

音声の変動量と「聴き取りにくさ」について — 残響環境下での高親密度単語の評価 —*

☆中村進(上智大・理工), 栗栖清浩(TOA), 安啓一, 荒井隆行(上智大・理工)

1 はじめに

音声明瞭性を評価する方法として、従来から *STI* (Speech Transmission Index) ^[1] が主に用いられている。しかし、その算出はインパルス応答に大きく依存していることから、音場の違いには敏感であるが、話者の違い等の音声自体が異なる場合やスピーカによる歪等の処理が非線形な場合において音声明瞭性の評価尺度として十分ではない事が報告されている^[2]。そこで、音場で付加される overlap-masking に着目した *SOR* (Signal to Overlap-masking Ratio) ^[3] によって、音声の明瞭性を推定する事を行ってきた^[2]。*SOR* は、ある特定の文章において主観評価値である「聴き取りにくさ (*LDR*: Listening Difficulty Ratings) ^[4]」を推定する事が可能である^[5]。しかし、様々な文章を対象とした場合においては、*SOR* から *LDR* を精度良く推定する事は困難な可能性がある^[6]。

そこで、*LDR* が等しいにも関わらず *SOR* が異なる原因を探るため、市販ソフト^[7]を用いて、音源から求めたラウドネスの *AM* 変調の度合いに相当する変動量 (*DLF*: Depth of Loudness Fluctuation) を求めた。その結果、変動周波数が 2.5~7.5 Hz の範囲において *LDR* が増加すると *DLF* が減少する関係が見られた^[8]。しかし、*SOR* や *DLF* との比較で用いてきた *LDR* は、当初 Morimoto et al. により提案された高親密度単語を対象とした *LDR* とは異なり、文章を対象としていた事が記されている^[9]。

そこで本研究では、Morimoto et al. の *LDR*^[4] にならい、4 モーラの高親密度単語を対象とし、物理量である *SOR*, *DLF* と音声明瞭性との関係について報告する。

2 聴取実験

音声の客観的明瞭性評価値である物理量と主観的明瞭性評価値である *LDR* の比較を行うため、音声明瞭性の主観評価実験を行った。以下に聴取実験の内容を示す。

2.1 刺激

実験刺激は、男女の話者による 20 単語 (*FW07*^[10]における高親密度 4 モーラ単語) に、8 種のインパルス応答 ($h_i: i=1\sim 4$, $h_i^*g: i=1\sim 4$, $RT = 0.5, 1.0, 1.4, 2.6s$) を畳込んだ音声を用いた。ここで、 h_i^*g は室の共振によるモード周波数を抑える音響調整前、 h_i が音響調整後を模擬している。これは、 h_i には共振がなかったため、フィルタ g によってモード周波数の成分を増幅させたためである。ゆえに、話者条件 (2 条件) × 残響条件 (4 条件) × 音響調整 (2 条件) の計 16 条件とし、各条件に 20 単語ずつ割り当て、全 320 の刺激を作成した。

2.2 聴取者

聴取者は、日本語を母語とする平均年齢 23 歳 (20~26 歳) の健聴者 40 名 (男性 31 名・女性 9 名) であった。

2.3 実験手順

実験は上智大学荒井研究室の防音室で行った。刺激は、PC に接続したオーディオインタフェース (*EDIROL UA-25EX*) を介し、耳覆い型ヘッドホン (*SENNHEISER HDA 200*) から提示した。

音声明瞭性の主観評価実験として、「聴き取りにくさ」の実験を行った。実験は、刺激を一度だけ提示し、刺激に対して 4 段階 (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) で評価させた。

* The relationship between the depth of loudness fluctuation and the listening difficulty ratings of a word set with high word familiarities under reverberant conditions, by NAKAMURA, Susumu (Sophia Univ., Japan), KURISU, Kiyohiro (TOA Corp., Japan), YASU, Keiichi and ARAI, Takayuki (Sophia Univ., Japan).

3 結果・考察

実験によって得られた LDR の結果を、物理量である SOR と比較した。ここで用いる LDR は、刺激毎に聴取者が聞き取りにくいと判断した（4段階中の2~4を選んだ）人数の割合を表している。そして、心理的な距離に対応している Z 値へと変換し、 Z_{LDR} で表した^[11]。 SOR と Z_{LDR} の関係を Fig. 1 に示す。 Fig. 1 より、 SOR が -5dB 付近において Z_{LDR} が 3、即ち標準偏差で 3 も異なっているため、 SOR から LDR を精度良く推定することは困難な可能性がある。

そこで、 Z_{LDR} が異なる場合について、刺激の時間変動特性を確認した。例として、 $Z_{LDR} = -2.0, -1.0, 0, 1.0, 2.0$ の刺激から 1 刺激ずつ選び、各変動特性を Fig. 2 に示した。ここで求めた変動特性は、音声を 1/3 オクターブ帯域に分割し、帯域毎のラウドネスから求める周波数特性である。そして、この時の周波数を変動周波数とし、各変動周波数における和によって DLF を求めた。すると、 Z_{LDR} が増加するほど、10 Hz 以下の変動周波数帯域において DLF が減少することが分かった。

聞き取りに影響を与える変動周波数帯域は、文章での結果^[8]と似た傾向であったため、別研究^[8]にならい、変動周波数 2.5~7.5 Hz における DLF の合計とした DLF_{BLS} を LDR と比較した。その結果を Fig. 3 に示す。図中に示した決定係数が 0.77 と、Fig. 1 で示した 0.55 に比べて高いことから、 DLF_{BLS} の変化によって明瞭性をある程度推定できる可能性が示された。

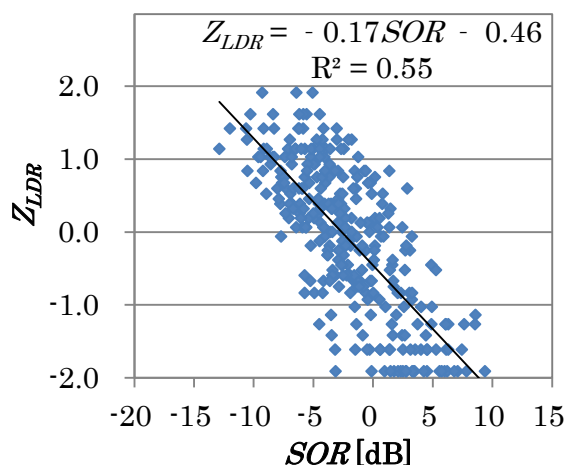


Fig. 1 SOR vs. LDR

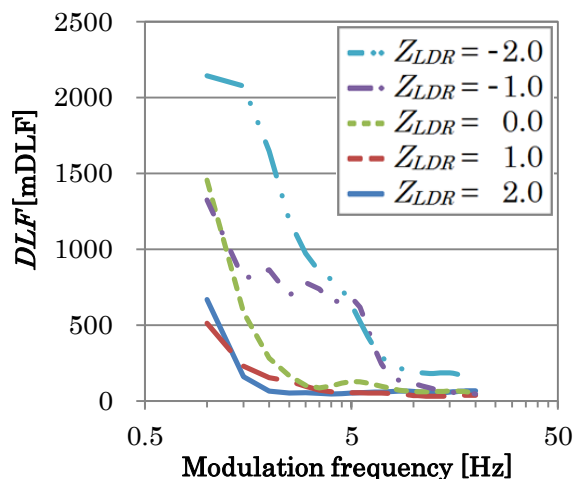


Fig. 2 Modulation frequency property

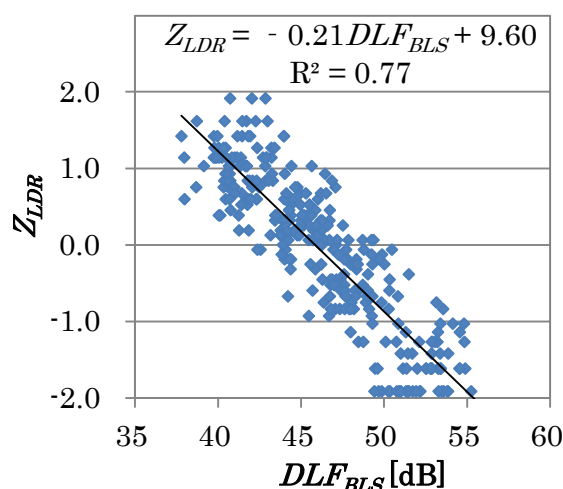


Fig. 3 DLF_{BLS} vs. LDR

4 おわりに

本研究では、高親密度単語を対象とし、拡声音から求める物理量と音声明瞭性との関係を調べた。その結果、帯域を限定した DLF_{BLS} と LDR の関係には、比較的高い相関がみられることが分かった。

参考文献

- [1] IEC 60268-16, Edition 4.0, 2011.
- [2] 栗栖 他, 建音研資料, AA2011-47, 2011.
- [3] T. Arai et al., AST, 28(6), 438-441, 2007.
- [4] Morimoto et al., JASA, 116(3), 1607-1613, 2004.
- [5] 栗栖 他, 音講論(春), 1225-1226, 2012.
- [6] 栗栖 他, 音講論(秋), 1251-1252, 2012.
- [7] 小野測器, Oscope Ver.2 マニュアル.
- [8] 栗栖 他, 音講論(春), 2-P-41, 2013.
- [9] 中村 他, 音講論(春), 1227-1228, 2012.
- [10] 近藤 他, 信学技報, 107(432), TL2007-62, 43-48, 2008.
- [11] 日本建築学会, AIJES-S0002-2011.