室内音響指標に基づく残響下音声認識性能の推定と評価

福森 隆寛†，森勢 将雅††，西浦 敬信†††
立命館大学　情報理工学部‡

1．はじめに

近年、音声認識技術の発展に伴いハンズフリー音声インタフェースの実現に高い注目が集まっている．しかし入出力間距離が長いと残響混入により音声認識性能が低下する問題がある．また近年，残響下音声認識手法に関する研究が盛んに行われているのに比例して，残響下音声認識性能の推定に関する研究にも注目が集まっている．現在，残響下音声認識性能を推定する手法として残響時間に基づく性能推定法が提案されている．しかし残響時間は同一室内で固有の値となるため音声認識性能を推定する残響尺度としては不適切である．また先行研究[1]において室内音響指標が残響尺度として有効であることが確認されている．そこで本稿では室内音響指標と回帰分析から更に高精度な残響下音声認識性能の推定を試みる．

2．音声認識性能推定のための従来の残響尺度

従来の残響尺度として残響時間(T60)[2]が用いられてきた．これは室内に放射された音の残響エネルギー密度が 60 dB 減衰するまでの継続時間長を表す．そして M. R. Schroeder により提案された 2 乗積分法[3]による残響測定法が提案され系の残響曲線はインパルス応答 $h(t)$を用いて式(1)に基づき容易に算出できるようになった．

$$\left\langle y_d^2(t)\right\rangle = N\int_t^\infty h^2(\lambda)d\lambda, \qquad (1)$$

ここで<>は集合平均，Nは単位周波数あたりのパワーを示す．$\left\langle y_d^2(t)\right\rangle$は残響曲線を示し，この曲線に基づき 60 dB 減衰するまでの時間が残響時間となる．音声認識性能は入出力間距離や位置関係に依存するため，室内固有の値をとる残響時間は残響下音声認識性能を推定する残響尺度としては不十分であることが問題視されている．

3．頑健な音声認識性能推定のための残響尺度

3．1　室内音響指標（ISO3382）

本研究は、これまでに入出力間のインパルス応答から音声認識性能と初期反射音量の関係について評価を行った[4]．その結果，約 25 ms から後続の反射音は音声認識性能の低下要因であることを確認した．しかし反射音量が大きくても音声認識性能が劣化しない系も確認できた．そこで従来の残響尺度を補う手法として室内音響指標(ISO3382)に着目した．これは音の初期部分の減衰状態を表現するために提案された指標である．この室内音響指標内の初期反射音と後続残響音のバランスの指標群に着目し，その中から式(2)に基づき算出される C 値を用いて音声認識性能との関連性を検証した．

$$C_{t_1} = 10\log_{10}\left(\int_0^{t_1} h^2(t)dt \Big/ \int_{t_1}^\infty h^2(t)dt\right), \qquad (2)$$

t_1は初期反射音と後続残響音の境界時間を示す．C 値は直接音と初期反射音のエネルギーに対する後続残響音のエネルギー比を表す．また音声認識性能と C 値の間に強い相関がある(特にt_1が約 30 ms)ことが先行研究[5]から確認されている．

3．2　室内音響指標を用いた音声認識性能推定

本稿では予め音声認識性能を推定するための学習データを用いて音声認識性能と C 値に対して環境ごとに回帰分析(回帰曲線：1 次，2 次，指数関数)を行う．回帰曲線は 1 次，2 次関数については最小 2 乗法より，指数関数については底を線形近似することで算出している．そして算出した回帰曲線を残響時間ごとに分類後，推定用インパルス応答から算出した残響時間と C 値を基に算出した回帰曲線から音声認識性能の推定を行う．学習セットに存在しない残響時間については線形補間によって曲線を近似して音声認識性能推定を行った．ただし指数関数は線形補間不能のため，今回は最近傍残響時間の回帰曲線から音声認識性能推定を行った．

4．室内音響指標を用いた残響尺度の策定

本研究では表 1(A)の 3 環境で残響尺度の策定を行った．C 値算出時のt_1は 30 ms に設定した．会議室の結果を図 1 に，各回帰曲線の相関係数を表 2 に示す．結果から各環境に対する指数関数の相関係数が 0.96 を超えており，高精度に分布を近似できた．

Estimation and Evaluation of Speech Recognition Based on Acoustic Parameters under Reverberation Environments
† Takahiro Fukumori, †† Masanori Morise, ††† Takanobu Nishiura
‡ College of Information Science and Engineering, Ritsumeikan University

表1：残響尺度策定と認識性能推定実験条件

(A) 残響尺度策定環境	和室(T60 = 400 ms, 72 RIRs) 会議室(T60 = 600 ms, 120 RIRs) 階段(T60 = 850 ms, 56 RIRs)
(B) 認識性能推定環境	研究室(T60 = 450 ms, 72 RIRs) 浴室(T60 = 650 ms, 28 RIRs) EV ホール(T60 = 850 ms, 120 RIRs)
音声・話者数	ATR 音素バランス 216 単語・14 話者
エンジン	Julius(クリーンモデル)

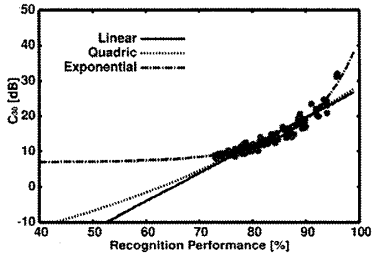

図1：C30と音声認識性能の関係(会議室)

5．残響下音声認識性能推定

5．1　独自収録のインパルス応答による推定

表1(B)の3環境で音声認識性能推定実験を行った．各環境について環境オープン・クローズテストを行った．推定精度評価には回帰曲線から算出した音声認識性能の推定値とテストデータの真値との差分を示す平均推定誤差を用いた．音声認識性能の推定結果を表3に示す．

残響時間が 450 ms 以外の環境は全ての回帰曲線において従来法よりも推定精度が向上した．残響時間が 450 ms の環境もオープンテストの推定精度が低下したが差異は小さい．またオープンテストの場合，指数関数で推定した際の平均推定誤差が全て 3 %以内であり標準偏差も他関数と比較して最小であった．特に高残響環境ほど高精度に音声認識性能を推定可能であった．そしてクローズテストの場合は 2 次関数で分析した際の平均推定誤差が低かった．ただし指数関数の場合も最大平均推定誤差が 2.22 %で他関数との大きな差異はなかった．したがって指数関数を残響尺度として用いることで頑健な音声認識性能の推定が可能であることを確認した．

5．2　CENSREC-4 を用いた性能推定

今回、雑音・残響環境下音声認識タスクの共通評価フレームワークである CENSREC から残響環境下音声認識評価用の CENSREC-4 を用いて策定した残響尺度(指数関数)の頑健性を表4の実験条件に基づき評価した．残響尺度策定環境は表1(A)と同様である．その結果，提案尺度に基づく推定誤差値は和室が 2.9 %(従来法：16.2 %)，リビングが 4.6 %(従来法：9.3 %)となり，共に従来法よりも高精度に音声認識性能を推定できた．

表2：相関係数

計測環境	T60 [ms]	相関係数 1次	2次	指数
和室	400	0.885	0.890	0.962
会議室	600	0.933	0.941	0.962
階段	850	0.957	0.968	0.966

表3：音声認識推定誤差
(**CT**：Close Test, **OT**：Open Test)

T60 [ms]	T60のみ CT	T60のみ OT	1次 CT	1次 OT	2次 CT	2次 OT	指数 CT	指数 OT
450	2.60	2.50	1.05	3.33	1.03	3.09	1.32	2.94
	3.10	**3.26**	**1.33**	**4.67**	**1.29**	**4.14**	**1.70**	**3.38**
650	5.36	6.13	1.95	4.09	1.97	3.87	1.99	2.71
	6.92	**7.18**	**2.36**	**4.56**	**2.37**	**4.35**	**1.70**	**3.24**
850	7.34	15.32	2.43	5.90	2.11	4.20	2.22	2.45
	8.80	**17.64**	**3.49**	**8.25**	**2.84**	**5.18**	**3.46**	**2.95**

上段は平均推定誤差[%]を，下段は標準偏差を示す．

表4：CENSREC-4 を用いた認識性能実験条件

認識性能推定環境	(1) 和室(T60 = 450 ms) C30 = 5.6 dB, 認識性能 = 54.3 % (2)リビング(T60 = 650 ms) C30 = 6.4 dB, 認識性能 = 65.3 %
音声・総発話数	連続数字音声(1～7桁)・4,004発話
話者数	104話者(女性：52名，男性：52名)
エンジン	Julius(クリーンモデル)

6．おわりに

残響環境下における室内音響指標と音声認識性能の関係について回帰分析し音声認識性能推定を試みた．その結果，各環境の C 値と音声認識性能の関係を指数関数で近似することで高精度に音声認識性能を推定できることを確認した．今後は周波数指標も含めて詳細な分析や評価を行うと共に音声認識性能推定に適した新残響尺度の確立を目指す．

謝辞　本研究の一部はグローバル COE，科研費による研究助成を受けた．また SLP 雑音 WG の諸氏に感謝する．

参考文献

[1] 平野良季，傳田遊亀，中山雅人，西浦敬信，"室内音響指標を用いた残響下音声認識性能の分析と推定"，音講論(秋季)，pp. 205-206，2007.
[2] 日本音響学会，"新版音響用語辞典"，コロナ社，2003.
[3] M．R．Schroeder，"New Method of Measuring Reverberation Time"，JASA，vol. 37，pp. 409-412，1965.
[4] 西浦敬信，傳田遊亀，"音声認識における初期反射音の影響についての検討"，音講論(春季)，pp. 205-206，2007.
[5] 平野良季，傳田遊亀，中山雅人，西浦敬信，"室内音響指標を用いた残響下音声認識性能の評価"，音講論(春季)，pp. 183-184，2008.