

## 信号対雑音比調整による単語リスト間の単語了解度差補正 ——親密度別単語了解度試験用音声データセット (FW07) を用いた検証——\*

近藤 公久<sup>\*1,†</sup> 坂本 修一<sup>\*2</sup> 天野 成昭<sup>\*1,††</sup> 鈴木 陽一<sup>\*2</sup>

【要旨】 親密度別単語了解度試験用音声データセット (FW03) [NII 音声資源コンソーシアム, 2006] は、日常の音声聴取能力を把握する目的で作製された。FW03 の四つの親密度ランクには 50 単語からなるリストが 20 リスト存在し、各親密度ランク内で音韻バランスがとられている。しかし、この FW03 を臨床的な場面に応用するためには、1 リストの単語数を縮減する必要があった。そこで 1 リストの単語数を 50 語から 20 語に縮減した FW07 が作製された [NII 音声資源コンソーシアム, 2007]。本稿では、単語数縮減の影響の検討と、試験語ごとの信号対雑音比 (SN 比) 調整による単語リスト間の了解度差補正効果の検証を行う。

キーワード 単語了解度, 認知閾, リスト間差, 聴力検査, 雑音下

Word intelligibility, SRT, List-difference, Hearing test, Noisy condition

### 1. はじめに

現在一般に実施される聴力検査は、聴力計 (audiometer) による純音聴力レベルを基本とし、音声レベルでの検査では、57-S [1] や 67-S [2] による単音節明瞭度が用いられることが多い。英語圏では、単語リスト (例えば PB) [3] を用いた単語了解度試験が従来から標準的に用いられている。また、雑音下における言語音声を用いた検査も広く用いられており、その手続き (例えば HINT) やそのための文リスト [4] が提案されている。日本においても補聴器適合評価用 CD (TY-89) [5] では単語リストと大沼ら [6] による文音声収録されている。しかし、TY-89 の単語リストは 2 音節及び 3 音節の単語が各 50 語しか存在しないため、異なる条件での単語了解度を比較するためには同じ単語の繰り返しが必要になる。また、リストに存在する単語は子供でも理解可能な容易な単語を用いており、単語の難易度による違いの検討は難しい。

Shiroma, Iwaki, Kubo, & Soli [7] により、雑音下における文章了解度検査のための文リスト (JHINT)

が作製され、JHINT は英語の HINT と同じ特性を持つことが示されている。JHINT には全部で 10 文からなる 24 リスト (計 240 文) が存在し、文の自然性や難易度及び音韻分布が考慮されている。しかし、HINT (JHINT を含む) に用いられている文の難易度は一定の範囲で同一になるように統制されており、難易度の違いによる影響を検討することは意図していない。

親密度別単語了解度試験用音声データセット (FW03) [8] は、単語のなじみの程度を表す「単語親密度 (音声提示)」[9] を用いて試験語の難易度が統制された単語リストである。単語親密度はそれぞれの単語に対する「なじみの程度」の主観的評定値 (評定者は 18 歳から 30 歳 (平均 24.3 歳)) であり、1-7 までの 7 段階評定値 (1 が最もなじみがない、7 が最もなじみがある) の平均値である。特定の文書に単語が使用される頻度 (単語出現頻度) も単語の難易度を表す指標として利用可能であるが、音声提示による単語親密度の方が日常生活場面における音声聴取能力を評価するためにはよりよい指標であると考えられる。

FW03 は、四つの親密度ランクにおいて 50 語のリスト 20 リストから構成されている。また、単語親密度別の各リストに含まれる単語の語頭及び語中の母音-子音連鎖の出現確率によるエントロピーを最大にすることで音韻バランスがとられている [10, 11]。この音韻バランスによって各リストの単語が含む音韻種の偏りは最小になっていると仮定できるため、同一親密度ランク内のリストに対する単語了解度は、ある程度均一化されると期待された。しかし、Amano, Sakamoto, Kondo, Suzuki [12] の単語了解度試験結果から、同一親密度ランク内のリスト間の平均了解度に統計的に有

\* Compensation for list-difference of word intelligibility by conditioning signal-to-noise ratio: Validation by using the familiarity-controlled word lists 2007 (FW07), by Tadahisa Kondo, Shuichi Sakamoto, Shigeaki Amano and Yōiti Suzuki.

\*1 日本電信電話(株) NTT コミュニケーション科学基礎研究所

\*2 東北大学電気通信研究所/大学院情報科学研究科

† 現在, 国際電気通信基礎技術研究所

†† 現在, 愛知淑徳大学

(問合せ先: 近藤公久 e-mail: tkondo@atr.jp)

(2012年1月31日受付, 2012年10月27日採録決定)

意な差があることが指摘された。この結果は、20 リストのうちのいずれかをそれぞれの聴取条件に割り当てた場合に、得られる単語理解度の差が条件間の差であるかリスト間の差であるかの区別が難しくなることを意味する。

一方で、Amanoら [12] は、FW03 を用いた健聴者による単語理解度試験結果から、異なる親密度ランク間の単語理解度試験結果を比較することで語彙性 (lexical information) の影響を考慮して単語理解度試験が可能であることを主張した。このことは、FW03 が日常生活における様々な場面の音声聴取能力を評価可能であることを意味している。しかし、このような日常生活における音声聴取能力を評価する場合には、補聴器のフィッティングや公共空間の音響設計の基礎データとして用いる場合など、検査に要する時間や負荷が問題となる場面が多い。特に、検査対象者が難聴者や高齢者を含む場合には負荷の問題は解決されなければならない課題である。FW03 では1 リスト 50 語で構成されており、条件の組み合わせの数  $\times$  50 語を聴取しなければならない。つまり、様々な条件下で試験を行う必要がある場合において1 リストの単語数の影響は検査に必要な条件が多ければ多いほど大きい。

そこで、各単語リストの単語数を20 単語にし、FW03 と同様に音韻バランスをとった FW07 が作製された [13]。また、FW07 では種々の雑音レベルにおける各試験語に対する正答率 (正答者数/聴取者総数) の50%点を各単語の聞こえ易さの基準 (認知閾値: Speech Recognition Threshold for a Word, 本稿では  $SRT_w$  と呼ぶ) として提示音圧を変える (以後、提示音圧調整と呼ぶ) ことによってリスト間の理解度差を補正する方法が提案されている。本稿では、FW07 を用いて、1 リストを20 単語にしたことによるリスト間差への影響と提示音圧調整によるリスト間差の補正効果を検討する。

## 2. FW07 の特性解析

### 2.1 FW07 リスト

FW03 の各親密度ランクに存在する50 語のリスト20 リスト、すなわち、1,000 語を用いて FW07 リスト (20 語のリストを20 リスト、計400 語) が作製された。その手順は以下のとおりである。なお、FW03 単語リストは、4 モーラ、0 型アクセント、四つの親密度ランク (low: 1.0–2.5, lower middle: 2.5–4.0, upper middle: 4.0–5.5, high: 5.5–7.0) から成る。従って、FW07 もこれらの特性を踏襲している。

#### STEP1: $SRT_w$ 推定

$SRT_w$  の推定には、Amanoら [12] が FW03 を用いて実施した単語理解度試験の結果を用いた。Amano

らの単語理解度試験では、聴取者男女32名 (20歳から37歳、平均27.1歳、全員健聴者) が参加した。FW03 には話者4名による50 単語  $\times$  20 リスト  $\times$  4 単語親密度ランクの4,000 語の音声収録されており、聴取者の内半数が話者 fhi と mis, 半数が fto と mya の音声を聴取した (話者名の先頭 f は female: 女声, m は male: 男声を現す)。聴取雑音条件は、親密度ランクによって異なり、親密度ランク low の単語リストに対しては  $-12$  から  $6$  dB (S/N), 親密度ランク lower middle に対しては  $-15$  から  $3$  dB (S/N), 親密度ランク upper middle と high に対しては  $-18$  から  $0$  dB (S/N) の間を  $3$  dB 間隔とした7種類であった。

この実験によって得られた結果を、各話者の各単語ごとに式 (1) で表される累積正規分布関数によってフィッティングした。フィッティングによって得られた回帰式上で、各単語に対し聴取者の50%が正答するSN比をその単語の  $SRT_w$  とした。すなわち、式 (1) において、 $SRT_w$  は  $p(x) = 0.5$  を満たす  $x$  であり、 $\mu$  に等しい。

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \quad (1)$$

#### STEP2: 対象単語の絞込み

FW07 リスト作製にあたっては、同じ親密度ランク内で特異的な特徴を示す単語は用いていない。すなわち、推定された  $SRT_w$  が95%信頼区間に入らない単語は対象外とされている。具体的には、各親密度ランクに含まれる1,000 単語のうち、low: 222 単語, lower middle: 101 単語, higher middle: 40 単語, high: 60 単語を対象外とした。また、FW03 作製時に使用した親密度データベースは  $\beta$  版であったため、正式版を用いたときに親密度ランクが変わってしまう5 単語 (low: 3 単語, high: 2 単語) は除外している。

表-1に、話者ごと、親密度ランクごとの単語 (400 語) の  $SRT_w$  の平均と標準偏差を示す。表-1 から  $SRT_w$  には話者間で平均値に異なりがある (最大で  $3$  dB 程度) が、この違いは、話者、男声/女声、あるいは、本実験では話者の条件は聴取者間要因であるので聴取者、のいずれに起因するものかは正確には区別できない。また、話者 fto の音声が他に比較して理解度が高いことがうかがえる。比較のため、FW03 における話者ごと、親密度ランクごとの単語 (1,000 語) の  $SRT_w$  の平均と標準偏差を示す。ただし、 $SRT_w$  が  $\pm 30$  を超えるものは異常値として計算から除外した。

#### STEP3: 音韻バランス

STEP2 で候補に残った単語を用いて FW03 作製時と同じ方法 (式 (2) を最大化) により音韻バランスの

表-1 話者ごと、親密度ランクごとの平均 SRT<sub>w</sub> 及び標準偏差 (ただし, SRT<sub>w</sub> が ±30 の範囲の単語のみ)

Narrator	Familiarity rank			
	low	lower middle	higher middle	high
Words in FW07				
fhi	-2.71 (N=400, sd=4.11)	-4.53 (N=400, sd=3.50)	-7.80 (N=400, sd=2.78)	-9.79 (N=400, sd=2.46)
fto	-5.49 (N=400, sd=3.01)	-7.35 (N=400, sd=3.08)	-9.58 (N=400, sd=2.39)	-10.97 (N=400, sd=2.21)
mis	-2.54 (N=400, sd=3.74)	-4.39 (N=400, sd=3.16)	-6.93 (N=400, sd=2.47)	-8.76 (N=400, sd=2.02)
mya	-2.87 (N=400, sd=3.91)	-4.86 (N=400, sd=3.42)	-7.41 (N=400, sd=2.59)	-8.64 (N=400, sd=2.58)
Words in FW03				
fhi	-1.66 (N=956, sd=7.14)	-4.06 (N=978, sd=6.50)	-7.57 (N=995, sd=5.33)	-9.56 (N=996, sd=4.94)
fto	-4.77 (N=964, sd=6.28)	-7.07 (N=989, sd=5.72)	-9.38 (N=990, sd=4.50)	-10.76 (N=984, sd=4.11)
mis	-1.49 (N=967, sd=6.76)	-4.23 (N=985, sd=6.17)	-6.93 (N=994, sd=4.92)	-8.65 (N=1,000, sd=4.10)
mya	-2.19 (N=975, sd=7.06)	-4.50 (N=989, sd=6.40)	-7.27 (N=998, sd=5.10)	-8.47 (N=996, sd=4.56)

表-2 親密度ランクごとのエントロピー ( $H_{total}$ )

list	Familiarity rank			
	low	lower middle	higher middle	high
FW03 (50 words)	183.4	187.6	186.8	182.2
FW07 (20 words)	154.0	156.8	156.8	154.2

とれた単語リスト (20 語の 20 リスト) が作製された。

表-2 に親密度ランクごとの 50 語リスト及び 20 語リストのエントロピー (式 (2)) を示す。表-2 から 20 語リストの方がエントロピーが全般的に低いことが分かる。しかし, このエントロピーの絶対値が低下していることと, リスト間での音韻の偏りが大きくなることとは必ずしも対応しない。なぜならばエントロピーの定義からリスト内の語数が減れば, とりうる最大エントロピーが小さくなるからである。

$$H_{total} = H_1 + H_2 \quad (2)$$

$$H_1 = - \sum_{l=1}^{20} \sum_m p(m) \log_2 p(m)$$

$$H_2 = - \sum_{l=1}^{20} \sum_v \sum_c p(v)p(c|v) \log_2 p(c|v)$$

$p(m)$ : 先頭モーラ  $m$  のリスト内での出現確率

$p(v)$ : 母音  $v$  のリスト内での出現確率

$p(c|v)$ : 先行母音が  $v$  のときの後続子音が  $c$  であるリスト内での条件付出現確率

### 2.2 平均単語理解度

FW03 に対して実施された単語理解度試験 [12] の結果から, すべての単語を用いたときの各リスト (50 語) の平均値 (図-1), 及び, FW07 リストに採用された単語による各リスト (20 語) の平均値 (図-2) を話者 mya を例に示した。なお, 図中の水平線は 50%点,

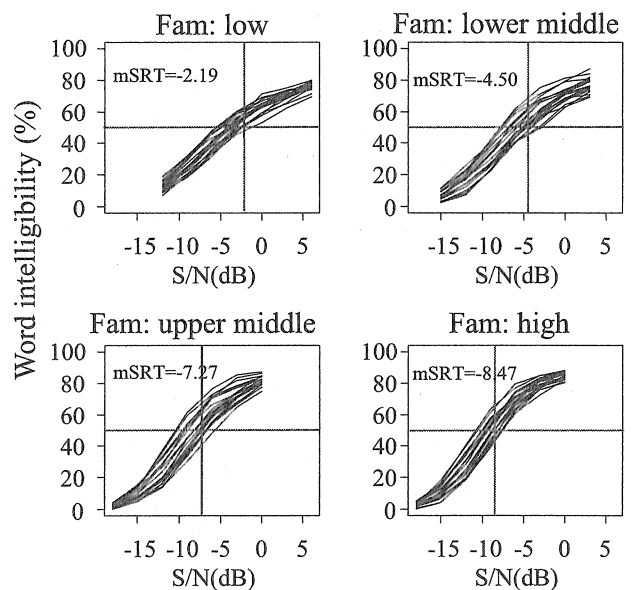


図-1 単語理解度 (50 単語リスト) \*話者 mya

垂直線は平均 SRT<sub>w</sub> を示している。図-1 と図-2 から FW03 の各親密度ランクに存在する 50 語のリスト 20 リスト及び選択された 20 語のリスト 20 リストの平均単語理解度にはリスト間でバラつきがある。しかし, そのバラつきは, 20 語リストにした影響で必ず大きくなるとは限らないことが分かる。

### 2.3 単語ごとの SRT<sub>w</sub> による提示音圧調整

同じ親密度ランクのリスト間で平均単語理解度に差があることは図-1 及び図-2 で示したとおりである。FW07 [13] では, 累積正規分布関数 (式 (1)) によるフィッティングパラメータ  $\mu$  ( $= SRT_w$ ) を用いて提示音圧を調整することによりリスト間の単語理解度を揃える方法が提案されている。以降, 各単語  $w_i$  に対して得られた推定パラメータを,  $\mu_i, \sigma_i$  で表して, FW07 の提示音圧調整法について示す。また, それらの平均値をそれぞれ,  $\mu_m, \sigma_m$  とし, 式 (1) において,  $\mu = \mu_m,$

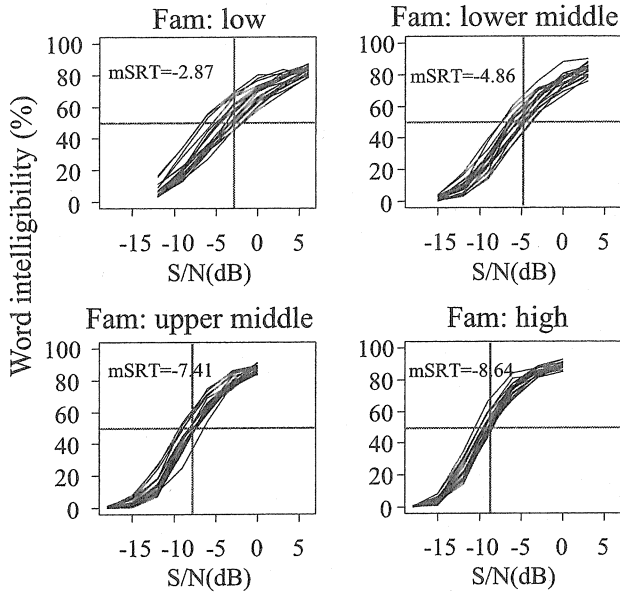


図-2 単語理解度 (20 単語リスト) \*話者 mya

$\sigma = \sigma_m$  としたとき得られる関数で描かれる曲線を推定平均単語理解度曲線と呼ぶこととし、これを基準として提示音圧調整量を決定する。

式 (1) から、基準となる推定平均単語理解度曲線の  $x = sn_0$  のときの値  $p_m(sn_0)$  [ $\mu = \mu_m, \sigma = \sigma_m$ ] と、ある単語  $w_i$  に対する単語理解度の推定値  $p_i(sn)$  [ $\mu = \mu_i, \sigma = \sigma_i$ ] とが等しいとき、 $z$  変換式から、

$$\frac{sn - \mu_i}{\sigma_i} = \frac{sn_0 - \mu_m}{\sigma_m} \quad (3)$$

の関係が成り立つ。式 (3) を変換すると、 $sn$  は、

$$sn = \mu_i + \frac{\sigma_i}{\sigma_m}(sn_0 - \mu_m) \quad (4)$$

と表せる。すなわち、ある単語  $w_i$  の単語理解度を、推定平均単語理解度曲線の  $sn_0$  のときの単語理解度に一致させたい場合、式 (4) を満たす  $sn$  で提示すればよいことになる。

しかしながら、式 (4) は基準となる  $sn_0$  によって調整量が異なるため複雑である。そこで FW07 では、 $\sigma_i = \sigma_m$  であると仮定して簡略化している (式 (5))。

$$sn = sn_0 + (\mu_i - \mu_m) \quad (5)$$

$SRT_w$  の定義から、 $SRT_w$  が大きい単語は大きな音圧で提示しないと正答率が 50% に到達しないことを意味する。式 (5) は、 $SRT_w$  の平均値からの差分だけ提示音圧を増減すればいいことを示している。FW07 では事前に  $SRT_w$  の平均から差が 95% 信頼区間外である単語は対象から除外されているため、各単語に対する提示音圧調整量はこの範囲に収まる。

### 3. 提示音圧調整後の単語理解度

#### 3.1 単語理解度試験

##### 3.1.1 刺激

親密度：親密度ランクは FW07 に存在する 4 ランク (low, lower middle, higher middle, high) であった。

音声：FW07 収録の女声 2 (fhi, fto), 男声 2 (mis, mya) の合計 4 話者の音声ファイルを音声刺激とした。

雑音条件：雑音は、FW07 [13] に収録されている雑音を用いた。この雑音は、白雑音を FIR デジタルフィルタを通して ITU-T G.227 勧告に示される特性を持たせたものである。

雑音は立ち上がりと立ち下がりに 50 ms の傾斜 (直線) をつけ、雑音が完全に立ち上がってから 300 ms 後に音声を開始し、音声終了後 200 ms で立ち下がりが開始するように FW03 収録の単語音声ファイルに重畳した。

雑音の音圧は等価騒音レベル ( $L_{Aeq}$ ) で FW07 収録の較正用 1,000 Hz 純音と同じ音圧で一定とし、単語の音圧をそれより 6 dB 低くして SN 比を  $-6$  dB とした。SN 比を  $-6$  dB としたのは、図-2 から、すべての親密度ランクにおいて理解度が 0 又は 100 (%) にならないレベルを選択したからである。

提示音圧：実際の提示音圧は、FW07 収録の較正用 1,000 Hz 純音を再生したときに、ヘッドホンカップラ (IEC711Coupler, BK4157) にコンデンサマイクロホン (BK41341/2inch マイクロホン) を装着しメジャリングアンプ (B&K2636) によって計測した音圧レベルが 60 dB となるように調整した。ただし、音声の音圧は、2.3 節で示した FW07 の音圧調整法と同じ  $SRT_w$  による音圧調整を行った。

提示装置：刺激は汎用のパーソナルコンピュータ (IBM R50e) から USB 接続された DA 変換器 (Sound Blaster Audigy 2 NX) により DA 変換され、ヘッドホン (Zennheizer HDA200) により各聴取者の左耳に提示された。

##### 3.1.2 聴取者

聴取者は成人健聴者男女 12 名、合計 24 名、平均年齢は 29.3 歳 ( $sd = 6.0, \min = 20, \max = 39$ ) であった。聴取者は全員、単語理解度試験 [12] とは異なる。なお聴取者のオーディオメータで測定した聴力は、全員 17.5 dBHL より良好であり、健常の範囲であった。本実験の聴取者は、 $SRT_w$  の推定に用いた Amano ら [12] の単語理解度試験聴取者とほぼ一致する (男女 32 名、20 歳から 37 歳、平均 27.1 歳、全員健聴者)。

##### 3.1.3 手順

聴取者は、各単語を聞いた後、パーソナルコンピュータのキーボードにより各自で聞こえたと思う単語をカ

表-3 各話者, 各親密度ランクごとのリストの平均単語理解度 (%), 標準偏差, 最小値, 最大値, 最小最大幅。イタリックは FW07 20word lists (adjusted) と分散が有意に差がないものを示す。

narrator	Familiarity rank			
	low	lower middle	upper middle	high
FW07 20word lists at -6 dB (adjusted) "mean (sd) / [min:max] ( $\Delta$ range)"				
fhi	16.84 (2.66) [11.25:20.83] ( $\Delta$ 9.58)	34.27 (2.76) [30.42:38.75] ( $\Delta$ 8.33)	71.26 (3.00) [64.58:76.46] ( $\Delta$ 11.88)	84.47 (3.03) [79.58:89.58] ( $\Delta$ 10.00)
fto	12.70 (3.16) [6.67:19.38] ( $\Delta$ 12.71)	24.76 (4.16) [17.29:32.08] ( $\Delta$ 14.79)	66.75 (3.68) [57.50:72.50] ( $\Delta$ 15.00)	82.77 (2.12) [78.33:87.08] ( $\Delta$ 8.75)
mis	14.69 (3.42) [7.50:21.46] ( $\Delta$ 13.96)	36.54 (3.66) [29.79:44.38] ( $\Delta$ 14.58)	75.26 (3.24) [68.54:81.67] ( $\Delta$ 13.13)	89.64 (1.97) [85.21:93.13] ( $\Delta$ 7.92)
mya	12.72 (2.96) [7.71:20.00] ( $\Delta$ 12.29)	28.06 (3.58) [23.13:36.04] ( $\Delta$ 12.92)	68.28 (4.31) [62.08:77.50] ( $\Delta$ 15.42)	83.88 (2.01) [78.33:87.92] ( $\Delta$ 9.58)
FW07 20word lists at -6 dB "mean (sd) / [min:max] ( $\Delta$ range)"				
fhi	39.69 (5.59) [30.63:52.19] ( $\Delta$ 21.56)	45.27 (5.82) [34.69:55.94] ( $\Delta$ 21.25)	66.03 (5.26) [59.69:79.69] ( $\Delta$ 20.00)	78.95 (3.15) <i>[71.56:84.38] (<math>\Delta</math>12.81)</i>
fto	56.13 (7.21) [43.44:72.50] ( $\Delta$ 29.06)	62.36 (4.89) <i>[54.06:70.63] (<math>\Delta</math>16.56)</i>	78.67 (4.54) <i>[72.81:86.56] (<math>\Delta</math>13.75)</i>	84.17 (2.68) <i>[78.44:89.06] (<math>\Delta</math>10.63)</i>
mis	33.23 (6.10) [23.13:51.25] ( $\Delta$ 28.13)	42.23 (6.80) [33.13:57.50] ( $\Delta$ 24.38)	61.53 (8.15) [44.06:76.25] ( $\Delta$ 32.19)	77.33 (4.43) [68.75:85.00] ( $\Delta$ 16.25)
mya	38.97 (8.29) [26.88:55.94] ( $\Delta$ 29.06)	47.78 (7.11) [35.63:60.63] ( $\Delta$ 25.00)	64.25 (5.28) <i>[54.06:74.69] (<math>\Delta</math>20.63)</i>	74.48 (4.01) [67.19:84.06] ( $\Delta$ 16.88)
FW03 50word lists at -6 dB "mean (sd) / [min:max] ( $\Delta$ range)"				
fhi	38.66 (4.37) [27.13:45.13] ( $\Delta$ 18.00)	45.49 (5.62) [35.00:56.63] ( $\Delta$ 21.63)	62.73 (5.44) [52.38:71.50] ( $\Delta$ 19.13)	72.76 (3.60) <i>[64.13:79.00] (<math>\Delta</math>14.88)</i>
fto	54.13 (4.77) [45.25:62.13] ( $\Delta$ 16.88)	61.03 (4.31) <i>[55.75:69.63] (<math>\Delta</math>13.88)</i>	73.05 (4.20) <i>[65.50:82.25] (<math>\Delta</math>16.75)</i>	79.13 (3.73) <i>[71.50:85.25] (<math>\Delta</math>13.75)</i>
mis	36.19 (5.16) [28.13:45.75] ( $\Delta$ 17.63)	44.94 (6.52) [35.38:59.13] ( $\Delta$ 23.75)	59.95 (6.95) [46.63:71.75] ( $\Delta$ 25.13)	71.49 (4.83) [63.88:80.13] ( $\Delta$ 16.25)
mya	40.27 (5.51) [31.50:48.00] ( $\Delta$ 16.50)	48.19 (6.88) [39.63:61.75] ( $\Delta$ 22.13)	61.69 (8.07) [47.88:76.38] ( $\Delta$ 28.50)	69.19 (5.19) [60.50:79.38] ( $\Delta$ 18.88)

タカナ入力した。各聴取者の全試行数は、話者 4 名, 親密度 4 ランク, 各親密度ランク 400 単語 (20 リスト  $\times$  20 単語), SN 比 1 条件 (-6 dB) の 6,400 試行であり, 話者ごとにブロック化して聴取者ごとにランダムな順番で提示した。なお, どの話者を先に聞くかは聴取者間でカウンタバランスをとった。

### 3.1.4 結果

聴取者の回答が提示単語と完全に一致しているものを正答, それ以外を誤答とした。話者ごと, 親密度ランクごとの平均単語理解度を表-3 に示す。表-3 の値は, 各リスト (20 語又は 50 語) の単語理解度の平均と標準偏差, 及び, 最小値 (min), 最大値 (max), 範囲 (range) を求めたものである。表-3 の上段に本実験結果を示し, あわせて中段には FW03 の単語理解度試験結果 [12] から FW07 の 20 単語リストに出現する単語に対する SN 比が -6 dB のときの単語理解度試験結果を用いて計算したもの, 下段には同様に, FW03 のすべての単語 (50 単語) を用いて計算した結

果を示した。表-3 から同一親密度ランクにおける単語理解度は, FW07 に提示音圧調整を行った場合に標準偏差が小さくなっており, 予測どおりリスト間の差が縮小していることが分かる。話者と親密度ランクの組み合わせごとに FW07 リストの提示音圧調整有りとなしの間で  $F$  検定を行った結果, 話者 fhi の親密度ランク high, fto の親密度ランク lower middle, higher middle, high, 及び, 話者 mya の親密度ランク upper middle を除き, 統計的に有意な縮小であることが認められた (片側で  $p < 0.05$ )。更に, 本結果と FW03 の 50 単語の場合の SN 比が -6 dB のときと比較した場合では, 話者 fhi の親密度ランク high, fto の親密度ランク lower middle, higher middle, high を除き分散が有意に小さい (片側で  $p < 0.05$ )。なお, FW03 の 50 単語リストと FW03 の 20 単語リスト間では, 話者 mya の親密度ランク upper middle を除き有意な分散の差は認められなかった ( $p > 0.05$ )。

また, 同一条件内の範囲 (表-3 中のカッコ内で示し

た range の値) は FW07 リストでは最大でも 15.42 ポイント (話者 mya の親密度ランク upper middle) であり, FW03 の 50 語リスト及び 20 語リストの最大差が 30 ポイントに及ぶものが存在することに比べると約半分に減少している。各条件の最大最小幅 (range) に対し対応のある  $t$  検定を行ったところ, FW07 リスト (調整有) の range は, 調整無しの場合, 及び FW03 の 50 語リストより有意に小さくなっていることが示された ( $p < 0.01$ )。

## 4. 考 察

### 4.1 語数縮減の影響

単語親密度と音韻バランスを考慮した単語理解度試験用単語リスト (FW03) を臨床的場面に応用するために, FW07 では試験に用いる単語リストの語数が 50 語から 20 語に縮減された。このため, 各リストの単語に含まれる単語理解度に影響を及ぼす特徴的な音韻の偏りが懸念された。しかし, 図-1, 2 のリストごとの単語理解度, 及び, 表-3 に示したリストごとの平均単語理解度の標準偏差から, 1 リストの語数を 20 単語に減らしたことでリスト間の差は増大していない。これは, 50 単語から 20 単語を構成する際に  $SRT_w$  のずれの大きい単語を除外したことによる効果が大きいと考えられる。

### 4.2 $SRT_w$ による提示音圧調整の効果

前節で述べたように, 1 リストの単語数を 50 から 20 にしたことによる単語理解度のリスト間差の増大はないものの, 図-1, 2 及び表-3 の中段と下段から読み取れるように, リスト間の単語理解度の平均値差が 30 ポイントを超える場合もあり無視できない。本研究では, この差を縮小するための  $SRT_w$  による提示音圧調整の効果を検討した。本実験では雑音の音圧を一定にしているので, 音圧の調整は SN 比の調整と同等である。その結果, 提示音圧調整の有意な効果が確認された。すなわち, 同一親密度ランクの 20 リストに対する平均単語理解度の標準偏差が有意に縮小され, リスト間の最大最小差 (range) も有意に縮小されていることが示された。ただし, 話者及び親密度ランクの組み合わせによって効果が認められない場合が存在し, すべての条件で期待される効果が認められたわけではない。しかし, 分散の縮小が統計的に有意な効果として認められなかった条件では, 調整前の標準偏差及び最大差が他に比較して小さく, 調整後のリスト間差が大きいわけではないので, 調整の効果を否定するものではないと言える。

本研究では,  $SRT_w$  による調整のみで, 傾き (式 (4) の  $\sigma_i$  と  $\sigma_m$ ) を考慮しなかった。Amano ら [12] では

SN 比が  $SRT_w$  の点での単語理解度の推定曲線の傾きがリストごとに異なることを示している。従って, 傾きの違い (式 (4) の  $\sigma_i/\sigma_m$ ) を考慮に入れば提示音圧調整による効果を上げることができると期待される。必要な精度によって, 式 (4) による傾きを考慮に入れるかどうかの判断が必要である。

また, 単語ごとに提示音圧調整を行っているため, 同じ聴取条件において, ある聴取者の聞き取り可能な単語と聞き取り不可能な単語の特性を解析することによって, その聴取者の単語の聞き取りに関与する要因を推察することが可能と期待される。

### 4.3 個人内のリスト間理解度差

前節において FW07 を用い,  $SRT_w$  による各単語の提示音圧 (SN 比) を調整することによって平均単語理解度のリスト間差を縮小できることを示した。ここで行った解析は複数の聴取者の平均値によるものであった。従って, 単語理解度を複数の聴取者の試験結果によって解析する場合において, 提示音圧調整が有効であることが確認されたと言える。例えば, 難聴者の一般的傾向を検討する場合や, コンサートホールや教室などの音響設計のために利用者の平均的特性を計測する場合に適用可能と考えられる。しかし, 補聴器のフィッティングや臨床的応用場面においては常に個人毎の検査結果による判断が要求される。そこで, 各話者の各親密度ランク内の 20 リストのすべての組み合わせの理解度差を聴取者ごとに求めたところ (各新密度ランクで 18,240 例), 親密度ランク low, lower middle, upper middle, high の順に 50%, 40%, 35%, 50%以上の組み合わせで 5 ポイント差 (正答数で 1 単語) 以下に収まる。しかし, 親密度ランク low と high で 20%以上, lower middle, upper middle では 30%以上の組み合わせで 15 ポイント差 (正答数で 3 単語) 以上となっていた (図-3 参照)。このことは, FW07 の単語リストから異なる二つのリストを割り当てて検査した結果の比較には注意が必要な場合があることを示している。この差には, 個人内の変動も含まれているため, すべてがリストに存在する単語の特性による差とは限らない。この点については, 検査法を含めさらなる検討が必要であり, 今後の課題としたい。

### 4.4 単語理解度による日常の聴力評価

「はじめに」で述べたように, 日常の聴力を評価する上で, 音声及び音声言語を用いた検査は重要な役割を果たすと考えられる。Amano ら [12] が FW03 リストを用いて示した親密度ランク間の単語理解度の違いは, FW07 リストを用いた単語理解度でも明確に示されている (表-3 参照)。従って, 親密度ランク間の単語理解度の違いから語彙的情報の影響を考慮した日常生活

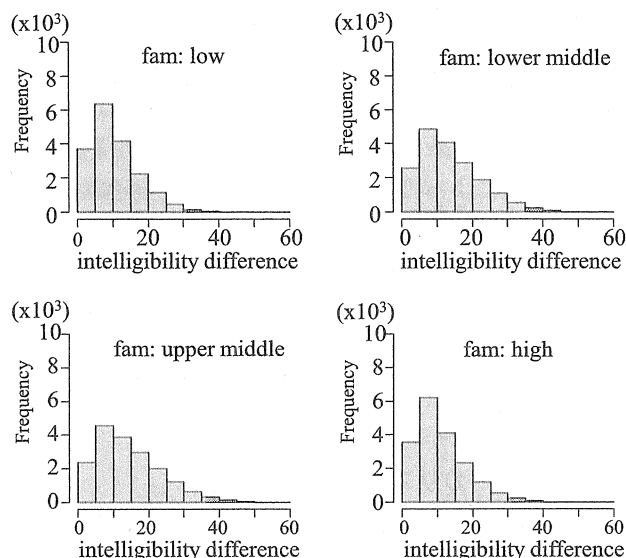


図-3 親密度ランク別のリスト間理解度差分布

における聴力の検討が、FW07 を用いれば FW03 より負荷を軽減して実施可能である。

また、Amano ら [12] が指摘しているように、日常会話を対象にした場合には聞き取りの対象の多くは親密度の高い単語であると考えられるため、高親密度語を用いた単語理解度試験を中心に、講堂での講演や教室での授業を対象にした場合には聞き取りの対象が常になじみのある単語ばかりとは限らないため、中程度の親密度語を用いた単語理解度試験を中心に行うなど、場面に応じた使い分けも有効と考えられる。

更に高次の言語情報による影響については、文を素材とした検査が有効であると期待される。しかし、文章の難易度やコンテキスト情報の統制などは単語を素材とした検査に比べると非常に困難である。現存する文を用いた聴力検査 (TY-89, JHINT) は連続発声音声の理解力を検査することを主な目的としているため、言語的情報の影響を検討することは難しく、FW03 や FW07 の目的とは異なる。このため、どちらか一方の検査結果があればよいというものではなく、異なる指標を与えてくれるものとする。

#### 4.5 提示音圧調整を用いる際の注意

前 4.2 節で  $SRT_w$  による提示音圧調整の効果によりリスト間の理解度の差を小さくすることが可能であることを述べた。しかしこの手法は異なる単語群からなるリスト間の単語理解度差を縮小することを目的としたものであり、結果の解釈において注意が必要である。例えば、本実験において、実験条件として  $-6$  dB (adjusted) としているが、実際には提示音圧 (SN 比) は単語ごとに異なるため、この  $-6$  dB という値は意味を持たない。親密度ランク high, upper middle, lower middle, low のとき  $SRT_w$  の平均は  $-9.42$ ,  $-7.83$ ,

$-4.93$ ,  $-2.54$  dB であり、 $-6$  dB はそれぞれ、 $3.42$ ,  $1.83$ ,  $-1.07$ ,  $-3.46$  dB だけ 50% 点から平均的にずれると予想される音圧 (SN 比) で提示されたことを意味する。

また、FW07 で示されている  $SRT_w$  は白色雑音を用いた結果から推定されたものであるため、実際の実験や検査で用いる雑音が白色雑音以外である場合にも適用できるかは検討が必要である。これには、雑音種による  $SRT_w$  の違いを検討しなければならない。この検討は別の機会に譲ることとしたい。

更に、本研究における  $SRT_w$  の推定には若年成人健聴者に対する単語理解度試験結果を用いた。従って、異なる集団に対してこの  $SRT_w$  の推定値が適用可能かどうか検討の必要がある。例えば、難聴者特有のある音韻に特化した聴取能力の低下なども考慮しなければならない。また、 $SRT_w$  の推定においてはある一定以上の SN 比条件では単語理解度が 100% になることを前提としているが難聴者ではこれは前提とできない [14]。以上のような場合においては単純な  $SRT_w$  による提示音圧 (SN 比) の調整は効果がないだけでなく、かえって結果を乱す可能性もある。臨床場面での応用においては注意が必要である。今後の検討課題としたい。

以上に加えて、実際の実験や検査の場面では、提示音声のラウドネスの違いが影響を与える可能性も指摘されている [15]。FW07 の単語選択において  $SRT_w$  が平均と大きく異なる単語は対象外としたが、それでも調整量が大きい単語ではラウドネスの違いが問題となる場合があることは否定できない。この点も今後の検討課題としたい。

以上から、 $SRT_w$  を基準とした提示音圧調整は、課題は残されているものの、単語リスト間の理解度差を補正に有効であると言える。また、 $SRT_w$  を基準に単語ごとに音圧を調整することが有効であるということは、調整を行わないで実験や検査を行った場合においても、事後検討で  $SRT_w$  を基準とした検討が有効であることを示唆している。

## 5. ま と め

FW03 をベースに各リストの単語数を 50 単語から 20 単語にして単語理解度試験を臨床場面に容易に適用することを目的として作製された FW07 の、単語数を縮減したことによるリスト間の理解度差に与える影響及び  $SRT_w$  を基準とした提示音圧調整のリスト間差の補正効果を検討した。その結果、1 リスト 20 単語とした FW07 リストは FW03 に劣ることなく、語彙情報の影響を考慮し音韻バランスのとれた単語リストとし



て単語了解度試験に用いることが可能であることを示した。また、 $SRT_w$  による提示音圧調整がリスト間の了解度差を補正する手法として有効であることを示した。更に、臨床場面に応用する際の注意事項を示すとともに、提示音圧調整の問題点を考察した。

### 文 献

- [1] 日本聴覚医学会, “57-S 語表” (1983).
- [2] 日本聴覚医学会, “67-S 語表” (1987).
- [3] P. Egan, “Articulation testing methods,” *Laryngoscope*, 58, 955–991 (1948).
- [4] M. Nilsson, S.D. Soli and J.A. Sullivan, “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,” *J. Acoust. Soc. Am.*, 95, 1085–1099 (1994).
- [5] 米本清, 立石恒雄, 木場興次, 倉内紀子, “補聴器適合評価用 CD (TY-89) 及び 57-S 語表の単音節明瞭度と音圧,” *Audiol. Jpn.*, 32(5), 429–430 (1989).
- [6] 大沼直紀, 中川辰雄, 今井秀雄, 岡本途也, “[日常生活文リスト] の検討—マルチトーカーノイズ下における了解度検査法の試行—,” *Audiol. Jpn.*, 32(5), 299–300 (1989).
- [7] M. Shiroma, T. Iwaki, T. Kubo and S. Soli, “The Japanese hearing in noise test,” *Int. J. Audiol.*, 47, 381–382 (2008).
- [8] 天野成昭, 近藤公久, 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 (FW03),” NII 音声資源コンソーシアム (2006).
- [9] 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, 東京, 1999).
- [10] 坂本修一, 鈴木陽一, 天野成昭, 小澤賢司, 近藤公久, 曾根敏夫, “親密度と音韻バランスを考慮した単語了解度試験用単語リストの構築,” 音響学会誌, 54, 842–849 (1998).
- [11] 坂本修一, 天野成昭, 鈴木陽一, 近藤公久, 小澤賢司, 曾根敏夫, “単語了解度試験におけるモーラ同定に対する親密度の影響,” 音響学会誌, 60, 351–357 (2004).
- [12] S. Amano, S. Sakamoto, T. Kondo and Y. Suzuki, “Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese,” *Speech Commun.*, 51, 76–82 (2009).
- [13] 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 (FW07),” NII 音声資源コンソーシアム (2007).

- [14] H. Sato, K. Kurakata and T. Mizunami, “Accessible speech messages for the elderly in rooms,” *Proc. 9th Western Pacific Acoustics Conf. (WESPAC IX)*, 199–204 (2006).
- [15] 長谷芳樹, 橋 亮輔, 阪口剛史, 細井裕司, “親密度別単語了解度試験用データセット (FW03) 単音節音節のラウドネス校正,” 音響学会誌, 64, 647–649 (2008).

### 近藤 公久

1986 年慶應義塾大学大学院管理工学研究科前期課程修了。同年, 日本電信電話(株)に入社。博士 (学術)。視聴覚知覚, 単語認知, 日本語の読みの研究に従事。2001 年より 2005 年まで NTT データ技術開発本部へ転籍, 2012 年より (株)国際電気通信基礎技術研究所に出向。日本音響学会, 電子情報通信学会, 日本心理学会, アメリカ音響学会等の会員。

### 坂本 修一

1997 年東北大学大学院情報科学研究科博士前期課程修了。同年(株)日立製作所入社, 2000 年から東北大学電気通信研究所。現在, 同所准教授。博士 (工学)。聴覚・マルチモーダル感覚情報処理過程に関する研究に注力。日本音響学会, 電子情報通信学会, 日本バーチャルリアリティ学会, 日本感性工学会等の会員。

### 天野 成昭

1985 年東京大学大学院人文科学研究科修士課程 (心理学) 修了。同年, 日本電信電話(株)に入社。1996 年東京大学より博士号 (心理学) 取得。同年米国インディアナ大学客員研究員 (兼務)。2010 年より愛知淑徳大学人間情報学部教授。日本音響学会, 日本心理学会, 米国音響学会会員。

### 鈴木 陽一

1981 年東北大・工・博士課程修了 (工博)。東北大学大計センター助教授等を経て, 99 年から同電気通信研究所教授。聴覚及びマルチモーダル知覚過程と工学応用研究に従事。日本音響学会佐藤論文賞 (92, 94 年), 同会長 (05~07 年)。アメリカ音響学会, 日本 VR 学会, 電子情報通信学会フェロー。