

令和 2 年 6 月 1 日現在

機関番号：14501

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K06671

研究課題名（和文）耳入力信号に着目した都市・建築空間における音情報の伝送品質評価に関する研究

研究課題名（英文）A study on evaluation of transmission quality of sound information in urban and architectural spaces focusing on ear input signals

研究代表者

佐藤 逸人 (SATO, Hayato)

神戸大学・工学研究科・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）：本研究の目的は、音情報の伝送品質を一般的な信号音と妨害音の強度比ではなく、人間が聴く音、つまり左右の耳に入力される音のみから評価する方法を研究することである。信号音の特徴として調波構造に着目し、聴覚モデルを用いて調波構造の基本周波数が連続的に抽出できた区間に基づいて音声伝送品質を評価する方法を検討した。その結果、音声と雑音と同じ方向から到来する場合は、本研究で提案する方法で音声了解度を予測できる可能性が示された。一方、音声と雑音が異なる方向から到来する場合における音声了解度の予測については、改善の余地が残された。

研究成果の学術的意義や社会的意義

本研究の学術的な特色は、聴取者の位置で得られる左右の耳への入力信号のみから音情報の伝送品質を評価する点である。一般に音情報の伝送品質の評価に用いられる信号音と妨害音の強度比は、稼働状態の空間では測定できないことが多く、長時間の測定を必要とする面的評価や様々な応用が考えられる実時間評価は難しい。また、強度比だけでは信号音の違いが伝送品質に及ぼす影響を完全には評価できない。これらの問題を解決することにより、音情報の伝送品質がさらに向上すれば、さらに安全・安心な都市・建築空間の構築が期待できる。

研究成果の概要（英文）：The purpose of this research is to study a method to evaluate the transmission quality of sound information only from the sounds that humans hear, i.e., the sounds input to the left and right ears, rather than from the ratio of intensity of signal and disturbing sounds that is widely used. The length the segment where the fundamental frequency of the harmonic structure can be continuously extracted using an auditory model proposed in this research was tested as an index to evaluate the speech transmission quality. The results showed that the method proposed in this research can predict the speech intelligibility when speech sound and disturbing noise come from the same direction. On the other hand, there is still room for improvement in the prediction of speech intelligibility in the case that the two sounds come from different directions.

研究分野：環境音響学

キーワード：音声伝送品質 耳入力信号 聴覚モデル ピッチ 両耳聴

## 1. 研究開始当初の背景

申請者は、都市・建築空間における音情報の伝送品質の評価方法と、その評価に基づく音情報の最適化に関する研究を長年進めてきた。一般に、音情報の伝送品質の評価指標として、情報を持つ信号音（音声、音楽）と、都市・建築空間における伝送品質を劣化させる妨害音（騒音、残響音）の強度比が用いられる。しかし、この強度比は稼働状態の空間では測定できないことが多く、長時間の測定を必要とする面的評価や様々な応用が考えられる実時間評価は難しい。また、強度比だけでは信号音の違いが伝送品質に及ぼす影響を完全には評価できない。

本研究では、このような「強度比に基づく評価」の限界をブレイクスルーする方法として、人間の音に対するパターン認識に関する研究を応用することを考える。都市・建築空間における信号音は妨害音と一体となって聴取者の耳に入力される。人間は音の要素の連続性や類同性などを手がかりとして、耳入力信号に含まれる信号音のパターンと妨害音のパターンを聞き分けること（音脈分凝）ができる。「パターン認識のしやすさ」を尺度化することにより、音脈分凝を音情報の伝送品質評価に応用できれば、評価に必要なのは耳入力信号のみとなる。つまり、信号音と妨害音を別々に測定する必要がなくなるため、測定の自由度が大幅に上昇する。また、信号音が異なれば耳入力信号も異なる。したがって、耳入力信号に基づく評価には、信号音の違いも反映できることが期待される。

## 2. 研究の目的

耳入力信号の音響特性からパターン認識のしやすさを尺度化し、音情報の伝送品質評価に応用するための基礎的研究を行なう。本研究では、音情報の音源として用いられる音声、鳥の鳴き声、楽音に共通する音響特性である調波構造に着目する。研究着手時は以下の3つを目的としたが、特に(1)の検討に時間を要したため、(2)までの検討にとどまった。(3)については今後の課題とする。

- (1) 研究全体の根幹技術となる、調波構造の「連続性」に基づくパターン認識のしやすさを尺度化し伝送品質の評価指標として応用する方法を明らかにする。
- (2) 信号音と妨害音の到来方向の違いが音楽内の方向定位精度に及ぼす影響を、左右の耳入力信号の調波構造の相互相関分析を用いて評価する方法を明らかにする。
- (3) 妨害音も調波構造を持つ場合に研究対象を拡大し、調波構造の「類同性」も合わせて考慮することにより、その影響を反映させる方法を明らかにする。

## 3. 研究の方法

まず、人間の聴覚モデルに基づいて、左右の耳入力信号から音の調波構造を抽出し、その連続性および類同性を評価するモデルを構築する。モデルの構築にあたっては、Lickliderによるピッチ抽出モデル[1]と、Lindemannの音像定位に関する両耳聴モデル[2]を参考とする。構築したモデルに音声と妨害音を混合した信号を入力し、その出力として得られた連続性あるいは類同性に対する評価を、音声と妨害音を混合した信号を聴取した際の音声了解度（音声为正しく聞き取れた割合と比較する。以上の方法により、構築したモデルによる音声了解度の予測の可能性を検討する。

## 4. 研究成果

### (1) 調波構造の「連続性」に基づくパターン認識のしやすさの尺度化

一般に音の調波構造は時間-周波数解析により高速に抽出できるが、本研究では人間の聴覚モデルに基づいて求める。提案モデルでは以下の順で計算を行う。

- ① 内耳における周波数分解：Bark尺度で2から22の帯域のIIRバンドパスフィルタバンクで模擬する。
- ② 有毛細胞の模擬：①の出力を半波整流処理した後にカットオフ周波数を800HzとするIIRローパスフィルタを通す。
- ③ 片耳聴モデル：Lindemannのモデルでは左右耳の入力信号を乗算するが、ここでは簡易に片耳聴モデルとして扱うため②の出力を自乗する。
- ④ 周期性の抽出：特定の周期に反応する神経細胞のモデルを考える。反応する周期の2倍の長さの固定長バッファを設定し、最新の③の出力を最後尾に追加する。次に、バッファの前半と後半の相関係数を計算し、半波整流処理を施す。
- ⑤ 調波構造の強度の算出：21帯域ごとの④の結果を平均する。このことにより、倍音成分まで含めて調波構造の強度を定量化する。

- ⑥ 調波構造の強度ピークの算出：④および⑤の処理を複数の周期について行い、強度がピークとなる周期を抽出する。この周期の逆数は音声の基本周波数の推定値に相当する。ピークは複数抽出し、その強度が最大ピークの75%以上であるものを基本周波数の候補として記録する。
- ⑦ 連続性の評価：5msの固定時間長の基本周波数のバッファを設定し、このバッファから線形回帰により次のサンプルの基本周波数を推定する。⑥の候補のうち、バッファから推定した基本周波数との差が5%未満のものがあればバッファの最後尾に追加し、なければ0を追加する。バッファがすべて0の場合は、⑥の候補のうち周波数が最大のをバッファに追加する。最終的な基本周波数の推定値は、バッファの50%以上が0ではない場合はバッファから推定した値とし、そうでない場合は0とする。20ms以上連続して0ではない基本周波数の推定値が得られた区間を有声区間とする。

以上の方法で得られた有声区間の長さは、調波構造の「連続性」の評価尺度と考えることができる。最終的には、妨害音が付加されていない音声（クリーン音声）について、有声区間の長さの統計量をあらかじめ求めておき、その統計量と妨害音を付加した音声から算出した有声区間の長さを比較することで、調波構造の「連続性」に基づくパターン認識のしやすさを尺度できると考えられる。本研究の範囲では、統計量ではなく、いくつかのクリーン音声のサンプルを対象とした。クリーン音声と雑音付加音声の有声区間を比較し、両者の基本周波数がクリーン音声のものと同じ（誤差5%未満）している区間が、クリーン音声の有声区間に占める割合（CI：Continuous Index）を評価指標として、音声了解度の予測可能性を検討した。

図1に単語「アマグモ」の音源（女声）を上述したモデルに入力し、有声区間を求めた結果を示す。図1(a)はクリーン音声、(b)はA特性重み付け音圧レベルで求めたSN比が0dBとなるようにクリーン音声にピンクノイズを付加した音源の結果である。縦軸は本研究で提案するモデルによる基本周波数の推定値であり、プロットがある区間が有声区間である。比較のために(b)には(a)の結果を灰色で併せて示す。(b)の有声区間の総時間長は(a)の65%程度まで短くなっており、ノイズを付加した影響が見られる。その一方で、(a)と(b)で基本周波数が一致していない区間も見られる。(b)の場合、CIは約0.60である。

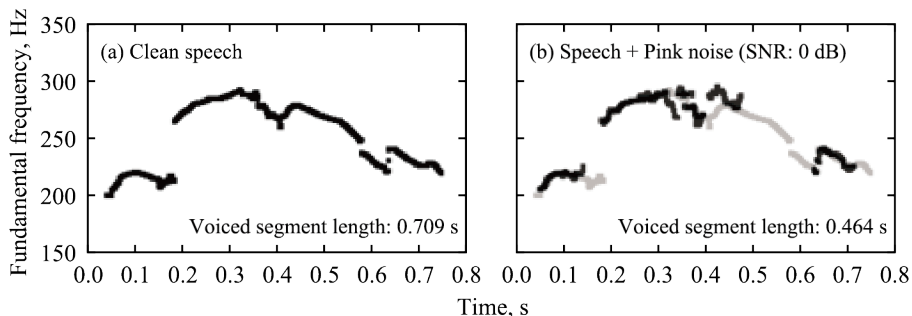


図1 提案モデルを用いて抽出した単語（アマグモ、女声）の有声区間と基本周波数

A特性重み付け音圧レベルで求めたSN比が-20dBから+20dBで5dBステップとなるようにクリーン音声にピンクノイズを付加した音源を作成し、それぞれについてCIを求めた結果を図2に示す。クリーン音声は前章で示した計算例で用いた音源と同じである。若干のゆらぎはあるが、CIとSN比の関係は-20dBから+20dBの範囲で変化するS字曲線を示した。実際の音声了解度の例として、Satoら[3]による予測式に基づく単語了解度とSN比の関係を図2に併せて示す（実線）。単語了解度が変化する範囲（-20dBから5dB）はCIのそれに内包されており、CIを用いて単語了解度を予測できる可能性が示された。

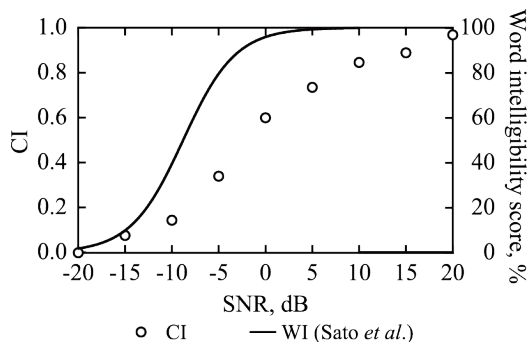


図2 CIおよび単語了解度（WI）とSN比（A特性重み付け）の関係

音声には、狭帯域ノイズでマスキングされてもマスキングされていない帯域が十分に広ければ了解度が大きく低下しないという傾向がある。この傾向をCIで反映できるかを確認するため

に、250Hz から 4kHz をそれぞれ中心周波数とする 1/1 オクターブバンドノイズをクリーン音声に付加し、CI を求めた。A 特性重み付け音圧レベルで求めた SN 比は 0dB と -20dB に設定した。結果を図 3 に示す。CI は SN 比やバンドノイズの中心周波数によらず 1 に近い値を示し、特定の帯域のみマスキングされても音声了解度と同様に CI も低下しないと言える。既往の音声了解度の物理指標では、各帯域に一定の重み付け係数を与えるため、特定の帯域をマスキングすると、その帯域の重要度に応じて必ず値が低下する。この点については、CI の方が既往の物理指標より実際の音声了解度とよく対応すると考えられる。ただし、1kHz を中心周波数とするバンドノイズを用いた場合のみ CI が 0.9 程度と若干低い。これは音節から音節の遷移区間において基本周波数の誤差が大きいことが原因であった。このような CI の低下が子音の異聴等に反映されるかどうかについては今後の課題とする。

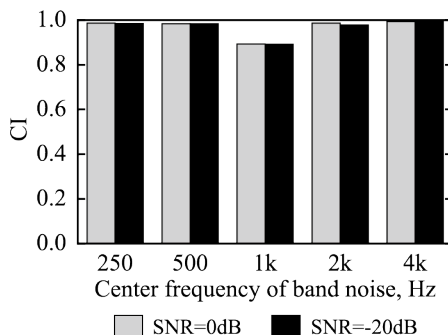


図 3 バンドノイズを付加した場合の CI

## (2) 信号音と妨害音の到来方向の違いが音案内の方向定位精度に及ぼす影響

(1)の研究で提案モデルを用いて音声了解度を予測できる可能性が示されたことから、このモデルを両耳聴モデルに拡張し、音声と雑音が異なる方向から到来することによる音声了解度の違いを CI で評価できるかについて検討した。

検討にあたり、(1)で用いた提案モデルを以下の通り変更した。

- ① 内耳における周波数分解：Bark 尺度で 1 から 23 の 23 帯域を対象とした。
- ② 有毛細胞の模擬：変更なし。
- ③ 両耳聴モデル：それぞれの周波数帯域において図 4 に示すように左右の信号を  $\Delta\tau$  ずつ遅らせながら乗算する。信号のサンプリング周波数は 48kHz とし、 $\Delta\tau$  は 1 サンプルとした。想定する両耳時間差の範囲  $\pm 1\text{ms}$  とし、合計で 49 チャンネルを分析対象とした。なお、Lindemann のモデルでは左右の信号を乗算した後に時間平均処理を行うが、ここでは次の周期性を抽出する処理で精細な時間波形が必要なため行わない。
- ④ 周期性の抽出：変更なし。
- ⑤ 調波構造の強度の算出：④の結果を平均する際の帯域の数を 23 に変更。
- ⑥ 調波構造の強度ピークの算出：基本周波数の候補として記録する条件を最大ピークの 50%以上に変更。
- ⑦ 連続性の評価：変更なし。

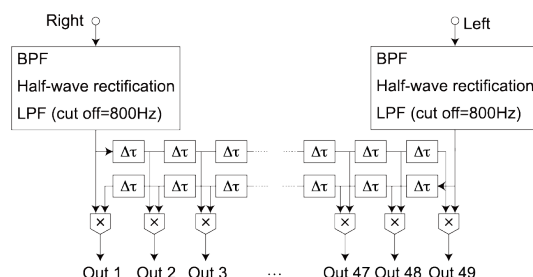


図 4 本研究で用いた両耳聴モデル（サンプリング周波数：48kHz）

東北大学電気通信研究所が公開している頭部伝達関数のデータベース [4] から KEMAR 擬似頭のデータを用い、左右それぞれの耳に入力される音声および雑音を模擬したバイノーラル音源を作成した。その音源を提案モデルに入力して CI を求めた。聴取者の前方に左右対称に配置した 2 つのスピーカを考え、左から音声、右からピンクノイズがそれぞれ到来する状況を模擬した。音声とピンクノイズの到来方向がなす角 ( $\Delta\phi$ ) は、 $30^\circ$  ステップで  $0^\circ$  から  $180^\circ$  の 7 種類とした。SN 比は 5dB ステップで +5dB から -15dB となるよう設定した。音声として 10 個の単語を用い、各

条件で10単語分のCIの平均値を求めた。図5に例として、 $\Delta\phi$ が $0^\circ$ と $120^\circ$ の結果を示す。同じSN比で比較した場合、 $120^\circ$ の方がCIが高く、提案モデルで音の空間特性による違いをある程度表現できていることがわかる。

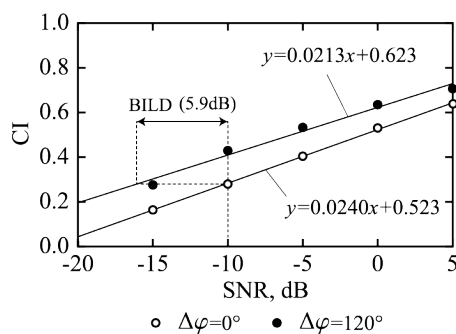
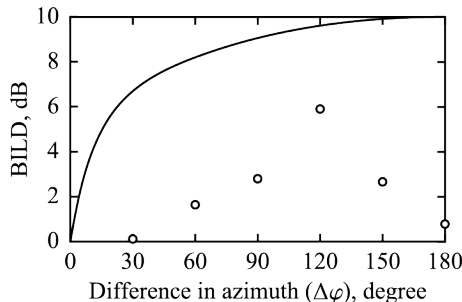


図5  $\Delta\phi$  と CI の関係

$\Delta\phi$ が $0^\circ$ 、つまり音声と雑音が同一方向から到来する場合の音声了解度を基準として、到来方向の変化による音声了解度の上昇量をSN比で換算した量を Binaural intelligibility level difference (BILD)とよぶ。このBILDを用いて提案モデルの妥当性を詳細に評価する。一般にBILDは音声了解度が50%程度になるSN比に基づいて算出する。ここでは $\Delta\phi$ が $0^\circ$ の場合、SN比が $-10$ dBで音声了解度が50%になると仮定し、この条件におけるCIと同じ値になるSN比をその他の $\Delta\phi$ において図5に示すような近似直線から求めることによりBILDを算出した。結果を図6に示す。図中の実線は、Blauert[5]が示したBILDの例であり、 $\Delta\phi$ が $30^\circ$ 程度までは急峻に上昇し、それ以降は緩やかに上昇して最大で $10$ dB程度に達する。一方、CIから求めたBILDは、 $\Delta\phi$ が $120^\circ$ までは上昇するが、上昇の傾きの傾向は音声了解度によるBILDとは異なり、さらに最大値も $5.9$ dBに留まった。また、 $\Delta\phi$ が $150^\circ$ あるいは $180^\circ$ の条件で最大値よりもBILDが低下するという音声了解度によるBILDとは異なる傾向が見られた。このような結果が得られた原因は、本研究で用いた聴覚モデルにおいて、両耳間レベル差の影響の考慮が不十分であること、人間の選択的注意による妨害音の抑制効果が考慮されていないことなどが考えられる。これらについては今後の課題としたい。



— BILD from *Spatial Hearing* (Blauert, 1996) ○ BILD estimated from CI

図6 BILD と  $\Delta\phi$  の関係

<引用文献>

[1] J. Licklider, A duplex theory of pitch perception, *Experientia*, 7, 128-134, 1951.  
 [2] W. Lindemann, Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals, *J. Acoust. Soc. Am.*, 80, 1608-1622, 1986.  
 [3] H. Sato *et al.*, Relationship between sound insulation performance of walls and word intelligibility scores, *Applied Acoustics*, 73, 43-49, 2012.  
 [4] <http://www.riec.tohoku.ac.jp/pub/hrtf/index.html>  
 [5] J. Blauert, *Spatial Hearing* (MIT Press, 1996), pp. 267.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 佐藤逸人
2. 発表標題 調波構造に着目した騒音下における音声了解度の予測
3. 学会等名 日本音響学会春季研究発表会
4. 発表年 2019年

1. 発表者名 佐藤逸人
2. 発表標題 調波構造に着目した音声了解度の予測：空間特性の考慮
3. 学会等名 日本音響学会秋季研究発表会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考