

様々な音声アナウンスのSORと「聴き取りにくさ」の関係について

○栗栖清浩 (TOA), 中村 進, 安 啓一, 荒井隆行 (上智大・理工)

1 はじめに

音声明瞭性の物理評価指標 STI は拡声システム (電気音響) の拡声音の評価指標としても多用されているが, イコライザ等で系の周波数特性を変化させたとき, 聴取印象が大きく変わるにもかかわらず, STI は全く変化しないという場合があることが分かっている. 例えば定在波のある室で, 定在波の周波数付近のゲインをイコライザで低減して拡声した場合, すっきりと聞き取りやすい拡声音になるものの, STI が変化しないため, そのイコライザの調整効果を STI による数値で評価することは出来なかった.

そこで著者らは, 室の共鳴や残響によるオーバーラップマスキングが, 拡声音の明瞭性を低減させているという視点に基づく物理指標 SOR (Signal to Overlap masking Ratio) を提案し, 拡声システムの調整前後で LDR (「聴き取りにくさ」) が改善するのを, SOR の改善量として客観的に評価できる可能性を示した[1][2].

提案した SOR は, 受音点における直接音成分と, それ以降に到来する成分のパワー比に概ね相当し, 直接音成分が大きくそれ以降の成分が小さいほど SOR の数値は向上する. また, 拡声システムがインパルス応答だけで完全に記述できる場合, SOR は C 値と等価であるが, コンプレッサ等の非線形処理が含まれる場合でも, SOR は処理の効果を数値の変化として表現することが可能である.

前報[1][2]では単一話者, 単一アナウンスという限定的な聴取実験の結果に基づき, SOR から LDR を予測する式を導出したが, 今回, SOR の汎用性を検討するため, 複数話者, 複数アナウンスの場合, SOR と LDR にどのような関係が見られるかを確認した.

2 聴取実験

「ATR 研究用音声データベース 503 文 B セット」[3]から 5 つ (話者 A), 「名古屋工大日本語音声データベース」[4]から 5 つ (話者 B) の合計 10 のアナウンスに, 10 種のインパルス応答を畳み込んだ 100 の刺激について「聴き取りにくさ」に関する聴取実験を行った.

10 種のインパルス応答とは, 残響時間 $T_{60}=0.2s, 0.5s, 0.9s, 1.4s, 2.6s$ の室を模擬したインパルス応答 5 種 h_j ($j=1\sim 5$) と, (170Hz で 14.5dB, 620Hz で 4dB のゲインを持つ) 室の共鳴を模擬したフィルタ g を h_j に畳み込んだ 5 種の合わせて 10 種である. これらは, 共鳴している室 $h_j * g$ と, イコライザを調整し共鳴が抑えられた h_j とを模擬している.

聴取者は日本語を母語とする 10 名 (20 代: 9 名, 40 代: 1 名) で, 防音室 (容積 27.9m³) において耳覆い型ヘッドホンを介しランダムな順番で聴取した刺激の「聴き取りにくさ」(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) を回答した.

3 実験結果と考察

Fig. 1 は各刺激に対して少しでも聴き取りにくいと回答された数 (評価 2~4 の合計) の割合を Z 値に変換した Z_{LDR} と各刺激の SOR との対応関係を示したものである.

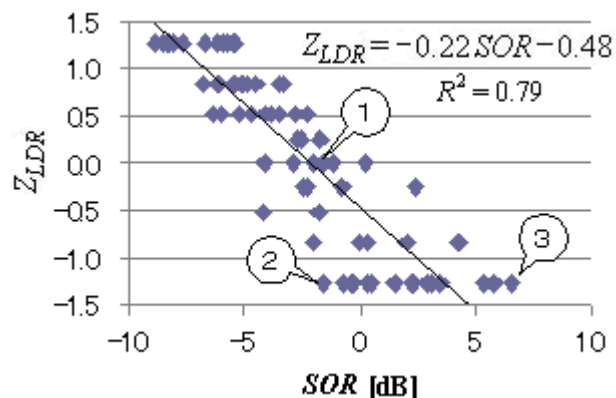


Fig. 1 Z_{LDR} (Z-score of LDR: Listening Difficulty Rating) vs. SOR and the regression line

* Broadcasted speeches in various acoustic conditions and their SOR (Signal to Overlap-masking Ratio), by KURISU Kiyohiro (TOA Corp., Japan), NAKAMURA Susumu, YASU Keiichi and ARAI Takayuki (Sophia Univ., Japan).

刺激が単一話者，単一アナウンスであった前報[1][2]に比べ，決定係数の低い回帰直線となっている．例えば聞き取りやすい場合 ($Z_{LDR} = -1.25$)， LDR には 10dB ほど開きがあり，聞き取りにくい場合 ($Z_{LDR} = 1.25$)でも 5dB 程度の開きがあるため，このままではアナウンスの種類が変わると，測定した SOR から LDR を精度良く推定することが出来ない．

参考までに今回の実験で用いた 10 アナウンスと前報の実験で用いた 1 アナウンスの合わせて 11 アナウンスのそれぞれについて， SOR と Z_{LDR} の回帰直線を描いた (Fig. 2)．

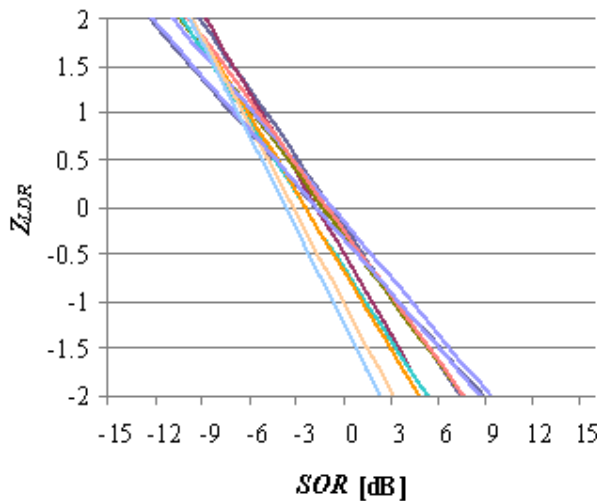


Fig. 2 Regression line for each speech

各回帰直線の傾きは異なるが，共通して図中の $SOR = -6$ dB, $Z_{LDR} = 1$ 付近を通っていて， $Z_{LDR} = 2$ に比べて $Z_{LDR} = -2$ における SOR のバラツキが大きくなっている．これはおそらく SOR が小さな領域では，残響時間が長いなど拡声音場として劣悪な環境であり，アナウンスが異なってもほぼ同様に聞き取りにくく (Z_{LDR} が大きく) なるのに対し， SOR が大きな領域では，拡声音の品質がある程度確保されており，各アナウンスの差が明確に聞き分けられたため， Z_{LDR} のバラツキが大きくなったものと思われる．

4 帯域別 SOR の試み

以上のとおり，複数のアナウンスの場合，これまでの広帯域 SOR では Z_{LDR} を精度良く推定することが難しい．

Fig. 3はFig. 1において SOR はほぼ同じだが Z_{LDR} が大きく異なる刺激①と②，そして刺激②と Z_{LDR} は同じだが SOR が大きく異なる刺激③のパワースペクトルを描いたものである．

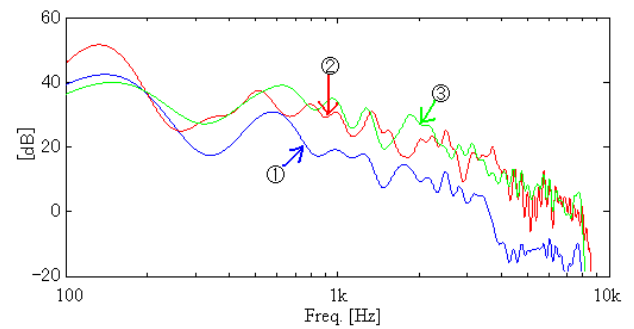


Fig. 3 Power spectra of stimuli ①~③ in Fig. 1

中高域において②，③は①より 10dB ほど高く，また①は 4kHz から急激に低下しており，①と②③を区別する手がかりを示唆している．

次に刺激①②③を 1/3oct.帯域に分け，各帯域の SOR を算出すると (Fig. 4)，①②と③の違いおよび，①②の細かな違いが議論できそうである．

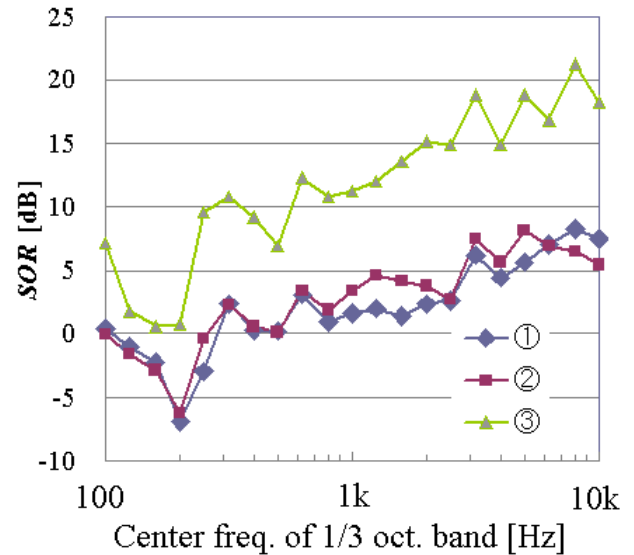


Fig. 4 1/3oct-band- SOR of ①~③ in Fig. 1

5 おわりに

従来の広帯域 SOR だけでなく各アナウンスのスペクトル及び帯域別 SOR を併用することで，Fig. 1がばらつくのを説明できる可能性が示唆された．

参考文献

- [1] 栗栖 他，建音研資料，AA2011-47，2011．
- [2] 栗栖 他，音講論(春)，1225-1226，2012．
- [3] 匂坂，浦谷，日音学誌，48，878-882，1992．
- [4] The Nitech Japanese Speech Database，<http://hts.sp.nitech.ac.jp/>