

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 6 日現在

機関番号：34310

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26350460

研究課題名(和文)音空間レンダリングによる防災行政無線の音声明瞭度予測・提示システムの開発

研究課題名(英文) Development of speech articulation estimation and presentation system for emergency broadcast system by sound field rendering

研究代表者

土屋 隆生 (Tsuchiya, Takao)

同志社大学・理工学部・教授

研究者番号：20217334

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究では、音空間レンダリングを用いて防災行政無線の音声明瞭度予測および提示が可能なシステムを開発している。レンダリング手法にCE-FDTD法の一つであるIWB法を採用し、任意形状に対応できるようにボクセラライザを開発した。サンプリング周波数を8 kHzに設定し、400×200×50 mの実空間のレンダリングを実施した。レンダリング結果を24chスピーカアレイシステムに出力したところ、現実的な音声明瞭度を提示できることが示された。

研究成果の概要(英文)：In this study, an estimation and presentation system of speech articulation for emergency broadcast system is developed by sound field rendering. The compact explicit-finite difference time domain (CE-FDTD) method is applied to the rendering. A voxelizer is developed for arbitrary boundary shape for the CE-FDTD method. Some demonstrations are made for a numerical model of 400×200×50 m with sampling frequency of 8 kHz by the sound field rendering. The rendering results are then reproduced by the 24ch loudspeaker array system. It is confirmed that the realistic speech articulation can be represented by this developed system.

研究分野：数値音響工学

キーワード：CE-FDTD法 防災無線 音空間レンダリング GPU ボクセラライザ

1. 研究開始当初の背景

東日本大震災は、津波により多くの犠牲者を出した。その多くは、避難の遅れが原因であると言われており、津波に限らず災害からの早期避難のための情報伝達法の検証が進められている。現在、情報伝達手段として中心的な役割を担っているのは防災行政無線(拡声子局)であるが、東日本大震災の場合、それによる音声聞き取りにくかったというケースが少なからず報告されている。その原因としては、地形や建築物による乱反射・回折などでロングパスエコーや残響が発生したことにより明瞭度が低下したケースと、複数の屋外スピーカからの音声が干渉することで不明瞭なものになったというケースが指摘されている。これらは、これまでの防災行政無線の設置設計が、スピーカ出力や音波の距離減衰といった振幅情報のみしか着目しておらず、位相などの時間軸情報が軽視されてきたことが原因である。したがって、聞き取りやすい防災行政無線網を構築するためには、時間軸情報を考慮した屋外音声明瞭度予測手法の開発が急務となっている。

2. 研究の目的

災害時の情報伝達手段として防災行政無線は有効であるが、背景でも述べたように地形や建築物による音波の乱反射・回折などに起因した音声明瞭度の低下が指摘されている。しかしながら、現時点で個々の拡声子局の設置状況に対応した音声明瞭度予測手法は確立されていない。本研究では、音空間レンダリングと呼ばれる音場解析手法を用いて、実際の地形や建築物に対応した音声明瞭度予測手法の開発を目的とする。さらに明瞭度を数値的に提供するだけでなく、レンダリング結果をスピーカアレイに出力することで音場そのものも3次元的に提供する音声明瞭度提示システムを構築し、防災システム設計者に対して音響的なフィードバックが可能なシステムを構築する。

3. 研究の方法

(1) 音空間レンダリング手法

音空間レンダリングに用いる数値計算手法として、CE-FDTD法を採用する。CE-FDTD法では、音圧 p に関する3次元波動方程式を差分法で直接離散化する。コロケートグリッド上で図1のような $2\Delta \times 2\Delta \times 2\Delta$ (Δ はグリッド間隔)の離散化領域(セル)を考え、セル内に存在する27の格子点を用いて波動方程式を離散化する。本解析では、空間は一様と考えるため x, y, z 方向のグリッド間隔を等しく Δ と考えている。CE-FDTD法も標準のFDTD法と同様に本質的には中心差分であるが、標準のFDTD法のように1つの座標軸に沿った離散化だけでなく、対角線方向や立体対角線方向も考慮して、次式のように離散化する。

$$p_{i,j,k}^{n+1} = 2p_{i,j,k}^n - p_{i,j,k}^{n-1} + \chi^2[(\delta_x^2 + \delta_y^2 + \delta_z^2) + a(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) + b\delta_x^2\delta_y^2\delta_z^2]p_{i,j,k}^n \quad (1)$$

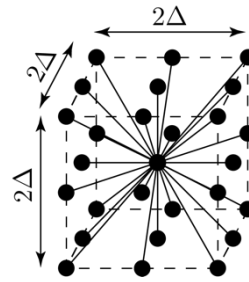


図1 CE-FDTD法のセル

ただし、位置 $(i\Delta, j\Delta, k\Delta)$ 、時刻 $t = n\Delta t$ (Δt は時間ステップ)における音圧を $p_{i,j,k}^n$ と表記している。また、 $\chi = c_0\Delta t/\Delta$ はCFL数、 δ^2 は中心差分に関する演算子で

$$\begin{aligned} \delta_x^2 p_{i,j,k}^n &\equiv p_{i+1,j,k}^n - 2p_{i,j,k}^n + p_{i-1,j,k}^n \\ \delta_y^2 p_{i,j,k}^n &\equiv p_{i,j+1,k}^n - 2p_{i,j,k}^n + p_{i,j-1,k}^n \\ \delta_z^2 p_{i,j,k}^n &\equiv p_{i,j,k+1}^n - 2p_{i,j,k}^n + p_{i,j,k-1}^n \end{aligned} \quad (2)$$

と定義される。式(1)において、 a, b は精度を制御するパラメータで、 $a = 0, b = 0$ とすると標準のFDTD法(SLF: Standard Leapfrog)となり、 $a = 1/4, b = 1/16$ とするとIWB(Interpolated Wide Band)法と呼ばれ、カットオフ周波数がサンプリング周波数の1/2という理想的な計算条件を実現できる。本研究ではIWB法を採用する。

(2) 任意形状への対応

音空間レンダリングを用いて実際の地形や建築物に対応した計算を実施するには、CE-FDTD法を任意境界形状に対応させる必要がある。そのために、本研究では音場モデルのボクセル化ソフト(ボクセライザー)を開発した。ボクセルは、画像分野における2次元画像を構成するピクセル(画素)に対応するもので、ボリュームとピクセルからつくられた造語である。CE-FDTD法では、3次元の物体形状を図1のようなボクセルの有無で空間形状を表現することになる。

3次元形状モデルの作成で用いられる3D CADの多くは、一般に表面(サーフェス)データのみを用いて形状を表現しているため、モデル内部が空洞になっている場合がほとんどである。それに対し、音空間レンダリングでは、モデル内部をボクセルで埋め尽くしたモデルが必要になるため、サーフェスモデルからボクセルモデルを作成する必要がある。最近では、3Dプリンタの普及に伴い、3次元ボクセル形状作成のためのボクセライザーやボクセルエディタも普及しつつあるが、多くは、それほど大きなボクセル形状を扱うことはできない。しかしながら、大規模な空間を扱う音空間レンダリングでは最低でも 10^9 ボクセル、大きなものでは 10^{11} ボクセル程度を扱う必要があるため、標準的なボクセルエディタ、ボクセライザーでは対応が難しい。そこで、本研究では3次元形状の表現で一般に使用されているサーフェスモデルからボクセルモデルを直接作成するためにボクセライザー-SUF2VOXを開発した。

ボクセライザー-SUF2VOXは、数値計算言

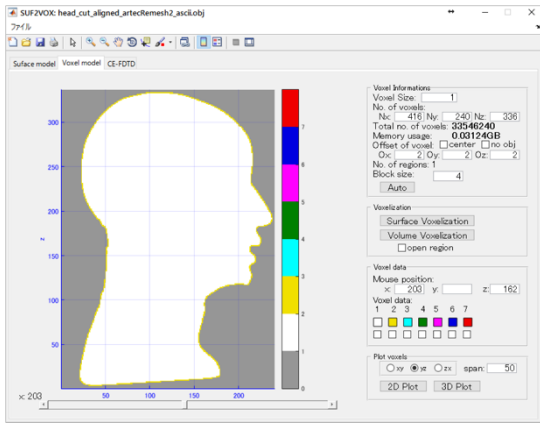


図 2 SUF2VOX の画面

語である MATLAB 上で開発した。入力する 3 次元サーフェスデータには、STL 形式と OBJ 形式に対応している。まず、読み込んだサーフェスデータ(3 角形)を離散化してボクセル空間に配置する。3 角形の 3 辺のボクセル化には Bresenham のアルゴリズムを使用した。その後、3 角形内部をボクセルで塗りつぶす処理を行う。これにより、サーフェスデータで表現された境界のみがまずボクセル化されることになる。つぎに、領域のボクセル化を行う。領域のボクセル化は、z 方向に間隔 Δ でスライスされた xy 断面ごとに境界内部を塗りつぶしていく。以上の処理をすべての断面について実施すれば、3 次元領域全体がボクセル化される。図 2 は、SUF2VOX の起動画面である。

(3)GPU クラスタの拡張

本研究では、既存の GPU クラスタシステムを拡張して計算に使用した。図 3 に構成を示す。まず、既存システムに 2 ノード追加して 10 ノードに拡張した。その上で、1 つのノードに 4 基搭載されている GPU を最新の NVIDIA 社製 GeForce TITAN X に換装した。これにより、全体で使用できるメモリ量を約 1.5 倍に引き上げた。したがって、システム全体では 40GPU が並列動作可能になった。ノード間の通信は InfiniBand を使用しているが、これも SDR から QDR に変更することで通信の高速化を図った。

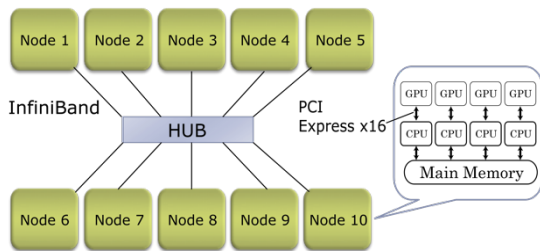


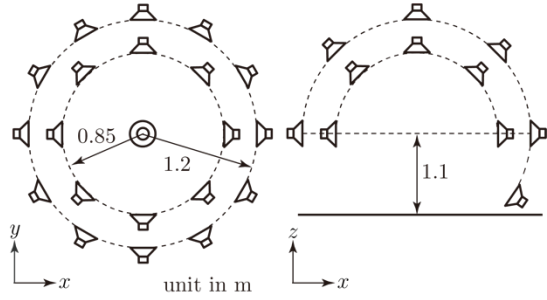
図 3 GPU クラスタ

(4)提示システムの構築

提示システムは、図 4 に示すような 24ch 半球ドーム型スピーカアレイで構成した。このシステムは、 $4.9 \times 4.0 \times 2.5 \text{m}^3$ の防音室内に 24



(a) 24ch スピーカアレイ



(b) スピーカ配置

図 4 24ch 提示システム

個のスピーカをドーム状に配置している。スピーカアレイに対応した半球状の仮想境界面をレンダリングモデル内に想定し、24 の仮想受音点で境界の外側から到来するインパルス応答を計算する。このインパルス応答に、防災無線の音声データを畳み込んで、スピーカアレイから出力することで、防災無線の放送状況を再現・提示する。

4. 研究成果

(1)レンダリング条件の検討

音空間レンダリングでは、メモリはサンプリング周波数の 3 乗、計算時間は 4 乗に比例するため、屋外の広範囲な領域を計算するためには、できるだけサンプリング周波数を低く抑える必要がある。一方で、帯域制限されたインパルス応答を用いると音声明瞭度指標の計算精度が低下することが予想されるため、サンプリング周波数やグリッドサイズと音声明瞭度指標との関係を検討する必要がある。

そこで、まず 1 方向に長い準 1 次元モデルを用いて、サンプリング周波数と明瞭度の関係を検討した。モデルは、 $500 \times 10 \times 10 \text{m}^3$ とし、一端に点音源を設置、390m の地点でインパルス応答を計算した。境界はすべて無反射条件とした。サンプリング周波数は、 $F_s = 6, 8, 10, 12 \text{ kHz}$ 、CFL 数は 0.98 として計算を行った。図 5 は、 $F_s = 8 \text{ kHz}$ のときの計算結果である。1.14 s 後から応答が始まっており、正しく計算が行えているのが分かる。ただし、数値分散の誤差の影響でリングングのような応答がずっと続いているため、可聴化すると耳障りな雑音が重畳されることになる。この誤差を軽減するには、ローパスフィルタ(L.P.F.)処理を施す必要がある。5 秒間の音声信号に計算されたインパルス応答を畳み込み、L.P.F. 処理後にスピーカで再生したところ、明瞭度を保ったままで再生可能な最低サン

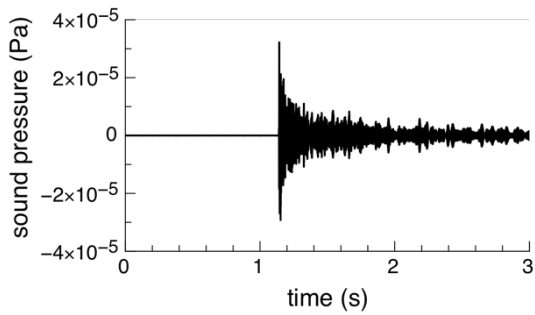
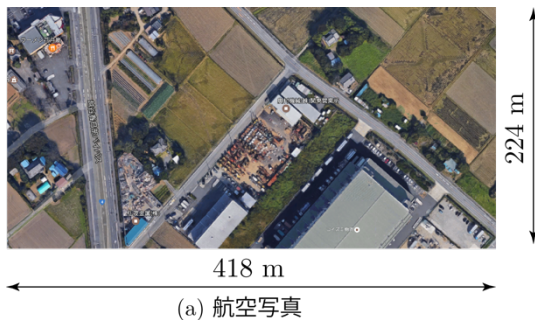


図5 インパルス応答の計算結果

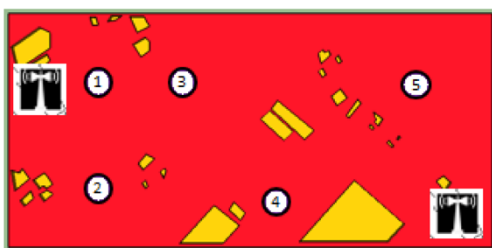
リング周波数は8 kHzであった。したがって、以後の計算ではサンプリング周波数を8 kHzとした。そのときのグリッド間隔 Δ は4.33 cmとなる。

(2) 実空間モデルによるレンダリング

つぎに、実際の屋外空間モデルを作成し、レンダリングを行った。モデル例として、防災無線マップが公開されている埼玉県春日部市の水角地区を採用した。図6は、水角地区の航空写真とレンダリングモデルである。図のように $418 \times 224 \times 50 \text{m}^2$ の空間についてモデル化を行った。図中のスピーカのマークは、防災無線の拡声子局の設置位置、丸数字は応答の計算地点である。スピーカは、高さ14 mの設置を想定している。計算条件は、地面の反射率を0.87、建物の反射率を0.99、他をすべて無反射に設定した。



(a) 航空写真



(b) レンダリングモデル

図6 屋外レンダリングモデル

図7は、①～⑤の各地点で計算されたインパルス応答である。すべての観測点でリングが発生しているが、これは先に述べた数値分散の誤差で、可聴化の際にはL.P.F.処理で除去可能である。観測点①～③では、波形を見ただけでも2つのスピーカの干渉が確認できる。また、観測点④では、2つのスピーカの干渉とともに、建物による反射や回折も

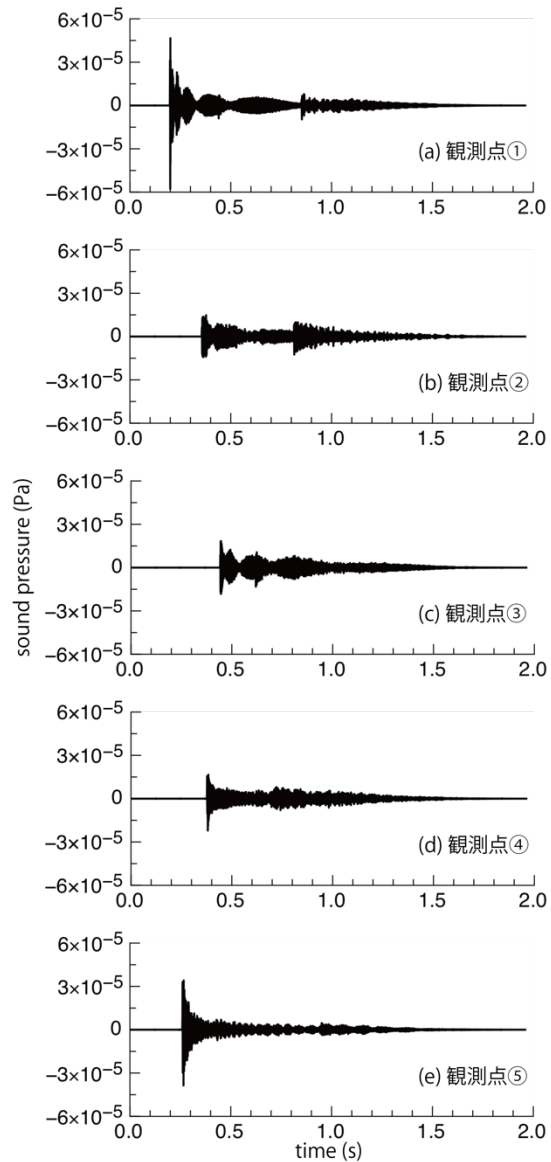


図7 各地点におけるインパルス応答の計算結果

加わり、複雑な応答になっている。

(3) スピーカアレイによる提示システムによる可聴化

最後に、上記で計算されたインパルス応答に5秒間の音声信号を畳み込み、24chのスピーカアレイシステムで提示した。その結果、波形でも示唆されたように、2つのスピーカと建物の影響による反射によって干渉が生じ、音声明瞭度が低下することが確認された。特に、観測点④では、反射や回折も加わりほとんど聞き取れないほどに明瞭度が低下した。

このように実空間モデルに対しても本システムの有効性を確認した。ただし、計算時間の都合上、聴取地点を5点しか設定できなかったため、地点間の移り変わりなどによる違いを心理実験するところまでは至らなかった。今後実用化する場合は、より多数の地点や繰り返し計算への対応などが必要となる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

- ① Yamashita, T. Tsuchiya, Y. Iwaya, M. Otani, and Y. Inoguchi, Reflective boundary condition with arbitrary boundary shape for compact-explicit finite-difference time-domain method, Jpn., J. Appl. Phys., 査読有, Vol. 54, 2015, pp. 07HC002.

[学会発表] (計 3 件)

- ① 土屋隆生, 音空間レンダリングを実装した VR システム ～自由視聴点レンダリングを目指して～, 電子情報通信学会技術研究報告, EA2016-76, 2017. 1. 25, 同志社大学(京都府・京都市).
- ② 土屋隆生, FDTD 法による任意境界音場計算のためのボクセラライザーの開発, 電子情報通信学会技術研究報告, EA2016-28, 2016. 8. 10, 東北学院大学(宮城県・多賀城市).
- ③ 杉浦恭輔, 土屋隆生, 大谷 真, 末永 司, 白木原太, リアルタイム畳み込みシステムを用いた音空間レンダラーの構築と評価, 電子情報通信学会技術研究報告, EA2016-17, 2016. 7. 8, ヤマハ豊岡工場(静岡県・磐田市).

6. 研究組織

(1) 研究代表者

土屋 隆生 (TSUCHIYA TAKAO)

同志社大学・理工学部・教授

研究者番号：20217334