

スピーチプライバシー保護を目的とする 声の音韻性を打ち消すマスキング音の生成法

小助川 莉沙[†] 若槻 尚斗^{††} 水谷 孝一^{††} 海老原 格^{††}

[†] 筑波大学大学院システム情報工学研究科

^{††} 筑波大学システム情報系

1. はじめに

近年、プライバシー保護への関心が高まっている。病院の待合室や薬局など、会話によるコミュニケーションが欠かせない現場においては特にプライバシーへの配慮が求められている[1]。本稿では、第三者に対して会話内容を聞こえにくくするための音を出力するサウンドマスキング技術に着目し、声の音韻性を打ち消すようなマスキング音の生成法を検討する。

2. 音声の生成とケプストラム解析

人間が発声する際、声を出すためのパルス列状の音源を声帯で作出す。呼吸流による振動により声の高さが変化し、音源波が唇から放射されるまでに通過する声道の形によって母音の波形が決定される。音声波形に含まれるスペクトル包絡は、声道の通過特性により決定され、主に音声における音韻性を反映する。一方、スペクトル微細構造は、音源の特性により決定され、話者の個性を反映する。ケプストラムは、波形の短時間対数振幅スペクトル $X(\omega)$ の逆フーリエ変換として定義され、スペクトル包絡とスペクトル微細構造を近似的に分離して抽出できる。

3. マスキング音の生成

3.1 音情景解析と声の音韻知覚 複数の音が存在する環境の中での音聴取に関して、人間が音を聴く仕組みや音同士の関係について分析する音情景解析[2]にもとづき、二つの音を一つの音脈として知覚させるような音の融合によるマスキングが試みられている[3]。著者らは、この知見に基づいて追試を行い、声のスペクトル包絡を変形して生成したマスキング音が、声の音韻性の知覚を難しくする効果を確認した。

3.2 生成手順 予備実験として、周波数 44.1[kHz]でサンプリングした音声/a/, /i/, /u/, /e/, /o/それぞれのケプストラムにおける次数 2~180 の成分を 0 として音声に戻したとき、上記 5 種類の母音が弁別できなくなることを確認した。本研究では、元の音声とマスキング音が合成された時に、この音と同じスペクトルになることを意図して、マスキング音を生成する。すなわち、上述のようにケプストラムの低次成分を除いて求めた目標音のパワースペクトルから元の音のパワースペクトルを引いたものをマスキング音のパワースペクトルとして振幅に変換してから、元の音声の位相スペ

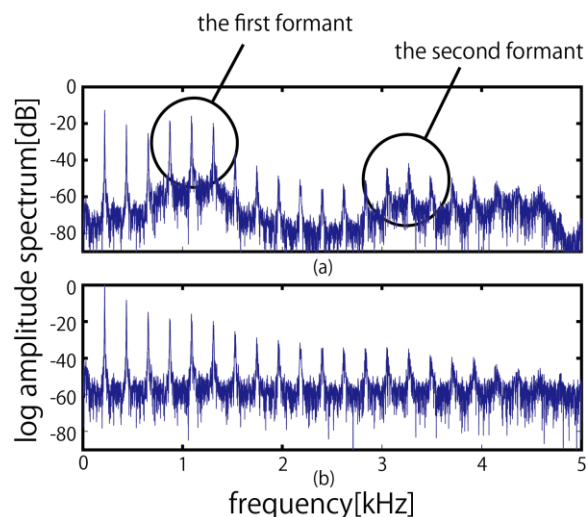


図 1 (a) 元の音「あ」の対数振幅スペクトル

(b) 元の音 + マスキング音の対数振幅スペクトル

クトルを乗じて逆フーリエ変換することにより、マスキング音を生成する。

4. 結果

一例として/a/の結果を図 1 に示す。元の音の対数振幅スペクトルを図 1(a)に、元の音と生成したマスキング音を足し合わせたものの対数振幅スペクトルを図 1(b)にそれぞれ示す。元の音とマスキング音を足し合わせた音のスペクトルは、元の音のそれに比べて平坦になっていることがわかる。すなわち、元の音のフォルマントは平滑化され、声の音韻性が打ち消されていることが分かる。

5. 今後の課題

今後は音韻知覚に関して更なる分析を行い、不快感の少ないマスキング音の生成を目指すとともに、子音についても分析することで、会話音に適応させることを目指す。

参考文献

- [1] 小山由美, 星和磨, 羽入敏樹, “調剤薬局におけるスピーチプライバシー保護—その 1 服薬指導時の会話漏洩の実態調査—”, 日本建築学会, 40133, pp.3, (2008)
- [2] A. S. Bregman, Auditory scene analysis: the perceptual organization of sound, MIT Press, Cambridge, MA, 1990.
- [3] 赤木正人, 入江佳洋, “音情景理解を応用した音声プライバシー保護”, 信学技報, EMM2011-59, Dec. 2011.