

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 6 日現在

機関番号：14501

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24760467

研究課題名(和文) 音声特徴量の保存度に着目した建築空間における音声品質の実時間評価法の開発

研究課題名(英文) Development of real-time methods of evaluating speech quality in rooms focusing on conservation degree of physical characteristics of speech

研究代表者

佐藤 逸人 (SATO, Hayato)

神戸大学・工学(系)研究科(研究院)・准教授

研究者番号：30346233

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：受聴点における音声の物理的な特徴量の保存度に注目することにより、実時間で室内空間における音声品質を評価する方法を開発した。具体的には、音声了解度の物理指標の測定所要時間の短縮方法、残響音付加音声のラウドネス推定方法、低SN比の条件における受聴点信号からの音声了解度推定方法を検討した。また、その応用の一例として、実時間評価を発話者に視覚的にフィードバックすることによって、過度な発話レベルの増大を抑えることができることを実証した。

研究成果の概要(英文)：By focusing on conservation degree of physical characteristics of speech at the listening point, real-time methods of evaluating speech quality in rooms, specifically, a method of reducing the time required for measurement of physical indicators of speech intelligibility, a loudness estimation method of reverberated speech, and a speech intelligibility estimation method from the listening point signal in conditions of low SN ratio were investigated. Furthermore, as an example of its application, it demonstrated that visually feedback of a real-time evaluation can suppress the increase of excessive vocal level.

研究分野：環境音響学

キーワード：音声品質 実時間評価 フィードバック

1. 研究開始当初の背景

音声によるコミュニケーションは、日常生活において必要不可欠であり、住宅における会話や、講演会、公共空間における案内放送等様々な形で行われる。室内における音声品質は、拡声設備や暗騒音、残響音といった伝送路の品質だけでなく、音声そのものの特徴、例えば発話速度や文節間のポーズ等の影響も受ける。しかしながら、これまでの研究では、伝送路の評価に重点が置かれ、評価に用いる音声を固定した上で、伝送路の物理的評価指標と音声品質の対応が検討されており、すべての要因を含む受聴点における音声から品質評価を行なう方法は確立されていない。さらに、現在のところ、音声品質の評価は、各種伝送系の設置後に、必要な品質をみたしているかどうかの確認に一度だけ行われることが多く、発話が行われているその時々における実時間評価はあまり注目されていない。

しかし、実時間評価については、潜在的な需要があると考えられる。例えば、発話者によるその評価をフィードバックすることにより、それを手がかりとして、受聴者にとって聞き取りやすい発話方法に変えることが考えられる。また、近年個人情報保護の観点から、会話の隣室への漏洩が問題視されるようになってきたが、隣室においてどの程度音声が聞こえてしまっているかを発話者に実時間でフィードバックすることができれば、発話者が声をひそめて話す等の対策が取ることが可能になる。以上のように、音声品質の実時間評価の応用範囲は広い。

2. 研究の目的

受聴点における音声の物理的な特徴量の保存度に注目することにより、実時間で室内空間における音声品質を評価する方法を開発する。また、その応用の一例として、特にその評価を発話者にフィードバックすることによって、発話の品質が向上するかどうかを明らかにする。

3. 研究の方法

以下の3つのテーマについて検討した。

(1) 音場の評価指標算出の所要時間：受聴点において観測される音から、音源-受聴点間の音声伝送性能の評価指標を算出する際に、測定誤差が許容範囲におさまる所要時間を検討した。

(2) 受聴点における観測信号に基づく音声品質評価指標の検討：実時間評価のためには、受聴点における観測信号から直接音声品質評価が可能な指標を算出する必要がある。本研究では、①残響音付加音声のラウドネス評価、②低SN比条件における音声了解度評価を検討した。

(3) 実時間評価のフィードバックが発話に与える影響：受聴点における音声品質評価を、発話者に実時間で視覚的にフィードバック

した場合に、発話がどのように変わるかを分析した。特に受聴点における音声をそのまま聴覚的にフィードバックした場合との比較に着目した。

4. 研究成果

(1) 音場の評価指標算出の所要時間

受聴点における音声信号だけでなく、伝送系への入力信号も得られる条件を想定する。大学内の教室1室を対象とした実測実験を行ない、音声明瞭度の評価指標として C_{50} を、クロススペクトル法で求めたインパルス応答から算出し、精度の高い C_{50} 算出の所要時間がどの程度であるかを明らかにする。さらに、比較的精度良く推定できている初期部分のみを切り出して C_{50} を算出することにより、所要時間の短縮を試みる。

測定対象とした教室の寸法は、幅 7.7m、奥行き 14.3m、高さ 2.8m である。講演者の位置に相当する場所に室の中央に向けて、マウスシミュレータを内蔵したダミーヘッド (B&K, Type 4128C) を設置した。ダミーヘッドの口元から 0.05m 離して入力信号を録音するマイクロホン (M1) を設置した。着座した受講者を想定し、教室の後方で床から 1.2m の高さに設置したマイクロホン (M2) で、受聴点における音声信号を録音した。

まず、 C_{50} の真値を求めるために、マウスシミュレータと M2 の間のインパルス応答を Swept-sine 信号を用いて測定した。その結果から、式(1)を用いて C_{50} を求めた。 $p(t)$ はインパルス応答の音圧波形を示す。

$$C_{50} = 10 \log_{10} \frac{\int_0^{50\text{ms}} p^2(t) dt}{\int_{50\text{ms}}^{\infty} p^2(t) dt} \quad (1)$$

その結果、 C_{50} は 4.2dB であった。以下ではこの値を真値として分析する。

次に、クロススペクトル法を用いて C_{50} を測定した。マウスシミュレータから無響室録音の音声 (音素バランス 1000 文) を再生し、M1 と M2 で録音を同期して行なった。M1 と M2 で録音した信号を、音場のインパルス応答長さに相当する 2^{15} サンプルの時間窓長を用いて 50% オーバーラップで順次に切り出し、それらの平均クロススペクトルからインパルス応答を求めた。この方法では新たにサンプルを切り出す度にインパルス応答が更新されるため、 C_{50} も更新される。なお、クロススペクトル法による測定の際は、教室の隅に向けて設置したスピーカから空調騒音を想定したピンクノイズを再生した。ノイズの再生レベルは、受聴点における SN 比が +15dB と ±0dB になるよう設定した。

図 1 に、クロススペクトル法による C_{50} と真値とした C_{50} の差が弁別閾とされる 1.1dB [1] 以下になるまでの所要時間を示す。SN 比が高い方が所要時間は短い、SN 比が ±0dB と低い場合は、20s 程度の時間を要するため、このままでは実時間評価に応用することは難しい。

C_{50} の測定誤差は、反射音の音圧が低くなるインパルス応答後部におけるノイズに起因すると考えられた。そこで、インパルス応答の 2 乗波形の初期部分 (~200ms) から求めた時間包絡線と、インパルス応答の時間軸上で 3/4 以降のエネルギー平均が一致する時間 t_1 を求め、この t_1 以降は信頼できない波形であると、式(2)により C_{50} を求めることとした。

$$C_{50} = 10 \log_{10} \frac{\int_0^{50\text{ms}} p^2(t) dt}{\int_{50\text{ms}}^{t_1} p^2(t) dt} \quad (2)$$

このアルゴリズム適用後の所要時間を図 1 に合わせて併せて示す。SN 比が $\pm 0\text{dB}$ と低い場合でも所要時間は 7s 程度と短縮された。所要時間は音場のインパルス応答の長さに依存すること考えられるが、7s はこの教室のインパルス応答長の 10 程度に相当する。言い換えれば、SN 比が $\pm 0\text{dB}$ 以上の条件では、インパルス応答長の 10 倍程度の時間があれば、 C_{50} をその弁別閾の範囲に収まる精度で推定可能である。

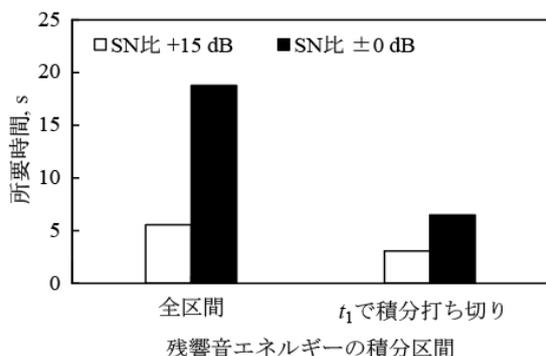


図 1 音声測定用信号とした場合のクロススペクトル法による C_{50} の測定所要時間

クロススペクトル法によるインパルス応答の測定は一般的技術であるが、音声明瞭度の物理指標を算出するために必要な時間が検討された例は見当たらず、本研究で得られた成果は、実用的なデータとして重要である。

(2) 受聴点における観測信号に基づく音声品質評価指標の検討

受聴点における音声信号のみが得られる条件を想定した検討を行なった。

① 残響音付加音声のラウドネス評価

音声には聴きやすいラウドネスが存在するため、その品質評価にあたっては、ラウドネスを定量化する方法が必要である。既存のラウドネスの物理指標である等価騒音レベルや Zwicker のラウドネスは、残響音付加音声のラウドネスと対応しない[2]が、音場のインパルス応答から求めた直接音と残響音のエネルギー比と等価騒音レベルを用いれば、残響音付加音声のラウドネスを評価できる[3]ことがわかっている。しかし、受聴点における音声信号しか得られない場合、クロススペ

クトル法は使用できないため、音場のインパルス応答を逐次求めることはできない。そこで、受聴点で得られる残響音付加音声の時間変動特性に着目し、ラウドネスとの関係を分析した。

既存の研究により、ラウドネスが既知である 22 種類の残響音付加音声について、本研究で新たに提案した時間変動強度 (F') を算出した。 F' が大きいほど時間変動が強いことを示す。

手順 1: 残響音付加音声に 1 オクターブバンドごとのバンドパスフィルタをかける。

手順 2: 音圧二乗波形を求める。

手順 3: 高速フーリエ変換し、その振幅スペクトルを求める。

手順 4: 振幅スペクトルを音圧二乗波形の積分で基準化する。

手順 5: 基準化した振幅スペクトルについて、変調周波数 0.5~Hz から 12~Hz の範囲の算術平均を求める。

7 つの中心周波数のオクターブバンドについて F' を求め、ラウドネスと比較した結果、1kHz のバンドが最も相関が高かった。1kHz の F' とラウドネスの対応を図 2 に示す。縦軸の値は調整レベルと呼ぶラウドネスと対応する値であり、この値が大きいほどラウドネスは小さくなる。図 2 より高い負の相関が見られ、ある特定の話者に限定された結果ではあるが、 F' により残響音付加音声のラウドネスを推定可能であること示す。

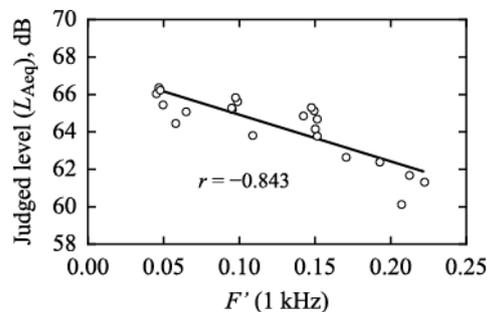


図 2 調整レベルと F' の対応関係

残響音の付加により、既存のラウドネスの物理指標と知覚されるラウドネスの対応が悪くなるという問題は、あまり認知されていない。その対応を補正する方法は、研究代表者らによる提案のみしか現状では見当たらず、価値のある知見であると言える。

② 低 SN 比条件における音声了解度評価

医療機関や金融機関などの個人情報を含む会話が行われる空間において、会話内容が漏洩する問題、すなわちスピーチプライバシーの問題が生じている。スピーチプライバシーの評価法として、音声了解度が古くから重要視されてきた。我が国における関連研究においても、音声了解度に着目したものが多く、例えば Sato ら[4]により、遮音性能、騒音の音圧レベル、音声レベルから高い精度で単語了解度を予測できる等単語了解度線が提案さ

れた。しかし、特に音声レベルについては現場において大きく変動することが想定され、スピーチプライバシーを確実に保護するためには、実時間による評価に基づいてマスキングノイズシステムを動的に調整する、あるいは話者が声をひそめる等の対応が必要である。そこで、受聴点における音の物理統計量を用いて音声了解度を予測し、スピーチプライバシーを評価できるかについて基礎的な検討を行った。

SN 比の低い騒音付加音声を受取る際には、時間一周波数平面上で瞬時的に SN 比の高い範囲の情報から、音声全体を類推していると考えられる。このような瞬時的に SN 比の高い範囲を音圧レベルの時間軸上の傾きの確率分布に基づいて検出し、この範囲が分析対象とする時間一周波数平面上の範囲全体に占める割合（音声検出率、SDR）を求めて、音声了解度の予測指標とする方法を考案した。

本研究では、事前に求めた暗騒音のみの場合の傾きの 95 パーセントタイル値を閾値とし、これを超える傾きが出現した場合に音声を検出されたと見なすものとした。21 の 1/3 オクターブバンド（中心周波数 100~Hz から 10~kHz）ごとに、騒音付加音声の音圧レベルの時変動の傾きと傾きの閾値を比較し、SDR を求めた。6 種類の SN 比の条件について、SDR を求めた結果と、文献[4]に示された SN 比と音声了解度の近似曲線の比較を図 3 に示す。音声了解度の予測曲線は、SDR とおおよそ同じ SN 比から上昇し始め、-30dB から -5~dB の範囲ではよく似た分布形状をとる。したがって、SDR は音声了解度とよく対応する可能性が高いと言える。

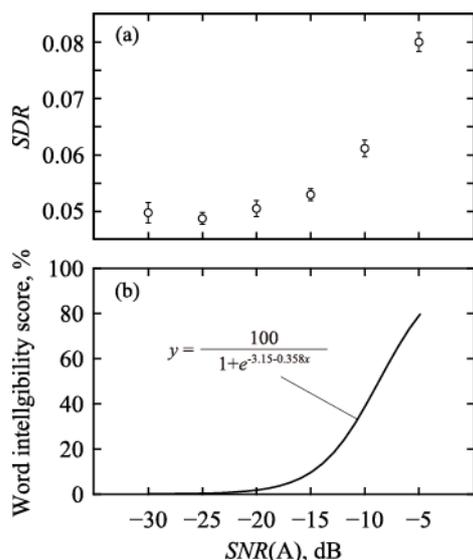


図 3 (a) SDR と SN 比の対応関係、(b) 単語了解度と SN 比の近似曲線[4]

SN 比が低い条件では、受聴点では音圧レベルを測定しても暗騒音の音圧レベルのみしか測定できず、一般に音声了解度の予測は

難しい。本研究で提案した方法は、受聴点における測定のみから音声了解度を評価するものであり、ここでは可能性を示したに留まったが、今後の発展的研究による実用化は需要が高いと考えられる。

(3) 実時間評価のフィードバックが発話に与える影響

音声品質評価を実時間でフィードバックした場合の発話レベルの変化を検討した。実験は無響室で行なった。被験者の正面にマイクと液晶モニタを設置した。マイクの位置は、被験者からの水平距離が 50cm、高さが 110cm である。音圧レベルはすべて等価騒音レベルで設定した。

聴覚フィードバックに関するシステムは以下の通りである。(a) マイクに入力した音声を、オーディオインターフェイス経由で PC に取り込む。(b) PC 上で音圧レベルの調整と騒音の付加を行う。(c) (b)の信号をイヤホンから提示する。聴覚フィードバックは音声の増幅率をパラメータとした。マイク位置における音圧レベルが 66dB の場合に、外耳道入口における音圧レベルが、76dB と 80dB となる 2 種類の増幅率を設定した。騒音は 70dB で一定とした。

視覚フィードバックに関するシステムは以下の通りである。(a) PC において取り込んだ音声を加工して出力する際に、同時に SN 比を算出する。(b) SN 比をメーターとして描画し、液晶モニタに表示する。メーターの色は SN 比によって変わり、SN 比が+10dB 未満では黒、+10dB 以上+22dB 未満では黄緑、+22dB 以上では赤で表示される。被験者には黄緑色を保つように発話させた。

表 1 に示す 6 条件をランダムに並び替えた試行を 1 試行とし、25 試行の実験を行なった。それぞれの条件については 30s ずつ発話させ、マイク位置における音圧レベルを記録した。被験者は 6 名である。

表 1 フィードバック条件

	音声の増幅率	聴覚フィードバック	視覚フィードバック
a	小	有り	無し
b	大	有り	無し
c	小	無し	有り
d	大	無し	有り
e	小	有り	有り
f	大	有り	有り

図 4 に結果を示す。多くの被験者において、増幅率によらず視覚フィードバックのみを提示した条件の発話レベルが他のフィードバックの条件よりも低い。したがって、視覚フィードバックより聴覚フィードバックのほうが発話レベルを大きくする効果があると言える。一方、増幅率が大きい条件では、聴覚フィードバックのみの条件と比較して、聴覚および視覚フィードバックの両者を提示した条件の発話レベルが低下した。これは、

視覚フィードバックでは適切な SN 比の情報を色で示したが、聴覚フィードバックではそのような情報が得られないため、過度に大きく発声してしまう傾向があると言える。ただし、被験者 H, J のように視覚フィードバックのみの場合の発話レベルが大きく低下する場合がある。特に被験者 J については、内観報告において視覚フィードバックを参考として発話することが難しいと述べている。つまり、被験者のフィードバックに対する感じ方が異なり、被験者にとってフィードバックに対する得手不得手がある点には留意する必要がある。

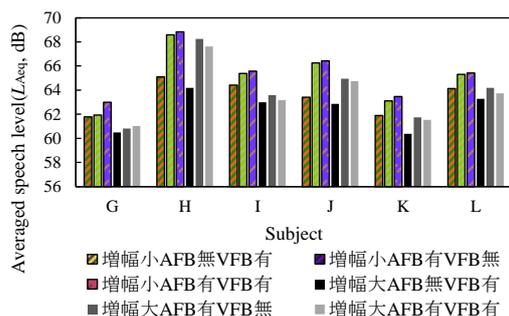


図4 条件ごとの発話レベルの平均

実時間で得られる音声品質評価を視覚的にフィードバックした場合に、発話レベルに明らかな違いが見られた。特に視覚フィードバックには過度な発話レベルを抑える効果が期待できる。今回は騒音付加条件のみであったが、今後残響音付加条件についても検討を進める予定である。

5. 主な発表論文等

[学会発表] (計6件)

- ① 吉本 雄大、佐藤 逸人、残響音を付加した音声の物理特性とラウドネスの関係、日本音響学会秋季研究発表会、2015.9.18、会津大学(福島県)
- ② 北村裕一、佐藤逸人、スピーチプライバシーの現場測定法に関する基礎的検討ー受聴点における物理統計量による評価ー、2015.9.16、会津大学(福島県)
- ③ 吉本 雄大、佐藤 逸人、残響音を付加した音声の自己相関関数がラウドネスに与える影響、日本建築学会近畿支部研究発表会、2015.6.27、大阪工業技術専門学校(大阪府)
- ④ 佐藤 逸人、森本 政之、吉本 雄大、残響音を付加した音源の物理特性とラウドネスの関係、日本音響学会建築音響研究会、2015.3.10、鎌ヶ谷市きらりホール(千葉県)
- ⑤ 松原 奏美、佐藤 逸人、室内空間における音声品質のリアルタイム評価に関する基礎的研究ークロススペクトル法による物理指標測定の所要時間ー、日本建

築学会近畿支部建築環境工学若手研究者研究発表・交流会、2014.11.26、近畿大学(大阪府)

- ⑥ 松原 奏美、佐藤 逸人、室内空間における音声品質のリアルタイム評価に関する基礎的研究ークロススペクトル法による物理指標測定の所要時間ー、日本建築学会近畿支部研究発表会、2014.6.22、大阪工業技術専門学校(大阪府)

6. 研究組織

(1)研究代表者

佐藤逸人 (SATO, Hayato)

神戸大学・大学院工学研究科・准教授

研究者番号：30346233