Vol.2016-MUS-113 No.12
2016/10/14

雑音下での少数サンプルによる了解度推定法

小林 洋介* 近藤 和弘** 坂本 修一***

*室蘭工業大学 大学院工学研究科 〒050-8585 北海道室蘭市水元町 27-1
**山形大学 大学院理工学研究科 〒992-8510 山形県米沢市城南 4-3-16
***東北大学 電気通信研究所/大学院情報科学研究科 〒980-8577 宮城県仙台市青葉区片平 2 丁目 1-1
E-mail: *ykobayashi@csse.muroran-it.ac.jp, **kkondo@yz.yamagata-u.ac.jp, ***saka@ais.riec.tohoku.ac.jp

あらまし 雑音下での音声了解度試験は多くの評価音を用いた被験者実験である。このため，被験者一人当たりの負荷が大きく安易に大規模な実験が行えないため，一人当たりの了解度評価単語数の削減が必要である。了解度を二項分布で表現すると，認知確率の等しい単語であれば 1 条件につき 1 単語で実験可能である。本稿では，20 単語による主観評価値を SNR と親密度を説明変数とするロジスティック回帰による推定関数を提案する。提案法は，全条件の平均二乗誤差で 0.068 と十分な性能を持つことが明らかとなった。

キーワード 音声了解度，了解度推定関数，単語親密度，雑音環境，ロジスティック関数

Speech intelligibility estimation method from few samples in noisy conditions.

Yosuke KOBAYASHI* Kazuhiro Kondo** and Shuichi SAKAMOTO***

* Graduate School of Engineering, Muroran Institute of Technology.
27-1 Mizumoto, Muroran, Hokkaido, 050-8585, Japan
**Graduate School of Science and Engineering, Yamagata University.
4-3-16 Jonan, Yonezawa, Yamagata, 992-8510, Japan
***Research Institute of Electrical Communication / Graduate School of Information Sciences, Tohoku University.
2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan
E-mail: *ykobayashi@csse.muroran-it.ac.jp, **kkondo@yz.yamagata-u.ac.jp, ***saka@ais.riec.tohoku.ac.jp

Abstract Assessment of speech intelligibility under noisy conditions requires experimental subjects to listen to and evaluate many sounds. Thus, the burden on subjects is significant, and carrying out large-scale experiments is not easy. For lowering evaluation costs, the number of words on intelligibility evaluation word lists need to be reduced. By using binomial distribution to express intelligibility, conducting experiments with one word per condition is possible for all words whose likelihood of cognition is the same. In this paper, we proposed an estimate method from results of subjective evaluation useing only 20 words. The ability of estimator to predict subjective assessment values resulted in the root mean squared error of 0.068 on the overall condition average.

Keywords Speech Intelligibility, Intelligibility Estimate Function, Word Familiarity, Noisy Condition, Logistic Regression

1. はじめに

音声システムの評価尺度の中でも音声明瞭度試験は古く，1929 年に Fletcher らにより明瞭度試験法が提案され[1]，その後に Egan による音素バランスリストが利用されてるようになった[2]。英語圏では語中にのみ発生する音素を考慮するため単語による評価に移行し，Fairbanks による CVC 型単語を用いた Rhyme test[3]，1958 年に House らによる 6 つの CVC 型単語から聴取した 1 単語を選択する MRT(Modified Rhyme Test) [4]が提案され，最終的に Voiers による 2 単語を提示し 1 つを選択する DRT(Diagnostic Rhyme Test)[5][6]へ発展した。これらの音素バランス文，MRT 及び DRT は ANSI S3.2-2009 として標準化されている[7]。

日本語での了解度試験法は，被験者の評価語に対する難易度や評価音源の訓練効果による再現性が議論された。難易度に関しては，音声の音響的な情報の他に意味情報などの心的辞書を用いていると仮定し，単語の認知難易度指標の一つである親密度[8]を統制した親密度別単語了解度試験用音声データセット 2003(以下，FW03)[9]と，FW03 の雑音下での了解度が 50%となる聴取域値 SRT(Speech Recognition Threshold)を統制した同データセット 2007（以下，FW07）[10]がある。

評価音源の学習効果に関してはDRTを日本語化し，高親密度語に統制した二者択一式日本語了解度試験法（Japanese DRT）[11]があり，被験者の訓練効果の少ない評価が可能になっている。さらに近年では，他のアジア諸語の了解度評価法が各国の研究者によって開発される(例として中国語[12]とタイ語[13]がある)など，音質評価指標としての世界的に利用されている。

これらの了解度試験法の評価対象は，当初の電話網評価のみならず，聴力検査や合成音声，ホール音響，屋外拡声器，スピーチマスキングシステム，補聴器フィッティングなど幅広い。一方で，技術基準に明記された場合を除いて音響機器開発の現場では，複数の被験者を長時間拘束することになる了解度試験は行い難い。このため，STI(Speech Transmission Index)[14]やAI(Articulation Index)及びその発展系であるSII(Speech Intelligibility Index)[15]などの物理量から了解度値を推定できる指標は検討されているが，主観評価を置き換えられていない。さらに，現在の了解度評価手法は，補聴器フィッティングのように個人の了解度を求めたい場合と，各種音響システムのように想定ユーザの平均値を求めたい場合とを区別しないため，全被験者の平均了解度を求めたい場合も被験者個々の正確な了解度評価を行う必要があり実験コストがかかる。

著者はこの問題を解決するために，各種音響システムの研究開発向けに平均了解度を小規模な主観評価より効率的に求める手法を提案し，その適応範囲を検証している[16]。提案法は，被験者ごとの了解度値を正確に求め，その値を回帰分析するのではなく，了解度を「聴こえた」と「聴こえない」の二項分布の確率分布とした際の各音場の単語聴取ができた確率とした場合の統計モデルを構築するのに必要最小限度の評価実験とし，ロジスティック回帰による推定を行う。

本稿では，提案法の概略を述べた後に，行った主観評価とその推定に物理量である SNR(Speech to Noise Ratio)のみを用いた 1 変数モデルと心理量である単語親密度を加えた 2 変数モデルの推定性能を比較する。

2. 提案推定法の概略[16]

2.1. 検討する実験系

了解度の評価対象は非常に広いが，その目的は各環境における推定了解度関数を主観評価値と物理量などの説明変数との間で回帰分析より作成することである。本稿の対象は，音響システム開発時における了解度試験であるため，被験者ごとの正確な推定了解度関数ではなく，全被験者の平均推定了解度関数を対象とする。特に主音声とその妨害を SNR(Signal to Noise Ratio)でコントロールできる実験系を扱う。

2.2. 従来の了解度推定

これまでの了解度推定には，詳細な実験計画によって得られた主観評価値へカーブフィッティングすることが多い。例えば，FW03 の主観評価による了解度を SNR と親密度の 2 変数よる式(1)の推定了解度関数導出では，SNR と親密度値 F の 2 変数で了解度を推定している。この式は，10 人の被験者によって SNR が-12，-9，-6 と-3 dB の FW03 の低親密度から高親密度までを全て主観評価を行った結果の回帰分析から得られている。

$$Int.(SNR, F) = \frac{1}{\exp(0.91 - 0.53F - 0.25SNR)} \quad (1)$$

この他の了解度推定法に，日本語 DRT による了解度評価値を音声認識結果から推定する例[17]，雑音抑圧処理後の音声の了解度を複数の音声品質尺度から求めた例[18]などがあるが，複雑な信号処理を伴い，国際標準となった STI[14]のように計測機器の販売がない限り，音響機器開発現場で利用されにくい問題がある。よって，簡易な主観評価と計算で式(1)の様な推定関数を求める手法が必要である。

2.3. 了解度試験の評価語数

了解度は評価法ごとの評価音リストの評価音を提示し，そのうち何語を正しく聴き取ったかを評価する。提示単語によらず実験結果は，被験者の回答テキストと提示テキストとの比較による正答か誤答かの 2 値であり，提示総数のうち正答した値をカウントした非負の整数データである。上限値が明らかなため，了解度は 1 から 0 までの聴き取れた割合，または聴き取ることができる確率とみなせる。上限値が明らかなカウントデータは二項分布で表現でき，多数のデータがある場合には中心極限定理により正規分布で近似できるが，本稿では少ない評価単語による主観評価を提案するため，二項分布のまま考えていく。ある SNR 値での了解度の確率分布 $p(C \mid T, q)$は，全被験者への提示単語総数を T, そのうちの正答総数を C, T 個の評価語それぞれの聴き取ることのできる確率を q として式 (2)で表せる。右辺は T 個のデータから C 個のデータを選び出す場合の数を用いた取りうる了解度値の全組み合わせである。

$$p(C \mid T, q) = \binom{T}{C} q^C (1-q)^{T-C} \quad (2)$$

ここで， q が評価語に依存しないとすると実験条件ごとに 1 評価語の利用で良い。しかし，同一の評価語を複数回用いる場合，単語の学習効果があるため，これを排除する必要がある。本稿で想定するのは，1 回あたりの評価における条件数が多い場合である。この時の学習効果は，被験者が実際に聴取する総数に依存するため，$p(C \mid T, q)$を評価条件毎に独立とすれば条件毎に 1 評価語とできる。聴取の難易度には親密度[8]が

影響することが明らかとなっており[9]，さらに親密度をコントロールしたFW07は，推定了解度50%であるSRTが一定になるように音圧補正されている[10]ため，FW07の評価単語は少なくとも了解度が50%になる条件においてはqが一定である。さらに高親密度単語リスト間ではSNRが同一の条件での了解度差が小さい[10]。よって，全てのSNR値においてFW07の高親密度単語は単語によらずqが一定と仮定できる。

本稿では，FW07の高親密度群より1リスト分の20単語で20条件を評価する了解度評価を提案し，この結果より了解度推定関数を導出する。

2.4. 推定関数の導出法

二項分布を用いた統計モデルにロジスティック回帰があり，二項分布と式(3)のロジットリンク関数を用いる。ロジットリンク関数と線形予測子が等しくなる条件で解くと，式(5)のロジスティック関数が得られ，主観評価値とそれに対応する説明変数xの観測値にあてはめ最尤推定することで回帰係数βが定まる。

$$\mathrm{logit}(q_i) = \log \frac{q_i}{1-q_i} \tag{3}$$

$$z_i = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \cdots + \beta_i x_{i-1} \tag{4}$$

$$\mathrm{logistic}(z_i) = \frac{1}{1+\exp(-z_i)} \tag{5}$$

本稿では，説明変数xに音場条件を設定するための物理量としてSNRをx_1とした1変数モデルと，式(1)と同様にSNRに加えて個々の単語の親密度値をx_2として加えた2変数モデルを比較する。

3. 評価実験

3.1. 主観評価の設定

FW07の本来の利用法である実験条件ごとに1評価リストを割り当てる主観評価（以下，リファレンス実験）と提案する推定関数の導出に必要な実験条件ごとに1単語とする主観評価を行った。評価条件を表1に示す。評価音声と妨害雑音はFW07に含まれる音源であり，付属の校正音源でSNRを計算した。

リファレンス実験はSNR値ごとにFW07の評価リストを一つ分の20単語の評価を行うが，実験規模を考慮して2 dB間隔とした。評価に用いる11リストはFW07付属の高親密度語リストからランダムに割り当てた。提案法は実験範囲のSNR値について1 dB間隔の20段階の評価で1セットだが，リストの影響を確認するため3リストをリファレンス実験に用いないリストから選択した。実験条件ごとの単語割り当ては全被験者で同一の音源としたが，リファレンス実験では女声話者と男性話者で条件ごとに割り当てるリストが異なるように設定した。提案法は，男女の話者で同一のリストを用いたが，SNR値ごとの割り当て単語は異なるように設定した。

表1　主観評価条件

条件	リファレンス	提案法
SNR	-20~0 dB (2 dB 間隔)	-19~0 dB (1 dB 間隔)
妨害雑音	スピーチノイズ	
発話者	女声(fhi), 男声(mis)	
親密度	高親密度単語	
リスト数	11	3(3 関数分)
被験者数	10 名(共通)	

主観評価は室蘭工大に設置した防音ブース内にて被験者10名(全員が20代前半の学生)で行った。評価音は，ラップトップコンピュータに接続したオーディオインタフェース(Roland, UA-25EX)からヘッドホン(Sennheiser, HDA-300)を用いて全評価音源を被験者へランダムにダイオティック提示した。被験者は，専用GUIを用いてカタカナで回答入力した。本実験の前に別の音源を用いてGUIの操作練習を行った。評価音の提示レベルは，校正音源を用いた予備実験により決定し，全員で共通とした。回答の正誤判定はFW07付属のテキストとのマッチングで行ったが，二重母音部の入力に長音を使用した場合および同音異字はどちらも正解とした。

3.2. 推定関数の性能評価指標

提案法の予測精度の導出には，式(6)のリファレンス実験による了解度値とそれぞれの推定関数による予測了解度値とのRMSE(Root mean squared error)で評価する。式の分母はリファレンス実験で評価した総数であり，nは実験条件（SNR及び親密度値），$sub.(n)$はリファレンス実験による了解度値，$pred.(n)$は推定関数を用いた予測値である。

$$\mathrm{RMSE} = \sqrt{\frac{\sum (sub.(n) - pred.(n))^2}{total\ number\ of\ responces}} \tag{6}$$

RMSEに加えて，作成した推定関数のSRTを求め，リファレンス実験値にカーブフィッティングしたSRTとの差を求める。SRTはZ_i=0の値なので，1変数モデルではSRTを式(7)で求め，2変数モデルでは式(8)とし，親密度x_2は各推定関数の作成に用いたサンプルの平均値とした場合の値を比較に用いる。

$$\mathrm{SRT} = -\frac{\beta_1}{\beta_2} \tag{7}$$

$$\mathrm{SRT} = -\frac{\beta_1 + \beta_3 x_2}{\beta_2} \tag{8}$$

3.3. リファレンス実験の結果

図1にリファレンス実験の結果を男女別に示す。図中の曲線は，SNRのみの1変数によるカーブフィッティングによる推定関数である。また，図の推定関数のRMSEとSRTを表2に示す。結果より，特に-14 dBで

男女差が 0.19 と女声の了解度が高く，次いで，-8 dB においては男女差が 0.09 あるが，こちらは男声の方が高い。この影響で，関数の立ち上がりは男性の方がわずかに急峻になっているが，SRT は 0.2 dB しか差がなく，FW07 のターゲットである，音源リストによらず安定した推定了解度を得ることができている。

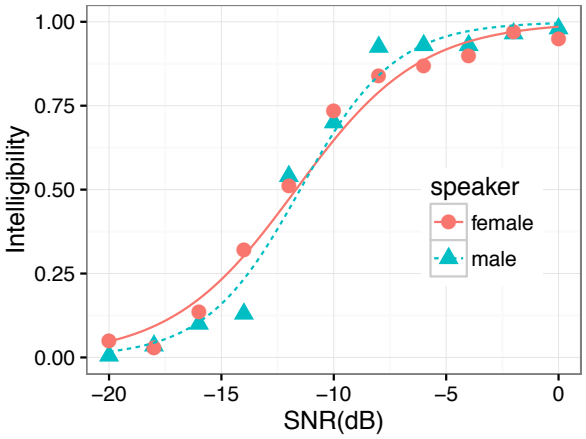

図 1　リファレンス実験の結果

表 2 リファレンス実験による 1 変数推定関数の精度

	female	male
RMSE	0.043	0.054
SRT (dB)	-11.7	-11.5

3.4. 1 変数モデルによる提案法の推定性能

図 2 に 1 変数モデルによる提案法の推定関数の例として，list A を用いた女声の推定関数と導出に用いた主観評価値のバルーンプロットを示す。図でバルーン上に示した主観評価値は正答を 1 に，誤答を 0 とし，それぞれの被験者数をバルーンの大きさとした。黒ドットはリファレンス実験による女声の主観評価値であり，この値と推定関数との差の平均が RMSE になる。図より，20 単語と少数の実験値から求めた推定関数であってもリファレンス実験の主観評価値の傾向をつかんでおり，十分推定できている。

作成した 3 リスト全ての推定関数を図 4 と 5 に女声と男声に分けて示す。図中の ref. は図 1 に示したリファレンス実験の値を用いた従来法による推定関数である。リストごとの RMSE と SRT を表 3 に示す。F は女声で M は男声の結果を示す。全体として女声の方が RMSE は小さいが，リスト C のみ男声の方が 0.019 ポイント小さい。また，SRT は全 6 条件の平均値を求めると-11.47 dB とリファレンス実験に近い値となるが，最大値と最小値の差は 1.9 dB ある。これら RMSE と SRT のばらつきについては，使用したリストに含まれる単語個々の認知難易度が影響していると考えられる。

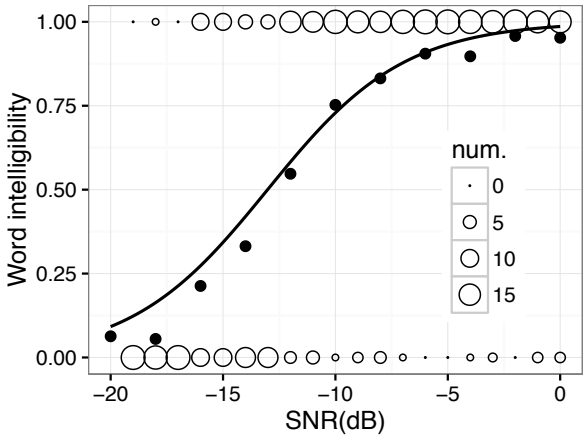

図 2　1 変数モデルの例（女声 list A）

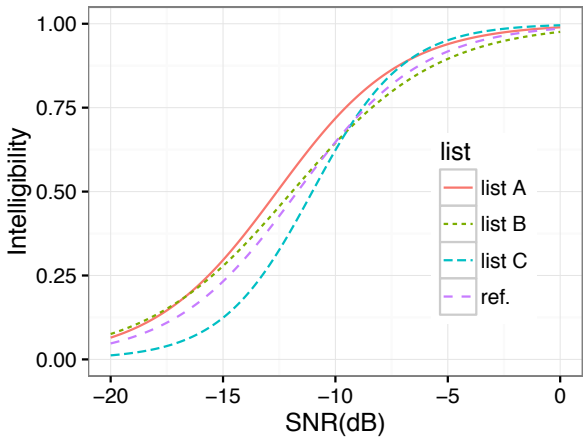

図 3　1 変数モデルによる推定関数（女声）

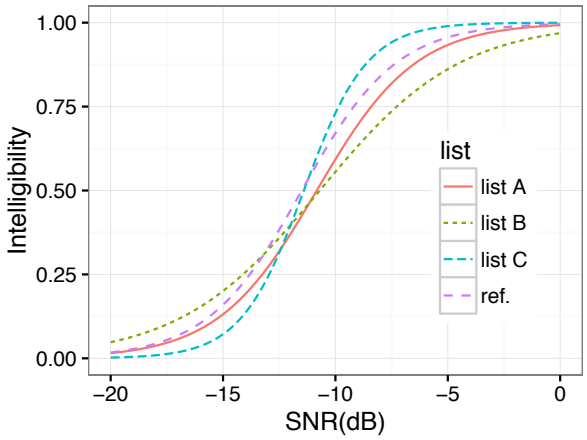

図 4　1 変数モデルによる推定関数（男声）

表 3　提案法 1 変数モデルの結果

	List A		List B		List C	
	F	M	F	M	F	M
RMSE	0.052	0.078	0.055	0.108	0.075	0.056
SRT (dB)	-12.6	-10.8	-11.9	-10.7	-11.0	-11.4

3.5. 2 変数モデルによる提案法の推定性能

1 変数モデルでは，提案法に用いるリストの影響がみられた。理由として，提案法は 1 条件に 1 単語のため，同一条件での推定関数の差は，単語の認知が統制できていないことが原因と考えられる。そこで，FW07 の設計に利用した単語ごとの認知難易度である親密度値を説明変数に加えた 2 変数モデルによって推定性能が向上するか比較する。推定関数の作成に用いる了解度と SNR 値は 1 変数モデルと同じ値を用いる。

図 5 に 2 変数モデルの関数例を示す。親密度値によって SNR と了解度の傾きが変化しており，親密度値が下がるにつれ，シグモイド関数の立ち上がる SNR の値が高い値にシフトしている。これは親密度が低くなるほど，同一の SNR では聴き取りにくくなることを示している。図 6，7 に女声と男声の親密度値 5.50 で切り出した推定了解度と SNR の関数をリストごとに示す。ここで用いた親密度値 5.50 は，高親密度と中高親密度の閾値であり，関数の形状差が出やすい値として採用した。図中の ref. は図 1 の結果の男女別の再掲である。結果より特定の親密度値で切り出した場合は，女声のリスト C，男声のリスト B のように図 3，4 と傾きが大きく異なるリストとあまり変化のないリストがある。

リストごとの RMSE と平均親密度値の SRT を表 4 に示す。1 変数モデルの結果と同様に，女声の RMSE が低く，男性は若干高めである。すべての条件で RMSE が改善しており，最も変化の大きいリスト C は女声が 0.007 ポイント，男声が 0.003 ポイント改善した。また，SRT は 1 変数モデルとほぼ同値で，最大値と最小値の差が 1.7 dB と 1 変数モデルよりも 0.2 dB 改善した。

表 4　提案法 2 変数モデルの結果

	List A		List B		List C	
	F	M	F	M	F	M
RMSE	0.049	0.075	0.055	0.107	0.068	0.054
SRT (dB)	-12.5	-10.9	-11.9	-10.8	-11.2	-11.4

4. 考察

4.1. 1 変数モデルと 2 変数モデルの差

表 5 に変数モデル別の平均 RMSE を男女別及び男女混合の平均に分けて示す。2 変数にすることで男女平均は 0.003 改善したが，その差はわずかであり，ターゲットとする開発現場での利用を考慮すると単純な解析が行える 1 変数モデルでも十分目的を達成できる場合が多いと考えられる。1 変数で十分であった理由として，SNR は主観評価のコントロール条件であり，変化幅が大きいのに対し，親密度値は評価単語を決定する際(2.3 節)に，高親密度単語のみに限定したため，説明変数として有効ではなかった可能性がある。

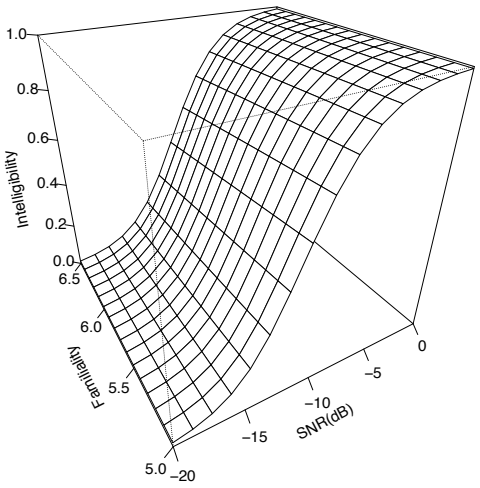

図 5　2 変数モデルの例（男声 list A）

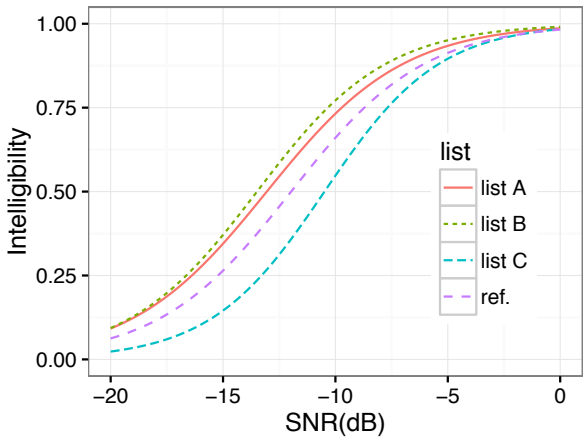

図 6　2 変数モデルによる推定関数
（女声-親密度値 5.50）

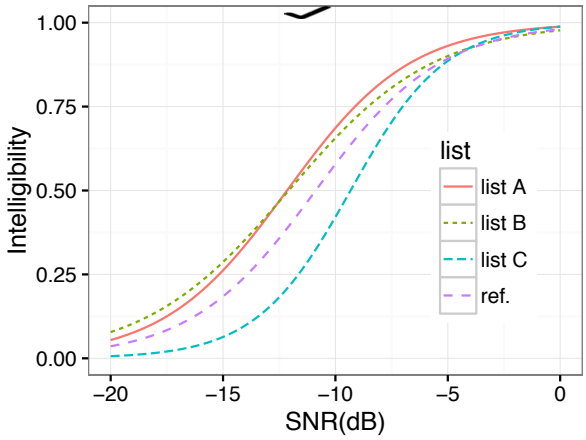

図 7　2 変数モデルによる推定関数
（男声-親密度値 5.50）

表 5　モデル別の平均 RMSE

	1 変数モデル		2 変数モデル	
	F	M	F	M
平均 RMSE	0.061	0.081	0.057	0.078
	0.071		0.068	

4.2. 提案法の話者差について

提案法は 1 変数モデル，2 変数モデル共に発話者差がある。図 8 に 1 変数モデルにおける SRT の男女差が最も大きいリスト A の結果を再掲する。図 1 のリファレンス実験の結果と比べると低 SNR における男声の了解度が低い点は共通するが，高い範囲での男女の入れ替わりが起きていない。リスト A は，図 3 より女声はリファレンスより了解度が高く，図 4 の男声はリファレンスより了解度が低くなっており，実験 SNR 値への単語の割り当て法を再検討する必要がある。

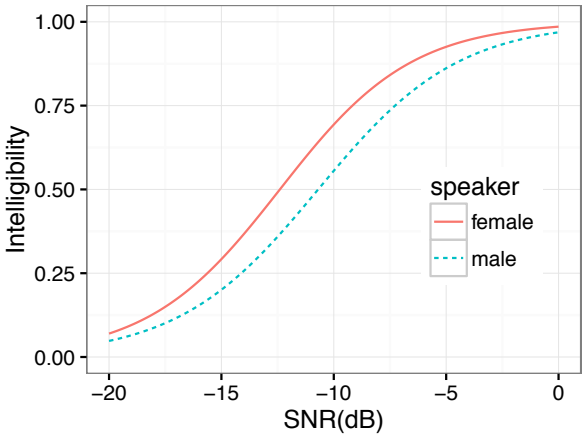

図 8 リスト A における 1 変数モデルの男女差(再掲)

5. まとめ

音声了解度の推定関数を効率的に求めるため，SNR 値ごとに 1 単語まで評価単語を極端に減らした主観評価結果のロジスティック回帰による了解度推定関数を SNR のみの 1 変数モデルと親密度値も加えた 2 変数モデルを作成し，その性能を比較した。その結果，2 変数モデルでは，平均で 0.003 ポイント RMSE が改善するが，提案手法の利用対象となる音響機器開発現場の統計解析を煩雑にすることを考えると 1 変数モデルでも十分な性能である。しかし，使用リストの影響により SRT は最大で 1.9 dB の差があり，利用単語の選定法を検討する必要がある。加えて，提案法の実システムの開発への応用のため，音声に伝送特性を畳み込んだ場合などの推定性能を評価する予定である。

謝辞

本研究の一部は，JSPS 科研費(16K21584)，(公財)人工知能研究振興財団 平成 27 年度研究助成，(公財)電気通信普及財団 研究調査助成及び東北大学電気通信研究所共同研究プロジェクト（H26/A14）の助成を受け実施した。関係者と被験者各位に感謝する。

文　献

[1] H. Fletcher and J.C. Steinberg, “Articulation Testing Methods,” Bell System Technical Journal, vol. 8, Issue 4, pp.806-854, October 1929

[2] J.P. Egan, “Articulation testing methods,”' Laryngoscope, vol. 58, pp. 955-991, 1948.

[3] G.Fairbanks, “Test of phonetic differentiation: The rhyme test,” J. Acoust. Soc. Am., vol. 30, no. 7, pp.596-600, 1958.

[4] A.S. House, C.E. Williams, M.H.L. Hecker, and K.D. Kryter, “Articulation testing methods: Consonantal differentiation with a closed-response set,” J. Acoust. Soc. Am., vol. 37, pp.158-166, 1965.

[5] W.D. Voiers, “Diagnostic Evaluation of Speech Intelligibility, in Speech Intelligibility and Speaker Recognition,” M.E.Hawley,ed., Dowden, Hutchinson & Ross, PA, 1977.

[6] W.D. Voiers, “Evaluating Processed Speech using the Diagnostic Rhyme Test,” Speech Tech., vol. 1, pp.30-39, 1983.

[7] ANSI, “Method for Measuring the Intelligibility of Speech Over Communication Systems,” Technical Report S3.2-2009(R2014), American National Standards Institute, 1989, reaffirmed 1995, 1999, 2009 and 2014.

[8] 天野成昭，近藤公久，日本語の語彙特性 第 1 巻 単語親密度，三省堂，1999.

[9] 坂本修一，鈴木陽一，天野成昭，小澤賢司，近藤公久，曽根敏夫，“親密度と音韻バランスを考慮した単語了解度試験用リストの構築，” 日本音響学会誌, vol. 54(12), pp.842-849, 1998.

[10] 近藤公久，天野成昭，坂本修一，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)の作成，” 信学技報 TL2007-62, pp.43-48, Jan. 2008.

[11] 近藤和弘，泉良，藤森雅也，加賀類，中川清司，“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 4 号, pp.196-205, Apr. 2007.

[12] Ian McLoughlin, “Subjective intelligibility testing of Chinese speech,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, Issue 1 pp.23-33, 2008.

[13] N. Saimai, C. Tantibundhit, C. Onsuwan and C. Wutiwiwatchai, “The roles of temporal envelope and temporal fine structure in speech synthesis for cochlear implants for tonal language speakers,” Proc. ICA 2013, vol. 19, 050074, 2013.

[14] IEC60268-16, “Sound system equipment --- Part 16: Objective rating of speech intelligibility by speech transmission index,” 2011.

[15] ANSI, “Methods for the Calculation of the Speech Intelligibility Index,” Technical Report ANSI S3.5–1997(R2012), American National Standards Institute, 1997, reaffirmed 2007 and 2012.

[16] 小林洋介，“雑音環境下での少数音声による推定了解度関数導出法，” 日本音響学会誌, 72 巻 6 号, pp.319-321, Jun. 2016.

[17] Y.Takano and K.Kondo, “Estimation of Speech Intelligibility Using Speech Recognition Systems,” IEICE Trans. Inf. & Syst., vol. E93-D, no. 12, pp.3368-3376, Dec. 2010.

[18] Y. Hu and P.C. Loizou, “Evaluation of objective quality measures for speech enhancement,” IEEE Transactions on Speech and Audio Processing, 16(1), pp.229-238, Jan. 2008.