 1 に示すように、原音声に対して一様に話速を変化させた音声の単語了解度は変化しなかった。この結果は、これまでの研究[1][2]と同様である。しかし、図 8 と表 1 に示すように、原音声に対して単語音声先頭部の話速を変化させた音声の単語了解度は向上した。

ここで、提案法を用いることで、単語了解度が向上した理由を考察する。図 3 と図 5 に表す音声のパワースペクトルを比較すると、音声先頭部を伸長した音声は、一様に伸

ⓒ 2019 Information Processing Society of Japan

長した音声と比べ先頭部の話速が遅くなっている。この為、今まで聞き取ることができていなかった第一モーラを、聞き取ることができたため聞き取り率が向上したと考える。また、先頭部以外の話速は一様に伸長したものと比べると早いため、先頭部以外の話速を遅くすることは、聞き取りの向上に効果的でないと考える。これらのことから、単語音声先頭部を伸長させることは聞き取りにおいて効果的であると考える。

5. おわりに

本研究の目的は、定常雑音環境下で話速を変化させることにより、聞き取り率を向上させることである。本研究では、聞き取りが悪かった単語音声先頭部を伸長させ、聞き取り率の向上を試みた。そして原音声と一様に伸長した音声、単語音声先頭部を伸長した音声を用いて単語了解度試験を行った。実験の結果、これまでの研究と同様に一様に音声を伸長させても定常雑音環境下で効果が無いという結果が出た。しかし、話速の先頭部を伸長させた音声では、定常雑音環境下でも音声の聞き取りの向上が示された。これは、音声の聞き取りの際に、音声の先頭部が聞き取りに重要であることが示唆される。

参考文献

[1] Martin Cooke, and Vincent Aubanel, ``Effects of linear and nonlinear speech rate changes on speech intelligibility in stationary and fluctuating maskers" The Journal of the Acoustical Society of America 141, 4126 (2017) Pages: 4126–4135
[2] 村上泰樹，芳崎大我，``一様な話速の変化が雑音環境下での単語了解度に与える影響，" 日本音響学会秋季研究発表会，2-P-43, Sept. , 2018
[3] M. Morise, F. Yokomori, and K. Ozawa, ``WORLD: a vocoder-based high-quality speech synthesis system for real-time applications," IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877-1884, 2016..
[4] M. Morise, ``D4C, a band-aperiodicity estimator for high-quality speech synthesis," Speech Communication, vol. 84, pp. 57-65, Nov. 2016..
[5] 近藤公久 , 天野成昭 , 坂本修一[他] , 鈴木陽一, 「親密度別単語了解度試験用音声データセット 2007(FW07)」電子情報通信学会技術研究報告.SP,音声 107(434), 43・48,2008-01-17

ⓒ 2019 Information Processing Society of Japan