

話速変換による単語了解度の音韻的特徴の検討*

○渋谷 徹^{*1+}, 小林 洋介^{*1}, 近藤 和弘^{*1}

(*1:山形大学大学院理工学研究科, +:日本電気通信システム株式会社)

1 はじめに

放送番組などの視聴において、話す速度が速いと音声聞き取りににくいと感じることがあるが、その解決手段のひとつに、話す速度を変える話速変換がある。話速変換の研究および実用化に関しては、「ゆっくり話すと分かりやすい」等の先行研究^[1]でもその効果が報告されている。一方、リアルタイム処理においては、話速伸張を行ったことによる絶対時間に対する遅れが生じてしまうという問題があるので、言語間のポーズ(“間”)の時間長制御に関する検討^[2-3]や、言語発声時の特徴を捕らえた“はじめゆっくり徐々に早く”の制御手法の検討^[4]が行われている。さらに、音声を伸張することによって生じる映像との相対遅延(リップシンク)の課題については、音声のみを伸張したときの映像と音声の“ずれ”の検知限や許容限が研究されている^[5]。また、話速変換を行わず明瞭度をあげる方法としては、音声定常部のエネルギーを抑圧する定常部抑圧処理^[6]や子音強調処理^[7]といった報告がある。

そこで本研究では、話速と聞き取りやすさ(単語了解度)の関係を、音韻ごとに捉えることにより、話速を遅くして聞き取りやすくなる(単語了解度が向上する)音韻や、反対に話速を早くしても聞き取りにくくならない(単語了解度が低下しない)音韻、といったように、音韻の特徴に着目した話速と了解度の関係について研究を行っている。

このたび、話速変換による日本語音声の単語了解度の変化を、ノイズを印加した状態における二者択一型の DRT 了解度試験^[8]で検証した。その結果、話速を遅くすると了解度が向上する音韻や話速を早くすると了解度の低下がある音韻、話速の変化の影響を受けにく

い音韻があることがわかり、音韻特徴によって話速変換と単語了解度の関係に影響があることが示唆されたので報告する。

2 単語了解度の評価試験

2.1 DRT 試験

主観音質の評価手法としては、受聴音の分かりやすさを評価する明瞭度・了解度試験 Diagnostic Rhyme Test (DRT) がある。DRT は語頭 1 音素のみ異なる単語対の候補の内一方を被試験者に聴取させ、単語対の内一方を選ばせる了解度試験方法で、熟練していない被試験者でも安定した了解度を計測できることが特徴である。今回、話速変換による日本語音声の聞き取りやすさの評価方法として、音声了解度の評価方法^[8-10]およびその推定方法^[11-12]、さらに各種外乱のある音声了解度の評価方法^[13-14]を用いて評価を行った。

2.2 二者択一音声の単語リスト

本研究では、文献[8]で報告されている日本語の 2 モーラ単語対リストを用いた。この単語対は、Voicing (有音性), Nasality (鼻音性), Sustention (継続性), Sibilation (不規則性), Graveness (拗音性) および Compactness (集中性) の 6 種類の音韻特徴に分類されており、それぞれ 10 単語対(各 20 単語)のリストがある。単語対の例を表 1 に示す。

2.3 話速変換方法

話速変換の音声加工には、フリーソフトの Audacity^[15]を用いた。原信号に対して 1.6 倍に時間を伸張したゆっくりの話速“1.6x-slow”と 0.6 倍の時間に短縮した早い話速“0.6x-quick”を同ツールで加工して作成した(図 1)。

2.4 ノイズ重畳

ノイズ環境下での了解度比較を想定し、外乱としてマルチトーカーノイズ(男女複数人

* Differences in the Effect of Speech Rate on Intelligibility in Artificially Speed-Altered Speech by Phonetic Feature, by Toru Shibuya^{*1+}, Yosuke Kobayashi^{*1} and Kazuhiro Kondo^{*1} (*1: Graduate School of Science and Engineering, Yamagata University, +: NEC Communication Systems, Ltd.)

表 1. DRT 単語対の例

音韻分類	音韻の特徴	語頭の子音対	単語対の例
Voicing	有声性 (有声音 /無声音)	z-s d-t g-k	ザイ-サイ ダク-タク ギン-キン
Nazality	鼻音性 (鼻音 /口音)	m-b n-d	マン-バン ナイ-ダイ
Sustention	継続性 (狭窄音 /閉鎖音)	h-k sh-ch s-t	ハシ-カシ シリ-チリ スキ-ツキ
Sibilant	不規則性 (粗擦音 /円熟音)	j-g ch-k sh-h	ジャム-ガム チャク-カク シキ-ヒキ
Graveness	抑音性 (抑音 /鋭音)	w-r p-t m-n	ワク-ラク パイ-タイ ムク-ヌク
Compactness	集中性 (集約性 /拡散性)	y-r k-p g-b	ヤク-ラク カイ-パイ ギン-ビン

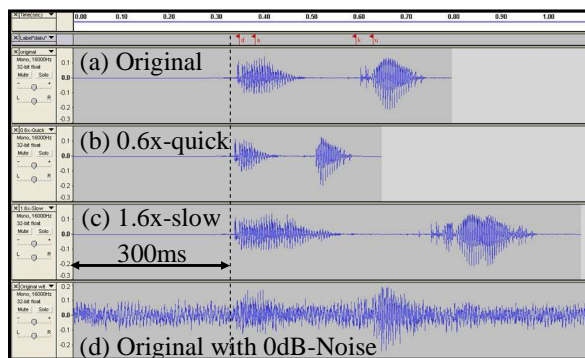


図 1. 話速変換の信号波形 “daku”

の混合音声雑音)^[16]を使用した。実験に使用する 120 単語の時間総和の信号レベルと同一レベルになるノイズレベルを 0 dB としてレベル統制を行った。実験に使用するノイズレベルは、原音に対して 0, 4, 8, 12, 16 dB のマルチトーカーノイズを印加したものとした。なお、発話者は 50 代女性の一種類とした。

2.5 音声提示方法

被試験者に提示する音声信号は、図 1 (d) に示すように、開始点はノイズ音とし、開始点から 300 msec 後に音声信号が重畳されるように時間統制を行った。被試験者へはランダムに再生し、ヘッドホンで提示した。

被試験者は「どちらの単語を発話したように聞こえたか？」を PC 上に表示される応答



図 2. 応答入力ダイアログボックスの例

る。なお、選択決定するまでの聞き直し回数は無制限とし、選択完了後に次の単語を提示している。また、途中休憩も可とした。

2.6 被試験者

被試験者は日本語を母国語とする健聴者の 20 代男性 6 名とした。なお、提示音圧は被試験者以外の健聴者が、すべての音源がはっきり聞こえるように予め調整したレベルを被試験者全員が使用した。また、事前に試験方法の説明と模擬体験のトレーニングを行った後に本番の試験を実施している。

2.7 試験条件のまとめ

DRT の試験条件についてまとめたものを表 2 に示す。なお、試験に用いた提示音声は合計 2,160 単語である。

2.8 正答率の算出

本試験の結果である単語了解度は正答率で表す。なお、二者択一方式の偶発性を避けるため、正答率は式(1)により算出している。すなわち、全数正解で 100%、全くでたらめ (半数正解) に答えると 0% となる。

$$\text{正答率} = (\text{正答数} - \text{誤答数}) / \text{総提示数} \times 100 [\%] \quad \cdots (1)$$

表 2. DRT の試験条件

項目	条件	種類
単語リスト	2モーラの単語対 6種類の音韻特徴	120単語
話速変換	original (原音) 1.6x-slow (ゆっくり) 0.6x-quick (早く)	3種類
ノイズ種類	マルチトーカーノイズ	1種類
ノイズレベル	0, 4, 8, 12, 16 dB, ノイズなし (原音)	6種類
被試験者	20代男性 (健聴者)	6人

合計：2,160単語

3 試験結果と考察

3.1 全平均における単語理解度

図3はノイズ量に対する全音韻特徴の正答率の評価結果を話速ごとに示したものである。この図から、0.6x-quickはoriginalや1.6x-slowに比べて理解度が低下していることが分かる。

0.6x-quickとoriginalの2つの話速における、ノイズレベル5種類(0, 4, 8, 12, 16dB)の繰り返しのある二元配置の分散分析の結果は、 $[F(1, 300) = 13.27, p < 0.01]$ で有意であった。一方、1.6x-slowとoriginalでは大きな傾向の違いは見られなかった。

3.2 音韻特徴ごとの単語理解度

図4～9は、図3を音韻特徴別に分けて示した図である。

図4に示すVoicing(有声性)では、0.6x-quickの場合に理解度が低下する傾向がみられるが、1.6x-slowとoriginalでは大きな違いがないことがわかる。同様の傾向がGraveness(拗音性)にもみられた(図8)。なお、0.6x-quickとoriginalのノイズレベル5種類での分散分析の結果は、それぞれ、 $[F(1, 50) = 29.17, p < 0.01]$ 、 $[F(1, 50) = 18.46, p < 0.01]$ で有意であったことから、話速が早いときに理解度の違いがあるといえる。

一方、話速が遅い場合に理解度の違いがみられたのは、Compactness(集中性)の場合であった。図9から、1.6x-slowの理解度がoriginalおよび0.6x-quickより高いことがうかがえる。1.6x-slowとoriginalおよび1.6x-slowと0.6x-quickのノイズレベル5種類での分散分析の結果は、それぞれ、 $[F(1, 50) = 4.25, p < 0.05]$ 、 $[F(1, 50) = 9.71, p < 0.01]$ であったことか

ら、話速が遅いときの了解度には差があると考えられる。

他の音韻特徴であるNasality(鼻音性)、Sustention(継続性)、Sibilation(不規則性)においては、話速の違いによる単語理解度の有意差がないことがわかる(図5～7)。

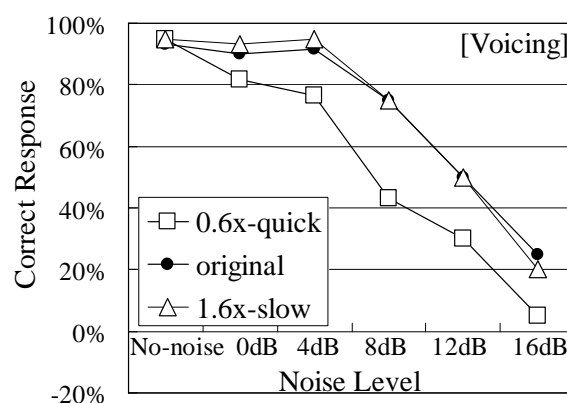


図4. DRT 評価結果 [Voicing]

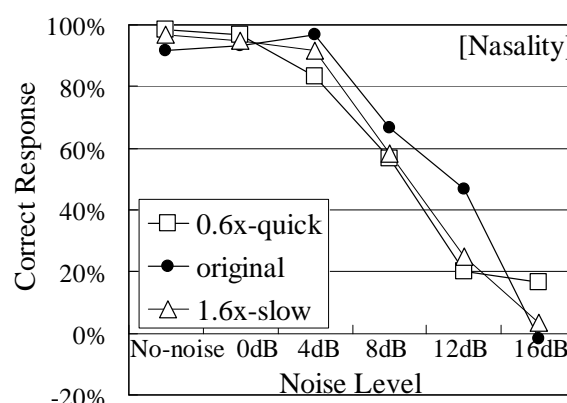


図5. DRT 評価結果 [Nasality]

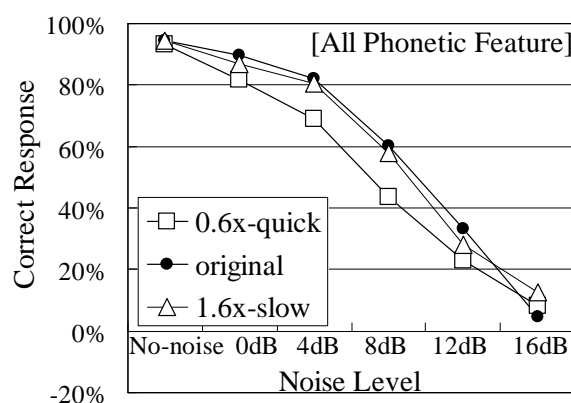


図3. DRT 評価結果 [全音韻特徴]

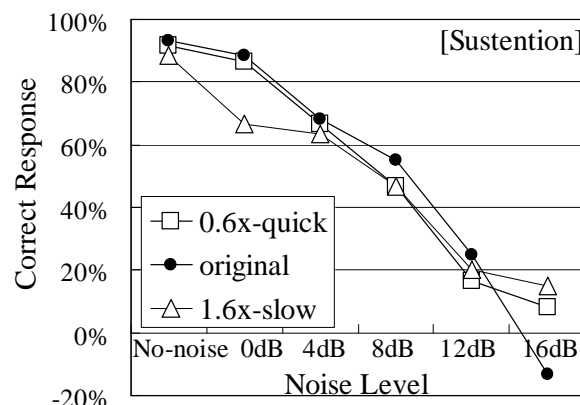


図6. DRT 評価結果 [Sustention]

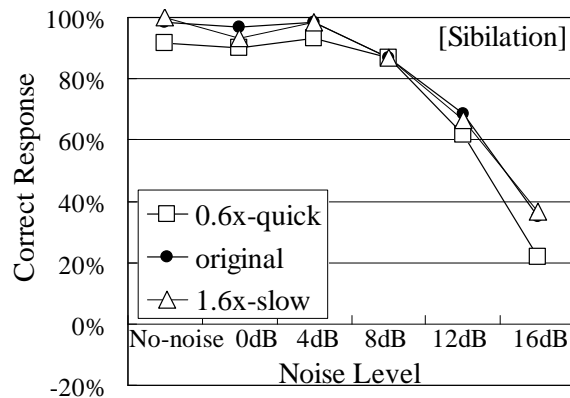


図 7. DRT 評価結果 [Sibilant]

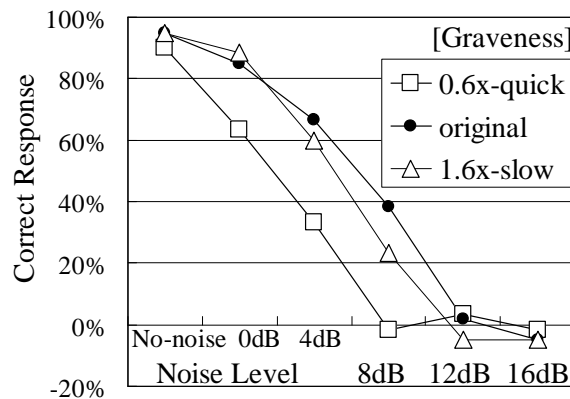


図 8. DRT 評価結果 [Graveness]

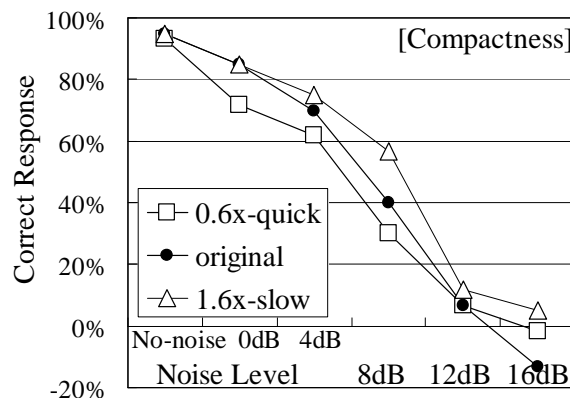


図 9. DRT 評価結果 [Compactness]

また、図 4 の Voicing と図 8 の Graveness、図 9 の Compactness において、ノイズレベル 4 dB～12 dB 付近の単語理解度を比較してみると、1.6x-slow は 0.6x-quick に対して、約 4 dB に相当する単語理解度の向上を示していることが分かった。このことから、話速は音声

表 3. 単語理解度試験結果のまとめ

音韻分類	1.6x-slowと originalの 理解度比較	0.6x-quickと originalの 理解度比較
Voicing	—	×
Nazality	—	—
Sustention	—	—
Sibilant	—	—
Graveness	—	×
Compactness	○	—

○：理解度の向上が見られる
—：理解度の違いはない
×：理解度が低下する

理解度に影響があることがわかる。

3.3 音韻特徴のまとめ

表 3 は 1.6x-slow と 0.6x-quick における単語理解度の影響を音韻特徴別にまとめたものである。この表から分かるように、話速を遅くすると理解度が向上する音韻、話速を早くすると理解度の低下がある音韻、話速の影響を受けにくい音韻、があることが分かった。

ここで、理解度を向上させるための話速変換の観点で観察してみる。Compactness は話速伸張において効果があると考えられる。Voicing と Graveness は話速を短縮すると理解度が低下する傾向が見られたが、一方では、Nazality, Sustention, Sibilant, Compactness においては、話速を短縮しても理解度に大きな低下が見られない。このことは、音韻特徴を捉えた話速変換について、理解度を向上させるために伸張することのみならず、理解度を損なわずに時間短縮することが可能になることをも示唆している。

4 むすび

話速変換による日本語音声の単語理解度の変化を、ゆっくり(1.6 倍の時間伸張)、早い(0.6 倍の時間短縮)、原音の 3 種類の話速において、マルチトーカーノイズを印加した状態における二者択一型の DRT 単語理解度試験で検証した。その結果、音韻に着目した話速変換が単語理解度に影響を及ぼすことが判った。

Voicing と Graveness では話速短縮で理解度が低下する傾向があり、Compactness では話

速伸張で了解度が向上しそうであること、他の音韻では話速の影響を受けにくそうであることなど、話速変換と了解度の関係がわかった。音韻特徴によって話速変換と単語了解度の関係に影響があることが示唆された。

また、ゆっくり (1.6 倍の時間伸張) において、Compactness 以外は了解度の向上が見られなかった。この原因は、今回使用した音源は、発話者がゆっくり、かつ、はっきり発話した音声だったことから違いが出にくかったと推定している。今後は、この原因についての解析や、音韻ごとの話速変換の効果などを確認する予定である。

謝辞

本研究にあたり度々の議論，ご支援を頂いた電気電子工学科の中川・近藤・高野研究室および情報科学科の小坂研究室の学生、職員一同に感謝する。

参考文献

- [1] 今井, “[招待論文] ゆっくり話せば話がはやい (話速変換技術)”, 信学技報 CQ2007-29 (2007-7)
- [2] 今井他, “話速変換に伴う時間伸張のリアルタイム吸収法”, 信学全大 D-694 (1999)
- [3] 今井他, “映像との同期を考慮した話速変換方式の一検討”, 信学技報 sp96-128 (1997-3)
- [4] 今井他, “ニュース音声を対象にした時間遅れを蓄積しない適応形話速変換方式”, 信学論文誌 Vol.J83-A、No8 (2008-8)
- [5] 沼畑他, “単語知覚における話者映像と時間伸張音声のずれの検知限および許容限”, 信学技報 HIP2007-139 (2007-12)
- [6] 安他, “高齢者を対象とした残響環境下における音声定常部抑圧処理の評価実験”, 信学技報 SP2004-154 (2005-03)
- [7] 安武他, “準実時間子音強調システム”, 信学技報 HP2005-94 (2005-12)
- [8] 近藤他, “二者択一型日本語音声了解度試験方法の検討” 日本音響学会誌, 63 巻, 196-205(2007)
- [9] Kondo, *et al.* “A Two-to-One Selection Based Japanese Speech Intelligibility Test and its Objective Estimation Methods,” *Advances in Communications and Media Research*, Nova Science Publishers, Hauppauge, NY (in print).
- [10] Fujimori, *et al.* “On a revised word-pair list for the Japanese intelligibility test,” *Proc. of the International Symposium on Frontiers in Speech and Hearing Research* (2006)
- [11] Kaga, *et al.* “Towards estimation of Japanese intelligibility scores using objective voice quality assessment measures,” *Proc. of the 4th Joint Meeting of the Acoust. Soc. of America and the Acoust. Soc. of Japan* (2006)
- [12] Takano, *et al.* “On Estimation of Two-to-One Selection-Based Intelligibility Score Using Speech Recognition,” *Proc. IEEE International Symposium on Consumer Electronics (ISCE2009)*, pp. 518-522, Kyoto, Japan, May 2009.
- [13] Kondo, *et al.* “Intelligibility Comparison of Japanese Speech with Competing Noise Spatialized in Real and Virtual Acoustic Environments,” *Acoust. Sci. & Tech.*, vol. 31 (2010)
- [14] Kitashima, *et al.* “Intelligibility of read Japanese words with competing noise in virtual acoustic space,” *Acoust. Sci. & Tech* (Jan. 2008).
- [15] Audacity: <http://audacity.sourceforge.net/>
- [16] Rice University: Signal Processing Information Base (SPIB), http://spib.rice.edu/spib/select_noise.html.