

拡声音声を用いて測定した物理量 SOR と  
「聴き取りにくさ」の関係  
— 拡声システムの調整結果を  
評価するツールとして —

栗栖 清浩  
(TOA)

中村 進 安 啓一 荒井 隆行  
(上智大・理工)

2011 年 12 月 21 日

主催 日本音響学会建築音響研究委員会

# 拡声音声を用いて測定した物理量 $SOR$ と「聴き取りにくさ」の関係

— 拡声システムの調整結果を評価するツールとして —

## Listening Difficulty of the Transmitted Speech and its $SOR$

An evaluation tool for acoustic system tuning

栗栖清浩<sup>†</sup>, 中村 進<sup>‡</sup>, 安 啓一<sup>‡</sup>, 荒井隆行<sup>‡</sup>

KURISU, Kiyohiro<sup>†</sup>, NAKAMURA, Susumu<sup>‡</sup>, YASU, Keiichi<sup>‡</sup>, ARAI, Takayuki<sup>‡</sup>

<sup>†</sup> TOA 株式会社      <sup>‡</sup> 上智大学

<sup>†</sup> TOA Corporation    <sup>‡</sup> Sophia University

### 内容概要

拡声システムによる音声の明瞭性を表す指標として国内外で  $STI$  (Speech Transmission Index) が広く用いられているが、その数値はスピーカ特性、数、そして室の建築音響的要因だけでほぼ決定されてしまうことから、システム設置後の調整でいくら拡声音の聴こえを改善できたとしても、 $STI$  はほとんど変化せず、音響調整の効果を評価することが困難であった。これに対し著者らが提案した  $SOR$  (Speech to Overlap-masking Ratio) は音響調整の前後で数値が変化することから、聴取実験により  $SOR$  と「聴き取りにくさ」の関係性を明らかにすることで、 $SOR$  に基づいた拡声音の「聴き取りにくさ」の推定が可能になった。

### 1. はじめに

公共の場やホール、講義室など様々な場所で、明瞭な音声伝達を目的に拡声システムが導入されている。その構成は音源、ミキサ、各種の信号処理器、増幅器、スピーカからなるが、これらを設置するだけで、必ずしも対象とするエリア内で期待通りの明瞭な拡声音を得られるわけではない。設置前のシステム設計段階でシミュレーションを行うなどして、エリア内において十分な音圧レベルを確保し、明瞭な音声となるよう音響設計がなされるが、通常は設置後に現場の状況に合わせた音響調整が必要となる。

音響調整は、小規模または単純なシステムでは音量を調整するだけで済むが、システムが複雑化しラインアレイスピーカ[1]などを多用している場合は、複数のスピーカの遅延時間を調整するタイムアライメントやマルチウェイのスピーカであればクロスオーバー周波数を調整する(例えば[2][3]など)など高度な調整が必要となってくる。また、室の共鳴(モード周波数)により拡声音が過度に響くことを避けるため、調整時に

モード周波数を特定し、その周波数の出力をフィルタで低減して拡声することなども行われる<sup>1</sup>。

この共鳴制御の処理を行うと過度な響きが抑えられるため調整後はスッキリとした拡声音になるのだが、単なるフィルタ処理であることから  $STI$  (Speech Transmission Index) [4]は殆ど変化しない。 $STI$  は国内外において明瞭性の指標として広く用いられているのだが、この例のように音響調整により拡声音の聴こえが良くなるにもかかわらず  $STI$  の値が変化しないことについて、 $STI$  では音響調整の良し悪しを評価できないとの現場の声、システムの周波数特性は  $STI$  では考慮されていないとの指摘[5]などがある。拡声システムを用いた場合の音声の伝送経路すなわち  $pathware$ [6]<sup>2</sup>は図1のように拡声機器系と音響系からなるが、音響系により  $STI$  がほぼ決定されてしまうことから、拡声機器系

<sup>1</sup> 室のモード周波数の特定とフィルタ作成を自動的に行う ARC 機能 (ARC: Automatic Resonance Control) を搭載した信号処理器を用いる。

<sup>2</sup> "channel"と呼んでも良いがここでは文献[6]に倣って "pathware"と呼ぶ。

の評価には向かないというわけである。

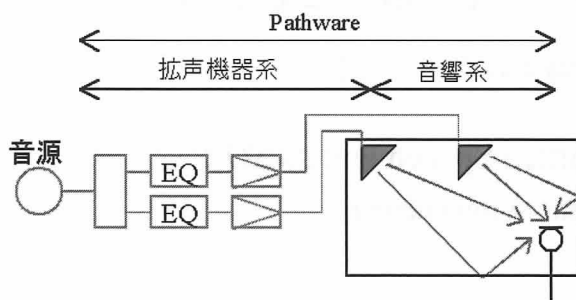


図 1 音声伝送の pathware

本報告はこのような状況に鑑み、拡声機器系の音声明瞭性に関する性能を評価することを目的に、著者らが過去に提案した  $SOR$  (Signal to Overlap-masking Ratio) [7] を応用した結果について述べたものである。音響調整を想定した場合の聴取実験から得た  $SOR$  と「聴き取りにくさ」の回帰式により、音響調整でどれだけ「聴き取りにくさ」が改善できたかを具体的な数値で示すことについて述べる。最後の議論では、今回用いた  $SOR$  と、従来から使われている各種明瞭性指標とを比較し、それらの類似性と差異について検討した。

## 2. 定常部抑圧処理の評価量としての $SOR$

残響下での拡声音の明瞭性を向上させることを目的に提案された定常部抑圧 (Steady-State Suppression, 以下  $SSS$  と略記) 処理[8][9][10]は、あらかじめ音声信号の定常部 (≒母音部) の振幅を抑圧して拡声することで、後続の過渡部 (≒子音部) へのオーバーラップ・マスキング[11][12]を軽減するものであり (図 2), ある範囲の残響条件下で単音節明瞭度が向上することが分かっている[10]。

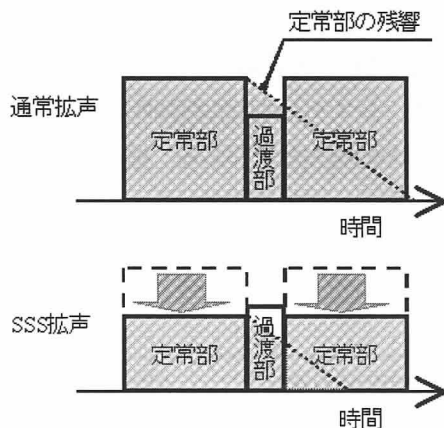


図 2 定常部抑圧処理 (SSS) の概念図

しかし、 $SSS$  は音声信号の包絡に応じて振幅を変化させるという非線形な処理であることから、その特性はインパルス応答上には現れず、すなわち  $SSS$  の効果はインパルス応答から算出する  $STI$  で表現することができない。Pathware の線形特性でしかないインパルス応答では、非線形な  $SSS$  を評価することができないわけである。しかし実際は  $STI$  が全く変化しないにもかかわらず、明瞭度が向上しているのも事実である。

そこで、受信点において  $SSS$  により改善されるオーバーラップ・マスキング量を用いて  $SSS$  の効果を評価する試みとして、 $SOR$  (Signal to Overlap-masking Ratio) が導入された[7]：

$$SOR = 10 \log_{10} \frac{\int_0^T |s(t) * h_{50}(t)|^2 dt}{\int_0^T |p(t) * h(t)|^2 dt} \text{ dB.} \quad (1)$$

これは、ターゲット信号  $s(t)$  及び明瞭性に寄与するとされている 50ms までの残響音をマスキ、ターゲットに先行するキャリア文  $p(t)$  の残響をマスカとして、マスキとマスカのパワー比、すなわち  $SN$  比を計算したものである。 $SSS$  により  $SOR$  が向上すると単音節明瞭度も向上する傾向が分かった[7]。すなわちインパルス応答から算出する  $STI$  は変化しないにも拘らず単音節明瞭度が向上しているのだが、これは音響調整により拡声音の聴こえが改善しているにも拘らず  $STI$  が変化しないのにことに類似している。そこで、この  $SOR$  が音響調整の評価指標として利用できないか、その可能性を探ることにした。

## 3. 実際の拡声音を用いた $SOR$

拡声システムは案内放送、BGM、非常放送など様々な種類の音声をお届けするが、決して白色ノイズや swept sine 等のテスト信号を拡声することを目的としていない。あくまで基本は音声 (音楽) の拡声であり、音響調整はその現場で実際に拡声される音声の明瞭にすることが目的であるので、新しく検討する指標の試験音は実際の拡声音を用いたい。Pathware 上の非線形成分を測定するもっとも代表的な試験音 (プローブ音) は実際の音声であると Steeneken & Houtgast[13]も述べている<sup>3</sup>。実際の音声といってもその種類は様々なので、

<sup>3</sup> しかし、論文[13]中では実際の音声を用いず、それに変わる試験音源を用いているようだ。

本報告では万が一の場合で重要性の高い、非常放送のアナウンス音(の一部)「職員の指示に従い、落ち着いて避難してください」を試験音に用いることとした。

### 3.1. 音声全体の SOR

SSS の効果はこれまで主に単音節明瞭度で評価していたので、当初提案した SOR も図 3 に示すように音声の一部(キャリア文に埋め込まれたターゲット単音節の部分)だけにおける overlap-masking の量を算出していた[7]。

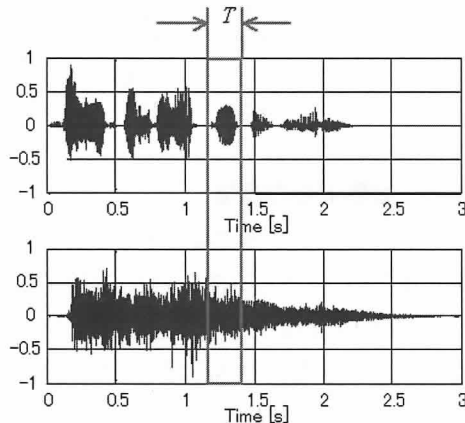


図 3 単音節 SOR (上: 直接音相当信号, 下: 拡声音)

しかし、今回は実際に拡声されることのある特定のアナウンス全体を対象としたことから、図 4 のように時間フレーム毎に算出した短時間 SOR を平均し、文章全体の SOR とみなすことにした。導出方法は次のとおりである：

- ① あらかじめ pathware のインパルス応答  $h$  を測定する、
- ② 音源  $s$  を拡声したときの受音点音圧  $p$  を実際に測定する、
- ③ 受音点における直接音相当成分(直接音及び 50ms までの残響音)  $p_d = s * h_{50}$  を算出する (\*: 畳み込み演算子)、
- ④  $i$  番目のフレームの  $E_i$  ( $p$  のエネルギー) から  $E_{d,i}$  ( $p_d$  のエネルギー) を差し引き、それを後部残響エネルギー(分散音エネルギー)  $E_{r,i}$  とする：

$$E_{r,i} = E_i - E_{d,i}, \quad (2)$$

- ⑤  $i$  番目のフレームの短時間 SOR <sub>$i$</sub> ：

$$SOR_i = 10 \log_{10} \frac{E_{d,i}}{E_{r,i}} \text{ dB}, \quad (3)$$

を算出し、これを全フレームで平均したものを文章全体の SOR とする：

$$SOR = \frac{1}{N} \sum_i^N SOR_i, \quad (4)$$

ここで、 $N$  は全フレーム数。尚、アナウンス中の息継ぎや間合いの関係で、聴取すべき直接音相当成分が殆ど存在せず、SOR <sub>$i$</sub>  が極端に小さくなってしまふことを避けるため、 $E_{d,i}$  に閾値を設定し、閾値以下であれば式(4)の平均処理から除外するようにした。

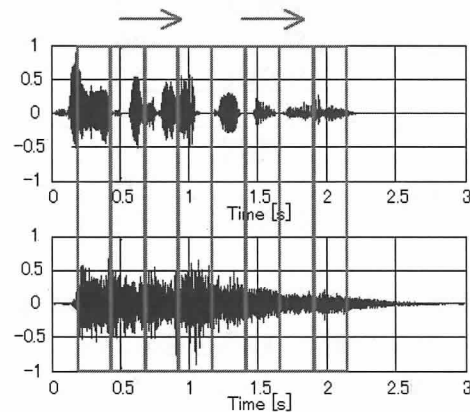


図 4 音声全体の SOR (上: 直接音相当, 下: 拡声音)

### 4. 実際の施設における測定

4 つの施設において、拡声音が可能な限り聴き易くなったと音響調整技術者の聴取で判断されるまで音響システムの調整を行った。調整前後の STI 及び SOR は表 1 のとおりである。尚、本報告における STI は、Schroeder の方法[14]によりインパルス応答から算出した MTF (Modulation Transfer Function)<sup>4</sup>を用いて、例えば文献[15]にあるような旧来の手法で求めたものであり、最新の手法[4]とは若干異なる。また、暗騒音による MTF の補正は行っていない。

各施設とも、音響調整により拡声音の聴こえ方が変化したにもかかわらず、STI は全く変化しないか、変化

<sup>4</sup> STI 算出の元になる MTF の測定は、変調ノイズによる手法とインパルス応答から数値演算でもとめる手法とがあり、規格[4]では変調ノイズ法を推奨している。Pathware が完全にインパルス応答だけで表現できる場合は、どちらの手法でも MTF に差が無いが、後述のように非線形な処理が含まれる場合、変調ノイズ法による MTF のみ非線形の影響が現れる。よって様々な非線形処理を搭載した拡声システムの STI を算出する場合、変調ノイズ法による MTF から算出することが望ましい。しかし、実際の音響調整の現場は時間的制約が多く、悠長に時間をかけた音響測定は許されにくい。それゆえ短時間に測定が可能なインパルス応答法による STI 算出しが事実上なされていらない。この現実を鑑み、本報告では STI の算出はインパルス応答法による MTF に基づくものに限定する。

したとしても  $JND$  ( $\approx 0.03$ ) [16] 以下であった。しかし、 $SOR$  は音響調整により数 dB の改善が見られることから、 $STI$  では測定が難しい音響調整の効果を  $SOR$  で表現できる可能性があると思われる。

表 1 調整前後の  $STI$  と  $SOR$

施設 (床面積)	調整前		調整後		変化量
	$STI$	$SOR$ [dB]	$STI$	$SOR$ [dB]	
小部屋 (30.9m <sup>2</sup> )	$STI$	0.70	0.70	0	
	$SOR$ [dB]	-7.2	-5.7	+1.5	
ホールロビー (300 m <sup>2</sup> )	$STI$	0.68	0.69	+0.01	
	$SOR$ [dB]	-5.0	-1.2	+3.8	
音楽ホール (313 m <sup>2</sup> )	$STI$	0.66	0.65	-0.01	
	$SOR$ [dB]	-1.4	1.2	+2.6	
体育館 (1,600 m <sup>2</sup> )	$STI$	0.49	0.47	-0.02	
	$SOR$ [dB]	-6.6	-6.0	+0.6	

### 5. 聴取実験： $SOR$ と「聴き取りにくさ」の関係

$SOR$  の数値と聴感との関係を明確にすることで、音響調整の結果、聴感的にどう変化するかを  $SOR$  から予測できるようにしたい。ここで通常の音響システムは、ある程度の  $STI$  が確保されるよう事前の音響設計に基づいて設置されていることから、よほど劣悪な残響環境でない限り、音響調整前でもそれ相当の  $STI$  は確保されていることが多い。そこで音声伝送性能 ( $STP$ ) [17] のクラス判定の元になる「聴き取り間違い」と「聴き取りにくさ」のうち、「聴き取り間違い」はあまり生じていないと考えることにし、ここでは音響調整により「聴き取りにくさ」がどのように変化するかを、以下に述べる聴取実験を通して調べた。

#### 5.1. 聴取者

聴取者は男性 31 名、女性 3 名の会社員で計 34 名 (20 代 8 名、30 代 11 名、40 代 8 名、50 代 6 名、60 代 1 名) で、平均年齢は 38.4 歳。各人の聴力は、年 1 回の健康診断で行われる聴力検査で正常の範囲内であった。尚、実験開始前に内容、危険性、個人情報の取り扱い等について十分な説明を行い、各人の自由意志に基づいて実験に参加する旨の同意書を全員から得ている。

#### 5.2. 実験用刺激の製作

##### 音源

非常時に拡声されるアナウンスの一部「職員の指示に従い、落ち着いて避難してください」を元音源  $s$  とした。将来のデータ比較のため、男声  $s_m$ 、女声  $s_f$  の 2 音源 (発話速度は男声：6.6 [mora/s]、女声：6.7[mora/s]) を用意した。

### Pathware：室のインパルス応答 $h$

Pathware のうち音源を拡声する空間に相当する特性として、表 2 に示すような残響時間  $T_{60}=0.2s, 0.5s, 0.9s, 1.4s, 2.6s$  の 5 つのインパルス応答  $h_1 \sim h_5$  を用意した<sup>5</sup>。これらは表 1 の体育館のインパルス応答を元に、減衰カーブを調整して作成したものである。

表 2 インパルス応答の仕様

	$T_{60}$ [s]	$STI$	$D_{50}$ 値[%]	$C_{50}$ [dB]
$h_1$	0.2	0.89	95.0	12.8
$h_2$	0.5	0.76	71.5	4.0
$h_3$	0.9	0.61	48.9	-0.2
$h_4$	1.4	0.51	33.6	-3.0
$h_5$	2.6	0.40	21.5	-5.6

### Pathware：音響調整の模擬

室のモード周波数の帯域を減衰させるという音響調整を模擬するため、図 5 の特性 (+14.5dB @170Hz, +4dB @620Hz) を持つフィルタ  $g$  を用意した<sup>6</sup>。室のインパルス応答  $h_1 \sim h_5$  に  $g$  を畳み込んだものが、室のモードが重畳されている状態を模擬していることから、 $h_i * g$  が音響調整前の pathware を、 $h_i$  が音響調整後の pathware を表している、where,  $i=1 \sim 5$ 。本報告に関する実験用刺激は、上記の男声  $s_m$  に pathware 特性 (調整前： $h_1 \sim h_5 * g$  の 5 つ、調整後： $h_1 \sim h_5$  の 5 つ) を畳み込んだものである (表 3)。

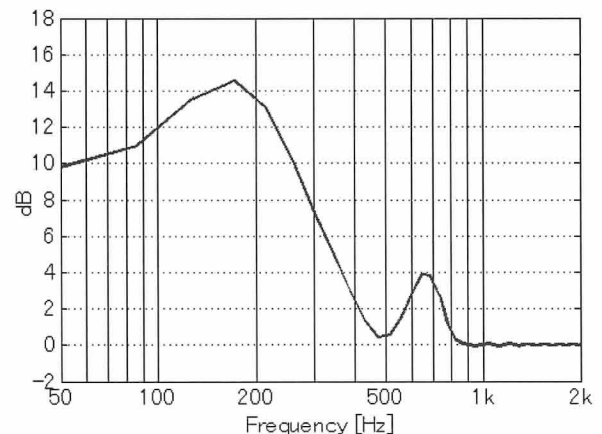


図 5 室のモードを模擬したフィルタ  $g$  の周波数特性

<sup>5</sup>  $T_{60}$  は 0.5k, 1k, 2kHz の 1/1oct. 帯域における  $3T_{20}$  を平均したものの。

<sup>6</sup> (中心周波数  $Q$ ) = (80Hz, 0.79), (100Hz, 0.99), (120Hz, 1.19), (180Hz, 1.78), (280Hz, 2.77), (620Hz, 6.12) の 6 つのフィルタを合成したものの。

表 3 音響調整前後の pathware

	pathware 特性				
音響調整前	$h_1 * g$	$h_2 * g$	$h_3 * g$	$h_4 * g$	$h_5 * g$
音響調整後	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$

男性  $s_m$  の他に女声  $s_f$  を用いた実験、更に、特殊な拡声として pathware 上に非線形処理を施した場合：

- ・音響分析ソフトウェア Praat[18]の PSOLA アルゴリズムにより発話速度を変化させたもの、
- ・故意に波形をクリップさせ歪ませたもの、
- ・文節ごとに無音区間を持たせた零挿入[19]、

の実験も併せて行った。このとき、pathware 上の非線形処理はインパルス応答には表れない。Pathware のインパルス応答に表れるものは、上述の  $h_1 \sim h_5$  及び  $g$  だけである。これら非線形処理を施した場合の結果の一部は後に示すが、詳細は別報[20][21]を参照されたし。

### 5.3. 実験装置及び手続

聴取者は防音室（容積=2.53m×2.53m×1.95m）内に置かれた椅子に着座しており、パソコンからオーディオ・インタフェース（EDIROL UA-25EX）を介した耳覆い型（circum-aural）イヤホン（Sennheiser HDA-200）で刺激が与えられる。被験者は各刺激に対して「聞き取りにくさ」を「聞き取りにくくない、やや聞き取りにくい、かなり聞き取りにくい、非常に聞き取りにくい」の4段階[22][23]で評価し、パソコン画面上の該当するボタンをクリックして回答した。本実験に先立ち残響時間の一番短い  $h_1$  を畳み込んだ男声  $h_1 * s_m$ 、女声  $h_1 * s_f$  を交互に3回ずつ計6回再生し、その間に聴取者自身で適切な音量に調整するよう指示した。尚、何を以って聞き取りにくいと感じるかは人によって異なると思われることから、残響時間が最長の  $h_5$  を畳み込んだ音声を事前に聴かせて「聞き取りにくさ」のダイナミックレンジを全聴取者で共通化してしまうことにならざるような操作は行っていない。

本実験において聴取者は、音響調整の前後を模擬した刺激と非線形処理を施した刺激の合せて160刺激を聴取し「聞き取りにくさ」を判断した。実験に要した時間は、事前の説明等を除いて一人当たり約30分であった。

### 5.4. 実験結果

理想的な音声伝送は100%聞き取り難くない状態であるべき、という森本ら[22]の主張に倣い、各刺激に対

する「聞き取りにくくない」以外の回答数（「やや、かなり、非常に」聞き取りにくい」の合計）の割合を、その刺激の「聞き取りにくさ[%]」と定義した。

図6は、各刺激の STI と「聞き取りにくさ」を表したものである。各刺激は音源に pathware の特性が重畳されたものだが、STI はその pathware のインパルス応答（pathware 特性の線形成分）から求めたものである。また、本報告では詳細には触れないが、非線形処理（女声の場合、発話速度変換、クリッピング、零挿入）を施した場合の結果も併せて記している。非線形処理を施した場合のインパルス応答は、処理を施していない場合の pathware と同一だとみなし、そのインパルス応答から STI を算出している。

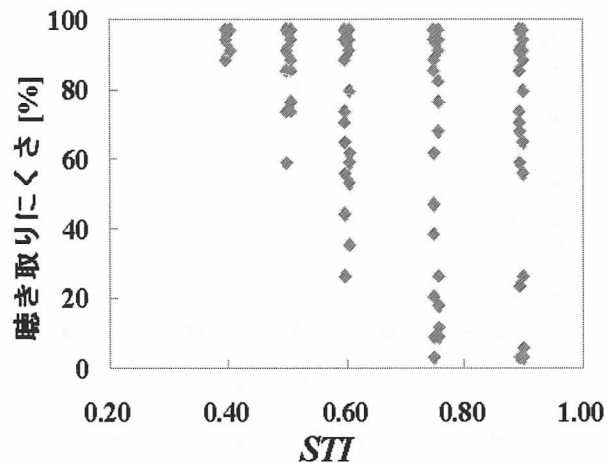


図 6 「聞き取りにくさ」 vs. STI

STI=0.4 のような悪条件では、音響調整や非線形処理を施しても殆どの方が聞き取りにくいと判断しているが、STI の値が高い領域では、同じ STI であっても音響調整や非線形処理により「聞き取りにくさ」が大きく変化していることが分かる。

次に非線形処理のデータを省いて音響調整前後の男声に対する実験結果のみを抜き出し、SOR と「聞き取りにくさ」の関係を表したのが図7である。「聞き取りにくさ」[%]を  $z$  値（標準得点[24]、標準型[25]）に変換すると、決定係数（寄与率）の高い「聞き取りにくさ」の  $z$  値から SOR への回帰直線を得た（図8）：

$$z = -0.33SOR - 1.4, \quad (5)$$

ここで、決定係数（寄与率） $R^2=0.9717$  であった。

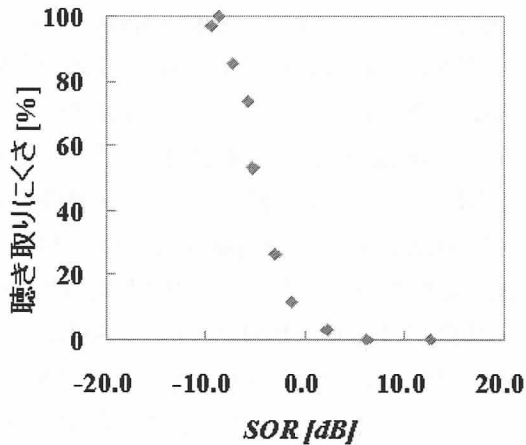


図7 「聴き取りにくさ」 vs. SOR :  
男声, 音響調整前後のみ

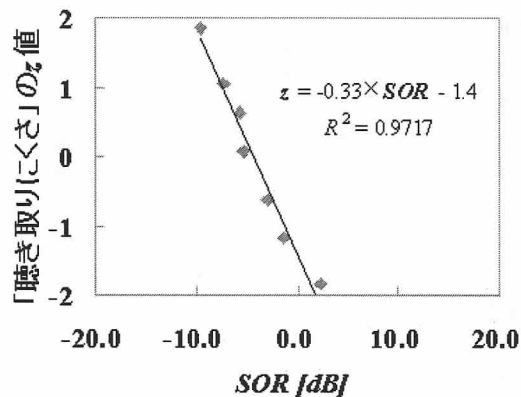


図8 「聴き取りにくさ」 z値 vs. SOR :  
男声, 音響調整前後のみ

得られた  $z$  値の式から「聴き取りにくさ (LDR) <sup>7)</sup> [%]の近似式[17]に変換すると

$$LDR = \frac{1 \pm \sqrt{1 - \exp\left(-\frac{2(-0.33SOR - 1.4)^2}{\pi}\right)}}{2} \times 100\% \quad (6)$$

ここで, 負号は  $SOR \geq -4.3$  のとき.

## 6. 「聴き取りにくさ」による音響調整結果の評価

式(6)を用いて先に測定した  $SOR$  (表 1) を「聴き取りにくさ(LDR)」に変換したのが表 4である。これまで, 音響調整により拡声音の聴こえが変化しても, その効果を  $STI$  の数値変化で表現できていなかったが,  $SOR$  測定を元に, どれだけ「聴き取りにくさ」が改善するかを示すことが出来るようになった。

<sup>7)</sup> LDR: Listening Difficulty Rating

表 4 調整前後の「聴き取りにくさ(LDR)」[%]

	調整前	調整後	改善量
小部屋	83.0	67.7	+15.3
ホールロビー	59.1	15.4	+43.7
音楽ホール	17.0	3.3	+13.7
体育館	77.5	71.1	+6.4

## 7. 議論

以上のように, これまで  $STI$  ではうまく表現できていなかった拡声システムの音響調整の効果を,  $SSS$  の評価量として提案していた  $SOR$  を用いた「聴き取りにくさ」で評価することが出来るようになった。ここで, これまで提案されている明瞭性に関する指標と比較しながら  $SOR$  の位置付け等について検討する。

### 7.1. 従来の明瞭性指標と $SOR$ の関係

$SOR$  は当初の(1)式及びそれを变形した(3), (4)式から分かるとおり, 「直接音及び 50ms までの残響音」のエネルギーと「それ以降の残響音」のエネルギーの比となっている。これは, 従来から使われている  $D$  値[26], 直接音成分を 50ms までとした  $C_{50}$  値[27][28], 残響指数  $RI$ [29] (残響対初期音比[30][31],  $H$ [32]), 反射比 (reflections ratio)[33]または直接音成分対散音エネルギーレベル差[29]と呼ばれる  $R$  などと同じ概念を表している:

$$SOR = C_{50} = R \quad (7)$$

$$= -RI = -H \quad (8)$$

$$= 10 \log_{10} \frac{D}{1-D} \quad (9)$$

ただし  $SOR$  との等号が成立するのは pathware が線形システムだけで構成されている, つまり pathware がインパルス応答だけで完全に記述される場合の話である。

例えば Thiele[26]は  $D$  値を

$$D = \frac{\int_0^{0.05s} I(t) dt}{\int_0^{\infty} I(t) dt} \quad (10)$$

と定義しているが, 音源から発せられたインパルス (時間幅約 1ms) をマイクロホンで收音し, その音圧を二乗したものを瞬時インテンシティ  $I(t)$  とみなしていることから,  $D$  値はインパルス応答  $h(t)$  から定義されることがわかる。その他  $C_{50}$  値や残響指数  $RI$  など全てインパルス応答から算出することを基本としている。通常, インパルス応答は相互相関法, swept sine (TSP)法で測定されることから[34], pathware が非線形成分を持

っていたとしても、得られるインパルス応答は pathware の線形成分しか表現していない<sup>8</sup>。ところが、このとき拡声された音声は pathware 上の非線形処理によりそれ相応の影響を受けていることから、その拡声音に基づく SOR は pathware 上の非線形成分の影響を評価することが可能である。

このように SOR は  $C_{30}$  値 (=RI) を特別な場合 (pathware が線形成分だけから成っているという特別な場合) として含んでいることになる。

## 7.2. Pathware 上の非線形処理

拡声システムにおける非線形処理として代表的なものはスピーカ保護のためのリミッタであろう。通常運用の範囲であればリミッタが作動することは稀だが、単純に振幅をクリップさせることで歪んだ拡声音となる場合がある。他に、拡声システムで使用されることはまだ多くは無いが、音声信号全体に渡って許される最大振幅 (例えばデジタル・フルスケール) に振幅を伸長するハイパー・コンプレッション[35]という処理がある。これは少しでも大きなラウドネスを提供することを目的とした処理で、特に音楽業界、FM ラジオ業界では盛んに用いられているとのこと。更に前出の SSS 処理は、今のところ限られた性能ではあるが、残響下での拡声音の明瞭性を向上させる技術として期待される。

このように拡声の pathware が非線形処理を持つことは、今後さほど珍しいものではなくなると思われる。そのためか、非線形処理された音声の STI[36]や圧縮音声の STI[37]の試みがいくつか報告されてきている。今後、SOR のように非線形処理による拡声音の違いを評価することができる指標の必要性が高くなってくると思われる。

## 7.3. 実際の音声を用いた測定について

今回提案した SOR はある特別なアナウンスを用いて測定及び聴取実験を行ったものであるので、当然、別のアナウンスで測定すると、同じ環境であっても SOR の値は異なってくるだろう。また、その数値と「聴き取りにくさ」の関係も、実際は概ね式(6)に近いものだろうが、厳密にはアナウンス一つ一つについて「聴

き取りにくさ」との回帰式が異なるであろうことも容易に想像できる。

実際の音声を用いた明瞭性評価を検討する場合、通常なら音声学的にバランスの取れた単語セット[38]を用いるなどして、様々な音声単語に対する一般化された明瞭性の度合いを議論するのが正攻法であろう。また、使用する元音声は限りなくクリアで訛りも無く誰もが「聴き取りにくい」ものであり、そのような音声を使って、音源である発信系や受聴者の能力を含む受信系と切り離して考えることが出来る伝送系の性能のみを評価するという姿勢[17]も、至極当然であろう。ただ、実際の拡声の現場で単語だけを拡声することは皆無であり、文章を拡声したときの明瞭性をどのように考えればよいか早く知りたいのも事実である。音声伝達の評価尺度について、個々のフレーズ、各音節の伝達特性にまで踏み込んだ尺度の必要性が過去指摘されており[39]、単語理解度との対応に優れた評価指標の提案[40]がなされるなど着実に研究は進んでいるが、更に文章ではどのようになるのか興味のあるところである。

文章の明瞭性など、あまりにも検討項目が多すぎて、どの状況でも普遍的に適用できる明瞭性指標など得られないかもしれない。しかし、例えば非常放送設備のように用途が限定されるような場合、普遍的な明瞭性でなくとも、限られたアナウンスの聴こえ方を具体的に調査することで、その pathware を評価できるのではないだろうか。そこで、やや勇み足的な試みではあったが、敢えて単語ではなくアナウンスを用いた測定及び聴取実験を行ってみたいわけである。今回用いたアナウンスは非常放送用のものであったが、更にいくつかこのシステムに特化したアナウンスを用いた測定及び聴取実験を行うことで、この限定用途システムの明瞭性を特徴付けることができるのではないかと、少なくとも、広範な学術的調査に基づくアナウンスの一般的な明瞭性評価手法が確立するまで、一つの目安として活用できるのではないかと考えている。

## 8. 終わりに

実際の拡声音を用いて測定した直接音相当成分と分散音の比である SOR を用いて、拡声システム設置後の音響調整の効果を数値で評価することを試みた。聴取

<sup>8</sup> 複数の周波数を同時に含む測定信号の場合、それらの差周波数や和周波数の歪成分の現れ方によっては、特徴的なインパルス応答となる場合があるが、それらの影響は小さいとみなしている。



実験より、ある特定のアナウンスでの結果ではあるが、SOR の数値から「聴き取りにくさ」を推定できることが分かり、これまで STI での評価が難しかった音響調整の効果を、「聴き取りにくさ」の改善量として示すことができた。また、従来から提案されているインパルス応答に基づく明瞭性指標との比較から、SOR はイコライザなどの線形処理ベースの音響調整はもとより、拡声システムに非線形な処理が含まれているときの明瞭性評価でも利用できそうである。

拡声システムのメーカーは、機器もさることながらそのシステムによる「音」を販売している。よって、その売り物である「音」がどのような品質であるか、適切な手法で評価し利用者に提示せねばならない。そこで、これまで提案されてきた様々な手法による評価結果を示してきたわけだが、果たしてそれらは、利用者の立場に立った、そして利用者にとって分かり易いものであったらどうか。実際に設計や調整を行うのはある程度の知識と経験をもった専門家ではあるが、「音」の品質評価にあたっては「音」のサービスを受取る当の利用者が不在とならぬよう、「音」を提供する側として今後留意してゆきたい。

## 謝 辞

非線形信号処理について関西大学の梶川嘉延教授に、聴取実験のデータ処理について神戸大学の佐藤逸人助教ならびに東海大学の程島奈緒講師に助言をいただいた。感謝します。

## 参考文献

- [1] 栗栖清浩, 尾本章, “音声情報伝達向上のためのデバイス - スピーカー -,” 騒音制御, vol. 33, no. 4, pp. 302-306, 2009.
- [2] 改訂新版 PA 音響システム, 藤岡繁夫 (編), 工学図書, 東京, 1996.
- [3] McCarthy B, Sound Systems: Design and Optimization, 2nd. Ed., Focal Press, Oxford, 2010.
- [4] IEC 60268-16, "Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index," Edition 4.0, 2011.
- [5] 佐藤洋, “1. 解説: なぜ音声伝送品質の評価は必要か,” シンポジウム 音声伝送品質の評価と設計 現状と今後 - 建築学会音声伝送 SWG 活動成果報告 (1999~2002), pp. 1-2, 1月, 2003.
- [6] Campbell D, "STI - Where did it come from and what does it do?" Syn-Aud-Con Newsletter, vol. 35, no. 1, pp. 10-17, 2007.
- [7] Arai T, et al., "Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking," Acoust. Sci. & Tech., vol. 28, no. 6, pp. 438-441, 2007.
- [8] 荒井隆行 et al., "音声の定常部抑圧の残響に対する効果," 音講論集, 2001 年秋, pp. 449-450, 2001.
- [9] Arai T, et al., "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," Acoust. Sci. & Tech., vol. 23, no. 4, pp. 229-232, 2002.
- [10] Hodoshima N, et al., "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments," J. Acoust. Soc. Am., vol. 119, no. 6, pp. 4055-4064, 2006.
- [11] Bolt R H, MacDonald A D, "Theory of speech masking by reverberation," J. Acoust. Soc. Am., vol. 21, no. 6, pp. 577-580, 1949.
- [12] Nabelek A K, Robinette L, "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am., vol. 63, no. 1, pp. 187-194, 1978.
- [13] Steeneken H J M, Houtgast T, "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am., vol. 67, no. 1, pp. 318-326, 1980.
- [14] Schroeder M R, "Modulation transfer functions: Definition and measurement," Acustica, vol. 49, pp. 179-182, 1981.
- [15] 中島立視, “音声の明瞭度指数 (STI) の測定,” 日音学誌, vol. 49, no. 2, pp. 103-110, 1993.
- [16] Bradley J S, Reich R, Norcross S G, "A just noticeable difference in C50 for speech," Applied Acoustics, vol. 58, pp. 99-108, 1999.
- [17] 日本建築学会環境基準 AIJES-S0002-2011 都市・建築空間における音声伝送性能評価規準・同解説, 2011.
- [18] Boersma P, Weenink D, Praat: doing phonetics by computer, <http://www.fon.hum.uva.nl/praat/>
- [19] Arai T, "Padding zero into steady-state portions of speech as a preprocess for improving intelligibility in reverberant environments," Acoust. Sci. & Tech., vol. 26, no. 5, pp. 459-461, 2005.
- [20] 栗栖清浩 et al., “拡声音のオーバーラップマスキング量 SOR と「聴き取りにくさ」の関係 - 拡声システムの調整結果を評価するツールへの応用 -,” 音講論集, 2012 年春, 2012 (in press) .
- [21] 中村進 et al., “拡声音のオーバーラップマスキング量 SOR と「聴き取りにくさ」の関係 - 発声系 (歪み・話速) と伝送系が変化する事例 -,” 音講論集, 2012 年春, 2012 (in press) .
- [22] 森本政之, 佐藤洋, 小林正明, “2. 音声伝達性能の主観評価指標としての聴き取りにくさ,” シンポジウム 音声伝送品質の評価と設計 現状と今後 - 建築学会音声伝送 SWG 活動成果報告 (1999~2002), pp. 3-8, 1月, 2003.
- [23] 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.
- [24] 心理学のためのデータ解析テクニカルブック, 森敏昭, 吉田寿夫 (編), p. 32, 北大路書房, 京都, 1990.
- [25] 藤沢偉作, 楽しく学べる統計教室, p. 56, 現代数学社, 京都, 1978.
- [26] Thiele R, "Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen," Acustica, vol. 3, pp. 291-302, 1953.
- [27] Bradley J S, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am., vol. 80, no. 3, pp. 837-845, 1986.
- [28] Reichardt W, Alim O A, Schmidt W, "Definition und

Meßgrundlage eines objektiven Maßes zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit bei Musikdarbietung," *Acustica*, vol. 32, pp. 126-137, 1975.

- [29] 永田穂 (編著), 建築音響 (音響工学講座 3), 日本音響学会 (編), p. 58, コロナ社, 東京, 1988.
- [30] Beranek L L, Schultz T J, "Some recent experiences in the design and testing of concert halls with suspended panel arrays," *Acustica*, vol. 13, pp. 307-313, 1965.
- [31] Schultz T J, "Acoustics of the concert hall," *IEEE Spectrum*, June, pp. 56-67, 1965.
- [32] Kuttruff H, *Room Acoustics*, 4th. ed., p. 225, Spon Press, London, 2000 (著: クットルフ, 訳: 藤原恭司, 日高孝之, 室内音響学, p. 221, 市ヶ谷出版, 東京, 2003).
- [33] Latham H G, "The signal-to-noise ratio for speech intelligibility - An auditorium acoustics design index," *Applied Acoustics*, vol. 12, pp. 253-320, 1979.
- [34] 橋秀樹, 矢野博夫, 環境騒音・建築音響の測定 (音響テクノロジーシリーズ 8), 日本音響学会 (編), pp. 80-88, コロナ社, 東京, 2004.
- [35] Vickers E, "The loudness war - Do louder, hypercompressed recordings sell better?" *J. Audio Eng. Soc.*, vol. 59, no. 5, pp. 346-351, 2011.
- [36] Goldsworthy R L, Greenberg J E, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3679-3689, 2004.
- [37] Chen S, Payton K L, "Compensating the STI to predict intelligibility for amplitude-compressed speech," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 49-52, 2005.
- [38] 坂本修一 et al., "親密度と音韻バランスを考慮した単語理解度試験用リストの構築," *日音学誌*, vol. 54, no. 12, pp. 842-849.
- [39] 戸井田義徳, "空間内における音声情報伝達," *日音学誌*, vol. 51, no. 4, pp. 312-316, 1995.
- [40] 翁長博, "残響音場の音声理解度に対応する物理指標の提案," *日音学誌*, vol. 66, no. 3, pp. 97-104, 2010.