音声認識技術を用いた明瞭性評価の試み －屋外拡声音の「聴き取りにくさ」とJulius尤度の関係－*

○栗栖清浩（TOA），川島佑亮†，安啓一，荒井隆行（上智大）

1　はじめに

屋内の残響場における音声明瞭性評価指標は様々なものが提案，規格化されている[1]。しかし屋外では過度の暗騒音，気象の影響，ロングパスエコー等，屋内音場で殆ど現れない音響障害が発生しているため，屋内を想定した従来の明瞭性指標をそのまま屋外拡声音の評価に用いることが難しい。

更に，STI などの現行の明瞭性物理指標は，基本的に音源信号と受音点信号の比較から劣化の度合いを数値で表現しているが，これは伝送系の優劣を評価しているものの，聴取者が聴いている音声そのものが明瞭かどうかの指標，即ち聴取者の立場に立った明瞭性評価指標になっていないのではないか，と著者らは考えている[2]。同様に，単語了解度や単音節明瞭度も，音源（正解）と受音点での回答を比べた正答率であり，人を使って伝送系を評価しているに過ぎないとも考えられる。

一方，屋外拡声の現場では伝送系に瑕疵が無くとも明瞭性が確保されないことが少なくなく[3]，伝送系の評価だけでは明瞭性を評価できないことがある。また，技術的，社会的な制約から伝送系の測定自体が困難なことも多いため，受音点で収音した信号のみから明瞭性を評価したいという現場の要求もある。

そこで今回，受音点の信号そのものの明瞭性を評価する手法確立の一助として，音声認識エンジン Julius[4]の解析結果から聴感印象の推定を試みたところ，ある条件の下では聴き取りにくさ LDR（Listening Difficulty Rating）[5]を Julius の尤度で精度良く推定できることが分かったので報告する。

2　聴取実験

屋外拡声を想定した試験音に対する聴取者の応答を確認する為，以下のような聴取実験を実施した。

[試験音]

キャリア文「これから流す単語は○○です」の○○に日本語 4 モーラ単語（FW07 の高親密度リストから 40 個, 中低親密度リストから 40 個）を 1 つずつ挿入した男声アナウンス 80 文（fs=44.1 kHz）に対し，40 種の処理を施した 3200 試験音を用意した。40 種の処理は以下の因子水準の組合せ 5×4×2=40 である：

因子①：ロングパスエコー5 水準（エコー数：0～4，エコー間隔：0.5 s，エコー振幅：直接音を 1 とした調和数列 1, 1/2, 1/3, 1/4, 1/5），

因子②：文節間ポーズ 4 水準（文節間に無音区間 0.0 s, 0.5 s, 1.0 s, 1.5 s を挿入），

因子③：歪＆帯域制限 2 水準（1. 何もしない，2. 歪：振幅を 100 倍して振幅 1 でクリップ，を与えた後，帯域制限：300～3,400 Hz を施す）

[聴取者]

聴取者は 40 名（20 代：6 名，30 代：11 名，40 代：11 名，50 代：9 名，60 代：3 名，♂：31 名，♀：9 名）で，内 2 名に片耳のみ若干の聴力低下の自己申告あり。

[試験音の提示]

防音室内において聴取者一人当たり 80 試験音（単語重複無し，40 処理を 2 回聴取）を耳覆い密閉型ヘッドホンを介して提示した。提示レベルは実験前練習中に聴取者が適切なレベルに調整し実験中は固定した。

[タスク]

試験音を一回聴取したあと，聞こえた単語のキーボード入力，及び PC 画面上に表示された聴取印象：1. 聴き取りにくくない，{2. やや，3. かなり，4. 非常に} 聴き取りにくい，から一つをクリックして回答させた。

[実験結果]

40 種の処理毎の単語の正答率と，LDR（処理毎の全回答数に対する聴取印象 2, 3, 4 回答

---

* Estimating the listening difficulty rating by the likelihood obtained from the speech recognition engine 'Julius' for outdoor mass notification speech, by KURISU Kiyohiro(TOA), KAWASHIMA Yusuke, YASU Keiichi, ARAI Takayuki(Sophia Univ.). †現在, 株式会社日立製作所勤務(Current affiliation: Hitachi, Ltd.)

合計数の割合）を算出した。三元配置分散分析の結果，正答率に対して因子①②③共に有意差が無かったが，LDR は 3 因子共に有意差あり，多重比較検定（Tukey の HSD 検定）でも有意差があった。

3　Julius による試験音の解析と LDR との関係

3200 試験音を全て Julius で解析した。ここで，Julius は事前学習等のチューニングを行っていない。今回の聴取実験では正答率は有意でなかったため Julius 認識結果（単語及び文章）の正答率には注目せず，尤度 P1BS（pass1_best_score）[4]に着目した。これは対数尤度で表された仮説のスコアであり，Julius が認識結果にどの程度自信を持っているかの度合いに相当し，これが聴取者の LDR に対応しているものと著者らは考えたからである。

Fig. 1 に 3200 試験音について P1BS と LDR の関係を示すが，残念ながらこれから LDR と P1BS の間に何らかの関係性を見出すことは難しい。

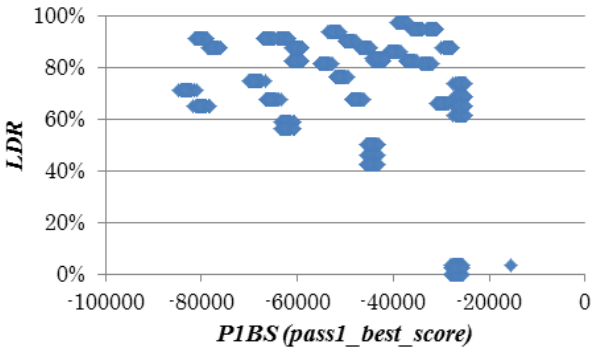

Fig. 1 LDR vs. the likelihood (P1BS) of speech recognition engine for all data.

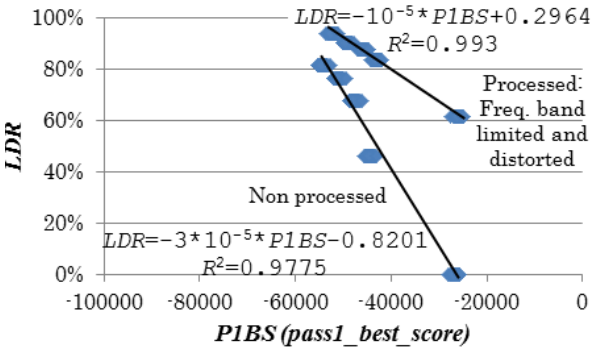

Fig. 2 LDR vs. P1BS for limited data: Factor②=0.5s.

しかし，例えば Fig. 2 のように因子②の文節間ポーズ水準 0.5s の試験音に対するデータに限定すると，因子③の「歪＆帯域制限」処理した試験音（Processed）としなかった試験音（Non processed）とで，それぞれ決定係数の高い回帰直線を得ることができた。

4　おわりに

屋外拡声を模擬した試験音での結果ではあるが，ある限定された条件の下で Julius 解析結果から LDR を精度良く推定することができた。音声認識エンジンを用いた例は，正答率を利用したものや[6][7][8], 認識結果の信頼度を尤度から推定するもの[9]等があるが，いずれも正答率に注目しているという意味で，伝送系の評価に関するものと考えられる。対して今回の報告は，尤度を音声認識エンジンの自信度合いと解釈し聴取印象と対応付けた点が特徴で，受音点の信号そのものの明瞭性評価手法につながるのではないかと期待している。（本研究は 2013 年度に実施された。）

謝辞

助言を頂いた近藤和弘先生（山形大学），小林洋介先生（都城高専）に感謝します。

参考文献

[1] 例えば IEC60268-16 の STI や日本建築学会環境基準 AIJES-S0002 の STP（音声伝送性能）など.

[2] 栗栖清浩，安啓一，荒井隆行，中村進，アナウンス音声から導出した物理量と「聴き取りにくさ」の関係－明瞭性の評価は音声そのものを評価すべきではないか？－，建築音響研究会資料 AA2013-36, 2013.

[3] 栗栖清浩，松本泰，山内昭弘，有賀成嘉，屋外防災拡声システムの現状と課題，音講論集，2013 年 9 月，1529-1532.

[4] 汎用大語彙連続音声認識エンジン Julius / Julian, rev. 3.3 (2002/09/11), http://julius.sourceforge.jp/book/Julius-book-3.3-ja.pdf

[5] Morimoto M, Sato Hi, Kobayashi M, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," J. Acoust. Soc. Am., vol. 116, pp. 1607-1613, 2004.

[6] Takano Y, Kondo K, "Estimation of speech intelligibility using speech recognition systems," *IEICE Trans. Inf. & Syst.*, vol. E93-D, no. 12, pp. 3368-3376, 2010.

[7] Kondo K, "Estimation of speech intelligibility using objective measures," *Applied Acoustics*, vol. 74, pp. 63-70, 2013.

[8] Arai T, "Time-reversed reverberation yields lower speech recognition rate by human and machine," *Acoust. Sci. & Tech.*, vol. 34, no. 2, pp. 142-146, 2013.

[9] 中川誠一，堀部千寿，音響尤度と言語尤度を用いた音声認識結果の信頼度の算出，情報処理学会研究報告，SLP-036, pp.87-92, 2001.