建築音響研究会資料
資料番号 AA2013-36

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？ ―

栗栖　清浩　（ＴＯＡ）
安　啓一　（上智大学）
荒井　隆行　（同上）
中村　進　（同上）

2013年8月2日

日本音響学会建築音響研究委員会

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？ —

The relationship between physical measures derived from broadcasted speech and its listening difficulty:

Should speech itself be evaluated, not the path it travels?

栗栖清浩†, 安 啓一‡, 荒井隆行‡, 中村 進‡*

KURISU, Kiyohiro†, YASU, Keiichi‡, ARAI, Takayuki‡, NAKAMURA, Susumu‡*

† TOA 株式会社　‡ 上智大学

† TOA Corporation　‡ Sophia University

内容概要

拡声音の明瞭性評価指数として著者らが提案した *SOR* (Speech to Overlap-masking Ratio) 及び音声の変動量に基づく指標について概説した. 従来の明瞭性指標は主に伝送系を評価するのに対し，音声の変動量に基づく指標は，受音点における拡声音のみを評価対象としていることから，受聴者寄りの視点に立った評価指標になっている. 従来の明瞭性指標は拡声音を提供する側には都合がよいが，このような聴取者のための明瞭性指標も必要ではないかという議論を展開した.

1. はじめに

拡声システムによる拡声音の明瞭性を客観的に評価する指標として，著者らは *SOR*（Speech to Overlap-masking Ratio）を提案し[1], その有効性を検証してきた[2][3][4]. *SOR* は残響環境下で拡声音を明瞭にするための定常部抑圧処理（Steady State Suppression 処理, 以下 SSS 処理と略記）[5][6][7]を数値評価するために提案されたものであったが, *STI*（Speech Transmission Index）[8]が不得意とするイコライザ調整による明瞭性改善の評価に（ある限定した条件の下で）使えることが分かった.

本論文では最初にこれまで著者らが提案及び検証してきた *SOR* と「聴き取りにくさ」の割合（Listening Difficulty Rating, 以後 *LDR* と略記）[9][10]の関係について概説し, 従来の明瞭性指標及び *SOR* が何を評価対象としているのかについて考察した. そして拡声音を提供する側は図 1 に示すような伝送系（pathware[11]）を主な評価対象としており，従来の明瞭性指標は必ずしも受聴者の視点に立った評価指標になっていないのではないかという議論を展開した. そこで受聴者寄りの評価指標の試みとして，拡声音の変動特性に基づく指標を考案し[12], *LDR* との対応関係を確認した.

最後に，受聴者の視点に立った明瞭性指標が受聴者の権利を守り，世の中の明瞭性向上に貢献する可能性について触れた.

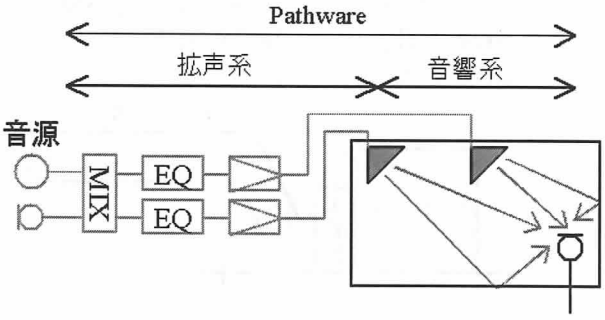

図 1 伝送系(Pathware) = 拡声系＋音響系[2][11]

2. *SOR* 概要

残響のある場でも明瞭に拡声する手法を評価する目的で提案された *SOR* であったが, ある条件の下で *LDR* の推定にも使えることが分かった. 以下に *SOR* の概要を記す.

2.1. SOR

SOR は SSS 処理の効果を数値で表現しようとして提案された[1]. SSS 処理は残響下での拡声音の明瞭性向上を目的としたもので，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）へのオーバラップ・マスキング（OLM）を軽減し，子音の知覚を改善することで明瞭性の向上を目指したものである（図 2）.

* 現在，ティアック株式会社勤務

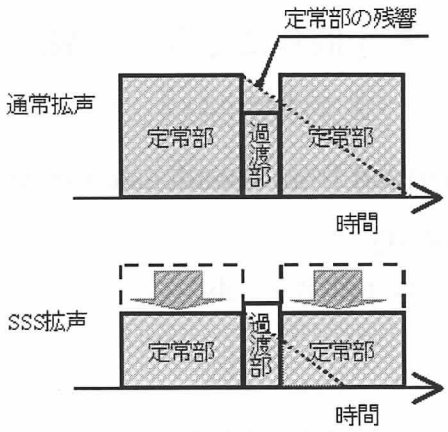

図 2 定常部抑圧(SSS)処理の概念図[2]

SSS 処理を施した信号を拡声すると，ある一定の範囲の残響条件下において単音節明瞭度の向上がみられた[7]が，当初は SSS 処理によりどの程度 OLM が軽減されたのかの検証がされていなかった．そこで SSS 処理の物理的な効果を評価するため *SOR* が提案された[1].

これは，受音点においてある単音節を観測するとき，ターゲット区間 *T* [s]内の単音節の直接音成分エネルギー *S* と，ターゲット区間に先行する音声の残響エネルギー *N* による *SN* 比で以って *SOR* としたものである（図 3 及び式(1)）．測定の結果，SSS 処理により *SOR* が改善すると単音節明瞭度も改善するという傾向が確かめられた[1].

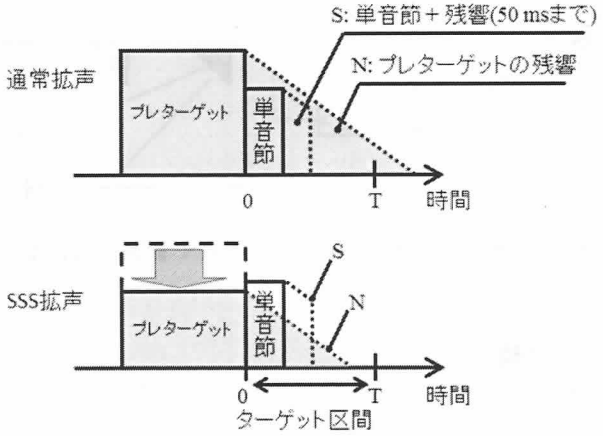

図 3 SSS 処理による *SOR* 改善の概念図

この時，直接音を補強し明瞭性に寄与すると言われている 50 ms までの残響[13][14]も直接音成分に含めている：

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

ここで，

s ：単音節部の信号，
p ：ターゲット区間より先行する信号，
h ：音響系のインパルス応答，
h_{50} ：50 ms までの h，
T ：ここでは 150 ms とした．

SSS 処理は入力波形の包絡線に応じて時々刻々と振幅を変化させるという非定常な信号処理であるので，振幅が定常な試験音を用いて測定したインパルス応答は，SSS 処理を同定することができない．仮に SSS 処理有り／無しの拡声システムのインパルス応答を比較しても，まったく同じか，又は振幅は異なるものの相対的形状は全く同じ応答になるものと思われる．よって C_{50},[1][15][16]や *D* 値[17]，そして Schroeder[18]の方法によりインパルス応答から導出した *MTF*（Modulation Transfer Function）に基づく *STI* 等の明瞭性指標は，原理的に SSS 処理の有無で数値は変化しない[2].

これに対し *SOR* は，包絡線が変化する実際の音声を試験音として用いることから，その音声に特化した結果ではあるものの，SSS 処理の効果を数値で示すことができる．尚，定義式(1)にはインパルス応答が含まれているが，特定の音声に含まれる包絡線変化及び SSS 処理の効果は音源（*p* 及び *s*）に含まれている．

2.2. 設置後の調整:「聴き取りにくさ」を改善する作業

拡声システムを導入する現場は様々だが，導入前の適切な音響設計により，多くの現場において聴取エリアの音圧は目標値をクリアしている．しかし，十分な音圧確保で「聴き取り間違い」があまり生じてなくとも，音響設計時に考慮されなかった多くのパラメータにより聴き取りにくい拡声となっていることがある．まさに「単語了解度がほぼ 100%に近い場合でも，『聴き取りにくさ』は 0%から 50%程度まで分布する」[19]という状況であり，よってこの「聴き取りにくさ」を改善することが拡声システム設置後に行う調整作業の重要な目的の一つとなっている．

調整方法の具体例として，イコライザを用いて伝送系の周波数特性を平坦に近づけることがある．例えば室が共鳴している場合，拡声システムのイコライザにより特定の周波数帯での音圧上昇を抑えて拡声する．これにより多くの場合「聴き取りにくさ」が改善されるのだが，以前より指摘されているとおり[20]，イコライザでシステムを調整しても *STI* はほとんど変化しない[2]．現状では明瞭性の数値評価は

---

[1] 音楽の明快さ，清澄さの指標として提案された「Klarheitsmass *C*（clarity index）」[15]は直接音成分の積分時間を 80 ms としていたが，ISO[16]において C_{te} としてまとめられ，*te* は 50 ms 又は 80 ms のどちらかを用いることとなった．

[2] ある帯域のゲインをイコライザで暗騒音が無視できないほど低減すると *STI* は悪くなるが，暗騒音が無視でき

主に *STI* を用いるよう発注者（施主）から指定されるので，調整によりいくら「聴き取りにくさ」が改善しても，その成果を客観的な数値で発注者に説明できないことがある．様々な音場における *STI* と「聴き取りにくさ」の割合 *LDR* の対応関係は実験的に確かめられている[19][21][22]ものの，このようなイコライザ調整による *LDR* の改善を *STI* と関連付けて説明することは難しい．

そこで，2.1 で述べたように *STI* が変わらない状況であっても単音節明瞭度の向上を *SOR* の改善と対応付けることができたのと同様に，イコライザ調整による *LDR* の改善を *SOR* の変化で説明できないか調査した．

2.3. *LDR* vs. *SOR*：単一アナウンスの場合[2][3][4]

2.1 に記したように単音節明瞭度がある程度確保されている拡声の現場の状況を踏まえ，単音節でなく，ある実際のアナウンスを試験音としたときの *LDR* を調査した．これに伴い，*SOR* の初出[1]では直接音成分 *S* 及び妨害成分 *N* はターゲット区間 *T* [s] 内に含まれるそれぞれのエネルギーであったが，セグメンタル *S/N*（たとえば文献[23]の SNR_{seg}）と同様に，アナウンスを連続した区間に分割し，各区間における *SOR* を平均してアナウンス全体の *SOR* とみなすこととした．

$$SOR_i = 10\log_{10}\frac{S_i}{N_i} \ \text{dB}, \qquad (2)$$

$$SOR = \frac{1}{N}\sum_{i}^{N} SOR_i \ \text{dB}, \qquad (3)$$

ここで，S_i，N_i は区間 i（全区間数 N）における直接音成分と妨害音成分のエネルギーである．尚，*SOR* が極端に小さくなることを避けるため，S_i に閾値を設定し，閾値以下であればその区間は式(3)の平均処理から除外することとした．

聴取実験の条件は次のとおりである．男声アナウンス「職員の指示に従い，落ち着いて避難して下さい」に 5 つの音響系（T_{60}＝0.2s, 0.5s, 0.9s, 1.4s, 2.6s）を畳込み，室が共鳴していない場合の拡声音刺激 5 つを用意した．更にそれら 5 つと室の共鳴を模擬したフィルタを畳込み，室が共鳴している場合の拡声音刺激 5 つを用意した．これら 10 刺激を，日本語を母語とする 34 名（♂：31，♀：3，年齢：20 代～60 代，平均年齢：38.4 歳）に，防音室において耳覆い型ヘッドホンを介してランダムな順番で呈示し，「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）を回答させた．その結果，図 4 に示す通り，各刺激の *SOR* と，各刺激に対する評価 2～4 の回答数の割合（「聴き取りにくさ」の割合 *LDR*）を *Z* 値に変換した Z_{LDR} との関係から，高い決定係数[24]（R^2=0.97）の回帰直線：

$$Z_{LDR} = -0.33\,SOR - 1.4 \qquad (4)$$

を得た．ここで，割合 *LDR*（確率密度関数）から *Z* 値（分布関数）への変換は，Williams の近似式[25][26]を用いた．また，$Z_{LDR}=\pm\infty$（*LDR* が 0%又は 100%に相当）となる刺激は回帰分析から除いた．

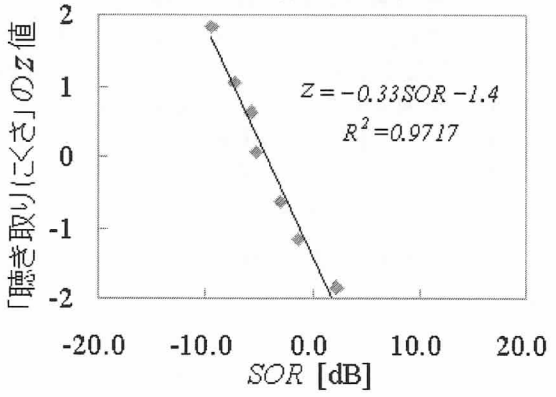

図 4　Z_{LDR} vs. *SOR*：単一アナウンス[2][3][4]

LDR は *Z* 値から割合に変換することで，

$$LDR = \frac{1 \pm \sqrt{1-\exp\left(-\frac{2}{\pi}Z_{LDR}^{\ 2}\right)}}{2} \times 100 \ \%, \qquad (5)$$

ここで，負号は $Z_{LDR} \le 0.0$ のときである．

以上のことを応用すると，ある特定の男声アナウンスに特化した *LDR* ではあるものの，拡声の現場において，*STI* では確認できなかった調整の効果を，*SOR* の向上又は式(5)から推定される *LDR* の向上で示すことが可能になった（表 1）．

表 1 調整前後の明瞭性指標の変化例[3]

施設（床面積）		調整前	調整後	変化量
小部屋（30.9m²）	*STI*	0.70	0.70	0
	SOR[dB]	−7.2	−5.7	+1.5
	LDR[%](推定値)	83.0	67.7	-15.3
ホールロビー（300 m²）	*STI*	0.68	0.69	+0.01
	SOR[dB]	−5.0	−1.2	+3.8
	LDR[%](推定値)	59.1	15.4	−43.7
音楽ホール（313 m²）	*STI*	0.66	0.65	−0.01
	SOR[dB]	−1.4	1.2	+2.6
	LDR[%](推定値)	17.0	3.3	−13.7
体育館（1,600 m²）	*STI*	0.49	0.47	-0.02
	SOR[dB]	−6.6	−6.0	+0.6
	LDR[%](推定値)	77.5	71.1	−6.4

2.4. *LDR* vs. *SOR*：複数アナウンスの場合[27]

次に，複数アナウンスであっても，*SOR* と「聴き取りにくさ」が決定係数の高い回帰式で関係付けられるかどうか調査した．アナウンスは男声 10 種で，

る範囲のイコライゼーションであれば，*STI* は変化しないと言ってよい．

「ATR研究用音声データベース503文Bセット」[28]，「名古屋工大日本語音声データベース」[29]から5つずつ選んだ．これらに2.3と同様の方法で室の共鳴が無い場合と共鳴を模擬した場合の10種のインパルス応答を畳み込み，聴取実験用に100刺激を用意した．これら100刺激を日本語を母語とする聴力に問題ない聴取者50名（20代：16名，30代：12名，40代：10名，50代：10名，60代：2名）に，2.3と同じ方法で提示し「聴き取りにくさ」を評価させた．*SOR*を独立変数，Z_{LDR}を従属変数とし，アナウンス毎に求めた回帰式の決定係数を表2に示す．また，$Z_{LDR}=\pm\infty$となる刺激を除いた全刺激から，決定係数R^2=0.92の回帰直線：

$$Z_{LDR} = -0.26SOR - 0.30. \qquad (6)$$

を得た（図5）．

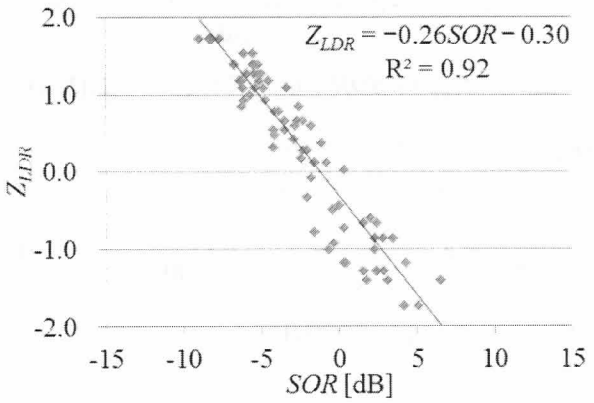

図5　Z_{LDR} vs. *SOR*：複数アナウンス [12]

決定係数は殆どのアナウンスで0.9以上（最大0.98），全刺激ではR^2 = 0.92であった．かなり高い決定係数ではあるが，たとえば*SOR* = 0 dBにおいて*LDR*が約1 SD（標準偏差）の範囲に分布している．*LDR*が1 SD異なると，音声伝送性能*STP*（Speech Transmission Performance）[19]が1クラス異なる場合があることから，このばらつきが実用上問題ないか更に検討が必要だと思われる．

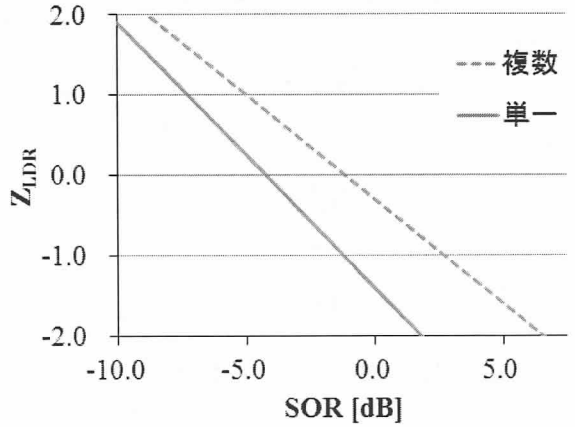

図6 回帰直線の比較：単一 vs. 複数アナウンス

尚，2つの回帰式：式(4)，式(6)を比べると，例えば「聴き取りにくい」と感じる人が殆どいないZ_{LDR} = −2.0（*LDR*≒2.0%）とするには，複数アナウンスは単一アナウンスよりも*SOR*が約5 dBも高くなければならない（図6）．これは，2.3で用いた単一アナウンスの周波数帯域に比べ，本節の10アナウンスの帯域が狭かったことによると思われるが，詳しい解析は今後の課題である．

3. 音声そのものを評価すべきでは？

以上の通り，拡声システムのイコライザ調整により音声の明瞭性が改善することを，*SOR*から推定される*LDR*の改善量で客観的に示すことができた．しかし，*STI*やC_{50}, *D*値，そして*SOR*等を測定する際，受音点における観測信号に加えて，音源（試験信号）を必要としており，これは人が明瞭性を評価するプロセスとは異なっている．以下，各種物理指標が評価しているものと評価の視点について考察する．

3.1. 各種物理指標が評価しているもの

まずC_{50}, *D*値について考える．これら2つの指標は共にインパルス応答から計算され，C_{50}は（50 msまでの）直接音成分とそれ以降のエネルギーの比，*D*値は全エネルギーに対する直接音成分の比であり，いずれも数値が高いと明瞭性が高い（たとえば文献[30]）．つまり，受音点における直接音成分の割合が高いと明瞭であることから，これらは，伝送系がどれだけ多く直接音成分を伝送できるかという伝送系の特性に注目した指標だと言える．

また*MTF*はModulation Transfer function[31]という名の通り，伝送系がどれだけ正確に信号の変調を伝送できるかを表していることから，それに基づく*STI*も，（算出過程において，音源の特性が若干考慮されてはいるものの）基本的に伝送系の特性を表しているものと考えられる．

他に，波形伝送でなくコーデック（Codec）の評価指標として，現在のところ最も洗練された手法である[23]というITU-TのPESQ（Perceptual Evaluation of Speech Quality）[32][33]があるが，これも伝送系への入力と出力の差からMOS（オピニオン平均）を推定するので，やはり結局は伝送系の特性を表しているのであろう．

*SOR*は，明瞭な音声（直接音成分）のエネルギーにどれだけ邪魔なエネルギーがオーバラップしているかを表しているが，それを調べる試験音は広帯域ノイズやスイープ音といった試験音ではなく，音声そのものである．よって，用いた音声に特化した結果ではあるものの，SSS処理や歪といった非線形処理が含まれても*SOR*の数値に反映される．

実際の音声を用いて明瞭性を評価する手法は他にも提案されている[34][35]．しかしそれらはVoIP（Voice over Internet Protocol）[36]のように音声が符号化されて伝送されるとき，従来の試験信号では適

切な測定ができないという理由で実際の音声を試験信号として用いているわけで，測定の目的は *STI* を求めることである．つまり結局は伝送系の特性を測定している．

3.2. 伝送系の評価だけで十分か？

このように明瞭性の指標として利用されているものは，その多くが伝送系の特性を評価するものだが，果たして受音点における明瞭性を議論するのに，伝送系の特性を評価するだけで十分であろうか？例えば $C_{50} = \infty$ dB，$D = 100\%$，$STI = 1.0$ などといった理想的な伝送系あったとして，それで受音点における音声が明瞭であると保障されるだろうか？それは，伝送系に入力される音声（音源）そのものが明瞭かどうかにも依存することだから，理想的な伝送系は受音点における音声が明瞭であるための必要条件ではあるが十分条件ではないだろう．たとえば，元々ひどく歪んだ不明瞭な音声が入力されると，理想的な伝送系は律儀にも歪んだ音声をそのまま受音点に伝送する．

3.3. 誰のための明瞭性評価？

明瞭性を評価する目的はもちろん明瞭な音声を届けること（の一助にする為）であるが，評価者の立場によって自ずと評価対象が異なってくると思われる．例えば，建築屋の立場からすると室の伝送特性に興味があり，また音響屋の立場からは拡声システムが明瞭に音声を伝送できるかどうかに興味があるであろう．そしてそれらを評価するのに C_{50}, D 値，*STI*, さらに PESQ といった伝送系を評価する指標は大変便利で有用だと思われる．しかし，音声を聴くのは聴取者であり，最終的には聴取者に届く音声が明瞭かどうか，という視点で明瞭性を評価することも必要ではないだろうか？

人が心理評価を下すまでのプロセス[37]において，評価対象となる信号は人の耳元に届く信号のみである，即ち，このプロセスに倣って聴取者の視点に立った評価指標を考える場合，評価対象となる信号は受音点で収録したものだけで，拡声システムに入力される音源信号は不要である．人は音源と受音点の音を比較するようなことをせず，受音点の音しか聴いていない．

4. 変動音解析の試み[12]

人が明瞭性を感じるプロセスと同様に，受音点で観測した信号だけを用いて，明瞭性に関係すると思われる物理量の導出を試みた．この時，伝送系のインパルス応答も分からなければ，そもそも音源がどのような信号であったかもわからない．つまり，受音点における直接音成分とそれ以降のパワーも不明，また，（インパルス応答を用いて推定される）拡声音を受音点信号から減算し暗騒音を推定することもできないという状況で，与えられるのは，それらがすべて重畳された信号のみ，という状況である．そこで，*MTF* を測定するとき音声のモデルとして変調ノイズを用いたこと，また，音声信号のうちある範囲の変調成分が明瞭性にとって重要であること[38]を参考に，受音点で観測した拡声音の変調に着目した解析を行った．尚，ここでは両耳処理を想定せず，1チャンネルで収音された信号を用いた．

4.1. 市販ソフトウェアによる変動量解析

図 7 は，例としてある音声アナウンスを市販ソフトウェア[39]を用いて解析したものである．横軸は 1/3 オクターブ帯域の中心周波数，縦軸はその帯域のラウドネスが変動している周波数を，等高線は変動の大きさを表している．この例では，音声の 1 kHz を中心とする 1/3 オクターブ帯域において，変動周波数 5.0 Hz で変動しているラウドネス変動量 *DFL*（Depth of Loudness Fluctuation）が約 4000 mDLF であることを示している．尚，用いたソフトウェア[39]では帯域別にラウドネスの包絡の周波数分析を行い，「変調の大きさ」ではなく独自の「変動量（単位 DLF 又は mDLF）」を算出しているので，ここではそれに倣った表現をしている．

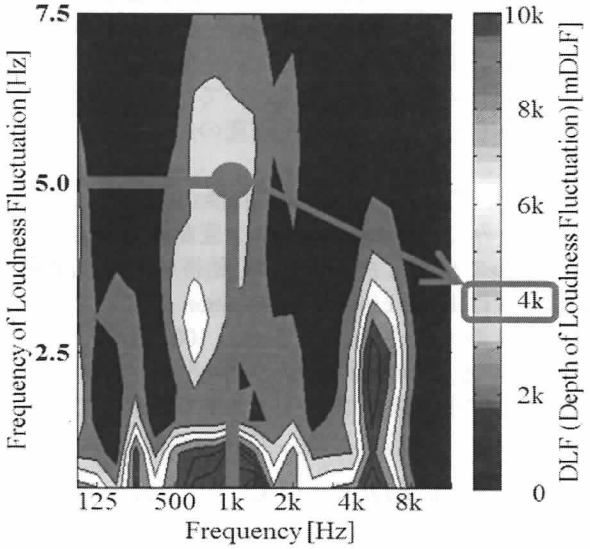

図 7 An example of the *DLF* analysis [12].

4.2. *LDR* vs. 変動帯域限定 *DLF* 和

次に，この *DLF* マップから明瞭性に関係する指標を導出する目的で，図 8 に示す方法で，変動量を中心周波数方向（図 7 では横軸方向，図 8 の *DLF* マップでは Freq.[Hz]方向）に加算し，変動周波数に対する変動量和 DLF_S（添字 S は和 summation の意）を求めた．

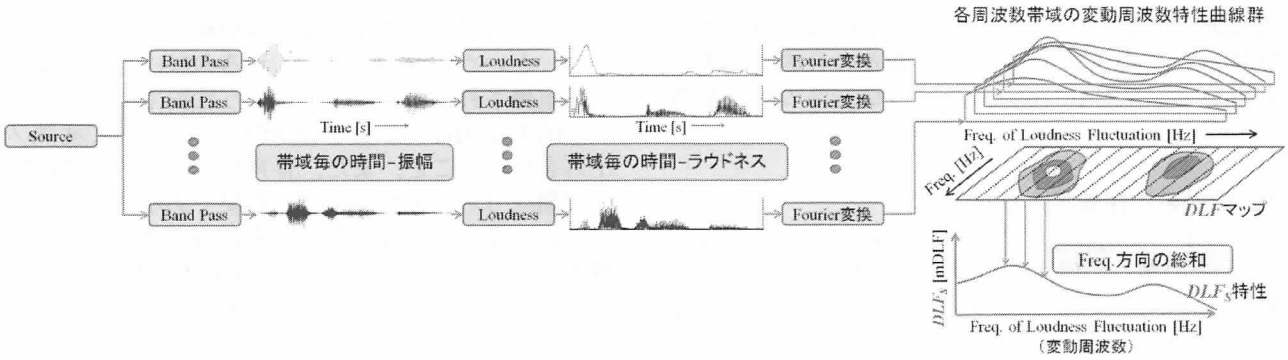

図 8 変動量 *DLF* のマップと変動量和 DLF_S の導出

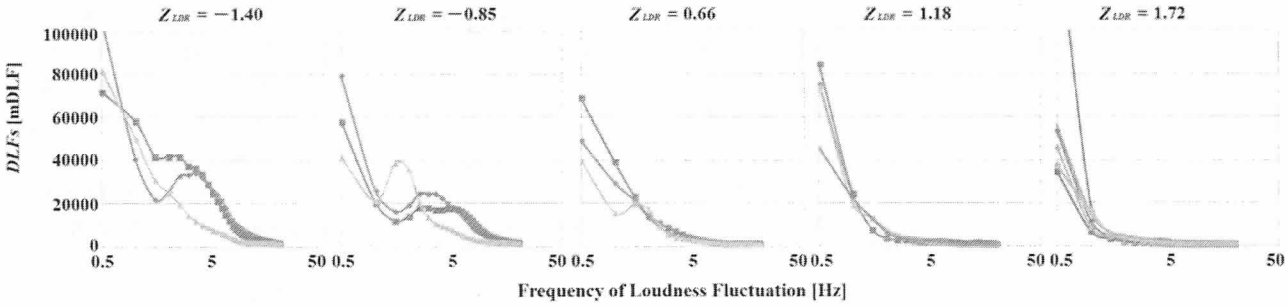

図 9 同じ Z_{LDR} を示す刺激の変動量和 DLF_S

2.4 で用いた 100 刺激から同じ Z_{LDR} を示す刺激をいくつか取り出し変動量和 DLF_S を見ると（図 9），Z_{LDR} の値が小さく聴き取りにくくない音声では，約 2～約 8 Hz の変動周波数において，特徴的な盛り上がりがみられ，これは既往研究の結果[40]とも一致している．

そこで，この盛り上がりの有無が明瞭性に対応しているとみて，変動量和 DLF_S を変動周波数 2.5～7.5 Hz に渡って積分した DLF_{BLS}（変動帯域限定 *DLF* 和，添字 *BLS* は Band Limited Summation の意）でこの盛り上がりの大きさを代表させ，100 刺激について DLF_{BLS} と Z_{LDR} の関係を見たのが図 10 である（尚，DLF_{BLS} は常用対数を 10 倍した dB で表示している）．

このとき，DLF_{BLS} と Z_{LDR} の間の回帰式は

$$Z_{LDR} = -0.28\,DLF_{BLS} + 13.5 \qquad (7)$$

で，決定係数 $R^2 = 0.93$ を得た．

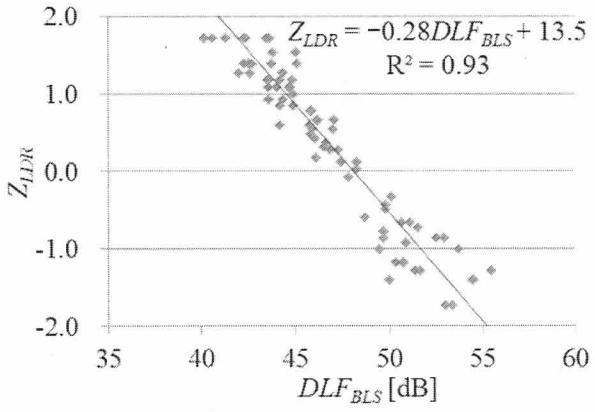

図 10 Z_{LDR} vs. DLF_{BLS} [12]

4.3. 変動量解析についての考察

3.1 でも述べたとおり，従来の明瞭性指標は主に伝送系の特性を評価していたのに対し，*DLF* は音声信号そのものを評価している点で興味深い．

決定係数だけで判断すると *SOR* を独立変数とした式(6)とほぼ同等である．しかし，*SOR* は全帯域のエネルギーを一緒くたにしているが，DLF_{BLS} は帯域毎に重み付けして加算することが可能など，まだチューニングの余地を残しており，さらなる展開が期待される．また，今の DLF_{BLS} のままでも，アナウンス毎に DLF_{BLS} と Z_{LDR} の回帰式を求めたところ，決定係数は *SOR* を独立変数とした場合に比べてほぼ全て同等かそれ以上であった（表 2）．

表 2 アナウンス別の決定係数

アナウンス	独立変数		
	SOR		DLF_{BLS}
sp01	0.94	<	0.98
sp02	0.92	<	0.96
sp03	0.98	>	0.97
sp04	0.98	<	0.99
sp05	0.95	<	0.96
sp06	0.89	=	0.89
sp07	0.92	<	0.98
sp08	0.93	<	0.96
sp09	0.95	<	0.96
sp10	0.90	<	0.98

今回，*DLF* を解析した市販ソフトウェア[39]は簡便に変動特性の傾向を見るには大変有用であったが，

*DLF*がまだ公にオーソライズされた単位ではないため学術的な議論に向いているとは限らない．詳細が公開されていて，且つ，今回紹介したDLF_{BLS}と同等な*AMI*（Average Modulation Indices）[3][41]などが利用できないか検討してみたい．

5. 終わりに

拡声システムのイコライザによる調整の効果が*STI*で評価できないという問題に端を発し，*SOR*から推定した*LDR*で評価する試み，そして拡声音の変動特性に着目した指標で評価する試みについて概説した．その過程で，各種明瞭性指標が何を評価しているかについて考察し，伝送系の評価ではなく音声そのものを評価する視点も必要ではないかとの問題提起をした．今回紹介した*SOR*やDLF_{BLS}がその答えであるとは言い切れないが，少なくともDLF_{BLS}は受音点信号のみを評価対象としていることから，かなり聴取者の視点に近い指標になっているのではないだろうか．

この他に，受音点の信号のみを用いて明瞭性を判断するという意味では，音声認識システムの応用が興味深い．Takano & Kondo[42]，Kondo[43]は騒音環境下において，人による単語了解度と，騒音に適切に適応した音声認識システムの認識率とが非常によく対応していることを示した．しかしArai[44]による残響環境下での実験では，まだ十分な対応が取れていないことから，音声認識システムを残響に適用させるなど，課題が残っていると思われる．

これまで，受聴者が音声を聴いて聴き取りにくい，不明瞭だ，何と言っているのか分からないなどと苦情を申し立てても，それらはあくまで主観的な報告であり，受聴者はそれを客観的に主張するツールを持ち合わせてなかった．伝送系を評価する指標が，建築屋，音響屋らが担当する範囲（室，拡声システム）の性能に問題が無いことを主張するツールであるように，音声そのものを評価する指標は，適切な品質の音声サービスを受けていないことを聴取者の視点から客観的に主張できるツールになり得る．

騒音計が騒音を発する側を監視することで，世の中の騒音を低減し，より良い音環境づくりに貢献した．同様に音声そのものを測定する「明瞭計」（又は「不明瞭計」？）なるツールがあれば，受聴者が適切な明瞭性の音声サービスを受ける権利が侵されてないか監視でき，世の中の，特に拡声音の明瞭性向上に役立つことであろう．

参考文献

[1] Arai T, Murakami Y, Hayashi N, Hodoshima N, Kurisu K, "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.

[2] 栗栖清浩，中村進，安啓一，荒井隆行，"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，" 建築音響研究会資料 AA2011-47, 2011.

[3] 栗栖清浩，中村進，安啓一，荒井隆行，"拡声音のオーバーラップマスキング量 SOR と「聴き取り難さの関係」：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 年春季, pp. 1225-1226, 2012.

[4] Kurisu K, Nakamura S, Yasu K, Arai T, "Signal to overlap-masking ratio of the broadcasted speech and its listening difficulty: An application for an evaluation tool of sound system tuning," *Acoust. Sci. & Tech.*, vol. 34, no. 5, 2013 (in printing).

[5] 荒井隆行，木下慶介，程島奈緒，楠本亜希子，喜田村朋子，"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.

[6] Arai T, Kinoshita K, Hodoshima N, Kusumoto A, Kitamura T, "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," *Acoust. Sci. & Tech.*, vol. 23, no. 4, pp. 229-232, 2002.

[7] Hodoshima N, Arai T, Kusumoto A, Kinoshita K, "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.

[8] IEC60268-16, Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index, Edition 4.0, 2011.

[9] 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，" in 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, 2003.

[10] 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.

[11] 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.

[12] 栗栖清浩，中村進，安啓一，荒井隆行，音声の変動量と「聴き取りにくさ」について：残響環境下での様々なアナウンスの評価，" 音講論集, 2013 年春季, pp. 1229-1230, 2013.

[13] Petzold E, *Elementare Raumakusik*, Bauwelt-Verlag Berlin, 1927.

[14] Haas H, "Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache," *Acustica*, vol. 1, pp. 49–58, 1951.

[15] Reichardt W, Abdel Alim O, Schmidt W, "Abhängigkeit der grenzen zwischen brauchbarer und unbrauchbarer durchsichtigkeit von der art des musikmotives, der nachhallzeit und der nachhalleinsatzzeit," *Applied Acoustics*, vol. 7, no. 4, pp. 243–264, 1974.

[16] ISO3382-1, Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces, 2009.

[17] Thiele R, "Richtungsverteilung und Zeitfolge Der Schallrückwürfe in Räumen," *Acustica*, vol. 3, pp. 291-302, 1953.

[3] 音声信号のうち変調周波数2～8 Hzの成分の大きさを平均したもので，複数人が同時に発話した音声から人数を推定するときに用いられた．

[18] Schroeder M R, "Modulation transfer functions: Definition and measurement," *Acustica*, vol. 49, pp. 179-182, 1981.

[19] 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.

[20] Mapp P, "Is STI a robust measure of sound system speech intelligibility performance?," *First Pan-American/Iberian Meeting on acoustics*, 2002.

[21] Sato Ha., Morimoto M, Sato Hi., "The relation between listening difficulty ratings and various objective measures in rooms," *Forum Acousticum*, pp. 1713-1718, 2005.

[22] Sato Hi, Bradley J S, Morimoto M, "Using listening difficulty ratings of conditions for speech communication in rooms," *J. Acoust. Soc. Am.*, vol. 117, no. 3, pt. 1, pp. 1157-1167, 2005.

[23] Kondo K, *Subjective quality measurement of speech: Its evaluation, estimation and applications,* Springer, 2012.

[24] 例えば, 南風原朝和, 心理統計学の基礎：統合的理解のために, 有斐閣, 2002.

[25] Williams J D, "An approximation to the probability integral," *The Annals of Mathematical Statistics*, vol. 17, pp. 363-365, 1946.

[26] 山内二郎(編), 日本規格協会, 統計数値表 JSA-1972.

[27] 栗栖清浩, 中村進, 安啓一, 荒井隆行, "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," 音講論集, 2012 秋季, pp. 1251-1252, 2012.

[28] 匂坂芳典, 浦谷則好, "ATR 音声・言語データベース,"日音学誌, vol. 48, no. 12, pp. 878-882, 1992.

[29] The Nitech Japanese Speech Database, http://hts.sp.nitech.ac.jp/

[30] Kuttruff H, *Room Acoustics, 4th. ed.*, Spon Press, London, 2000 (著: クットルフ, 訳: 藤原恭司, 日高孝之, 室内音響学, 市ヶ谷出版, 2003).

[31] Houtgast T, Steeneken H J M, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acustica*, vol. 28, pp. 66-73, 1973.

[32] ITU-T Recommendation P.862: Perceptual evaluation of Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

[33] Beerends J G, van Buuren R, van Vugt J, Verhave J, "Objective speech intelligibility measurement on the basis of natural speech in combination with perceptual modeling," *J. Audio Eng. Soc.*, vol. 57, no. 5, pp. 299-308, 2009.

[34] 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, pp. 3679-3686, 2004.

[35] Drullman R, van Wijngaarden S J, "New directions for a speech-based speech transmission index," *J. Acoust. Soc. Am.* (Abstracts), vol. 119, p. 3442, 2006.

[36] 例えば, Davidson J, Peters J F, Bhatia M, Kalidindi S, Mukherjee S, *Voice over IP Fundamentals (2nd ed.)*, Cisco Press, 2006.

[37] 森本政之, 室内音響心理評価のための物理指標について, 音響技術, no. 90, pp. 35-37, 1995.

[38] Arai T, Pavel M, Hermansky H, Avendano C, Intelligibility of speech with filtered time trajectories of spectral envelopes, *Proc. of the International Conf. on Spoken Language Processing (ICSLP)*, vol. 4, pp. 2490-2493, 1996.

[39] 小野測器, OS-0272 Oscope 変動音解析ユーザーズガイド.

[40] Greenberg S, "On the origins of speech intelligibility in the real world," *Proc. of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels*, pp. 23–32, 1997.

[41] Arai T, "Estimating number of speakers by the modulation characteristics of speech," *ICASSP*, pp. II-197 to II-200, 2003.

[42] 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.

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

[44] 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.