

主電動機磁励音を用いた鉄道乗車位置推定手法の設計と評価

堀 佑貴^{1,a)} 石田 繁巳¹ 内野 雅人¹ 安藤 崇央¹ 荒川 豊¹ 福田 晃¹

概要：近年、鉄道電車内の混雑が社会的な問題となっており、電車内混雑に対する対策として各車両の混雑度に基づいた案内によって乗車人数の平均化を図るなどの方法が検討されている。この実現を目指した研究として、鉄道電車内の混雑具合を推定する研究が存在するが、車両情報の付与を手動に頼っており車両情報が付与されない場合は車両ごとの混雑具合を推定することが困難となる。そこで本研究では、自動で乗車車両を推定する手法を提案する。提案手法では駅プラットフォーム上に立つ鉄道利用者の持つ携帯端末を用いて電車走行音声を取得し、乗車車両の推定を行う。観測音声から主電動機が発する磁励音のみを分離した上で、磁励音の変化を機械学習することで外的雑音に対して頑健な乗車位置推定を実現する。実環境で収集した音声データを用いて評価を行なった結果、乗車車両位置を正解率 78%、平均誤差 0.28 車両で推定できたことを確認した。

1. はじめに

近年、鉄道電車内の混雑が社会的な問題となっている。国土交通省の調査では、東京都の主要 31 区間のうちピーク時の混雑度が 150% を超える路線が 24 路線、180% を超える路線が 11 路線存在することがわかっている [1]。電車内混雑時の通勤が心身に様々な影響を及ぼすことは文献 [2] で報告されており、鉄道における小さな遅延の約 6 割が電車内混雑などの利用者起因のものであることも報告されている [3]。これら電車内混雑に対する解決策としては、各車両の混雑度に基づいた案内によって乗車人数の平均化を図るなどの方法が検討されている。

この対策の実現を目指した研究として、混雑具合を推定する研究が存在する [4,5]。これらの手法を用いることで高い精度で混雑具合を推定できるが、「どの場所の混雑具合であるか」を考慮した推定は行われていない。鉄道車両を対象に混雑度推定を行った研究では乗車車両の推定に関する検討も含まれているが [6]、あらかじめ車両情報を持った端末や手動での車両情報の付与が必須である。

乗車車両情報の付与が自動的になされるようになれば車両ごとの混雑度推定を可能とする範囲が拡大することにつながることから、マイクロフォンを用いて乗車車両を自動推定する研究が存在する [7]。しかしながら、この手法ではロングレールを採用した路線には対応できず、観測音声をそのまま使用しているため外的な雑音にも弱いという問題がある。

本稿では、電車に搭載された三相交流誘導電動機（主電

動機）が発する音を用いて乗車車両を推定する手法を提案する。文献 [7] と同様にプラットフォーム上に立つ鉄道利用者の持つ携帯端末を用いて電車が到着するまでの電車走行音声を取得し、主電動機が発する磁励音のみを分離した上で機械学習により乗車車両位置を推定する。主電動機が発する磁励音を利用することで線路やプラットフォームの違い、線路の継ぎ目通過音や周囲の人が発する雑音などの外的雑音に対して頑健な乗車車両推定を実現する。

主電動機の磁励音を分離する試みはこれまでに報告されていないことから、本稿では既存の音源分離手法を適用できるかを検証する。磁励音の発生原理が音声の発生原理と類似することに着目し、音声の音源分離に用いられる半教師あり非負値行列因子分解 (SNMF) 及び深層学習 (DNN) による音声強調手法 (SEDNN) を用いて磁励音の分離を試みる。分離した音声から周波数領域の各種特徴量を計算し、機械学習によって推定した乗車車両位置を比較した。実際に鉄道駅プラットフォームで収集した音声データを用いて提案手法を評価した結果、乗車車両位置を平均誤差 0.28 車両で推定できたことを確認した。

本稿の構成は以下の通りである。2. では、既存の乗車車両推定に関する研究やサービスを紹介し、それらの問題点を示す。3. では、提案する乗車車両推定手法のシステム概要を示す。4. では、磁励音分離ブロックで提案する 2 つの磁励音分離手法について説明する。5. では、乗車車両推定ブロックで提案する分離後の磁励音音声から乗車車両を推定するための手法を示す。6. で各ブロックに対する評価結果を報告した上で、最後に 7. でまとめとする。

¹ 九州大学大学院システム情報科学研究院
ISEE, Kyushu University, Fukuoka 819-0395, Japan
^{a)} hori@f.ait.kyushu-u.ac.jp

2. 関連研究

筆者らが調査した範囲では、磁励音を利用した鉄道の乗車車両推定に関する研究は存在しない。その他、乗車車両の推定を目指した研究や乗車車両の推定に応用可能な研究としては、スマートフォン内蔵のセンサを利用したものや、本研究と同様にマイクロフォンを用いた研究が存在する。

2.1 スマートフォン内蔵のセンサを利用した鉄道乗車位置推定

鉄道駅における乗車車両情報を取得する手法として GPS を用いた方法が考えられる。しかし、GPS は人工衛星からの電波を利用して位置推定を行うため、屋内では測定精度が悪くなるという特徴があり、鉄道車両内や鉄道駅屋での利用では正確な位置推定ができない。追加の機材を必要とせず GPS の精度を向上させるための研究もいくつかなされているが、十分な精度は得られていない [8]。

また、「AirStamp」という技術を利用した手法も存在する [9]。これは電車内に超音波ビーコンを設置し、超音波をスマートフォンで受信することで乗車している車両の推定を行う。この手法は BLE ビーコンなどでも実現が可能であり、実際に JR 東日本の山手線でも実装されている。一方で、車両や鉄道駅内に設備を設置する必要があるこのような手法は鉄道会社のみ実現が可能であるため鉄道会社の対応に依存し、コストの観点などからも全路線に対して適用することは現実的でない。

Bluetooth シグナルの RSSI (Received Signal Strength Indicator) を利用した乗車位置推定手法も提案されている [6]。ここでは、一部の近隣端末の持つ乗車位置情報をもとにその近隣端末の発した Bluetooth シグナルを受信し、RSSI から自分が同一車両に存在する確率を算出している。しかし、乗車位置推定に複数の携帯端末を必要とするため、利用者数の少ない路線での実現は現実的でないという問題点が存在する。また、この手法では一部の端末が自己の乗車位置情報を保持している必要があり、手動での車両情報の付与が必須である。乗り込んだ車両すべてにあらかじめ車両情報が付与された端末が存在するという想定は現実的と言えず、車両情報を持った端末が 1 つも存在しない場合は、車両情報を得ることができない。

2.2 マイクロフォンを利用した鉄道乗車位置推定

本研究と同様に、手動での乗車車両情報の付与を必要とせず、マイクロフォンを利用して乗車車両の推定を行う研究として、電車走行時のジョイント音を利用する手法が報告されている [7, 10, 11]。ジョイント音とは電車が鉄道のレール継ぎ目上を通過する際に発生する音であり、文献 [7] では、電車走行音声からジョイント音の発生した回数を計算し、総発生回数をもとに乗車車両を推定する。具体的には、駅のプラットフォーム上に立った状態で電車走行音を

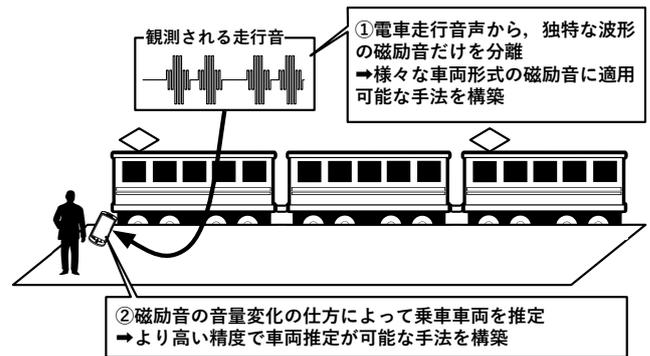


図 1 磁励音を用いた乗車車両推定システムの概要

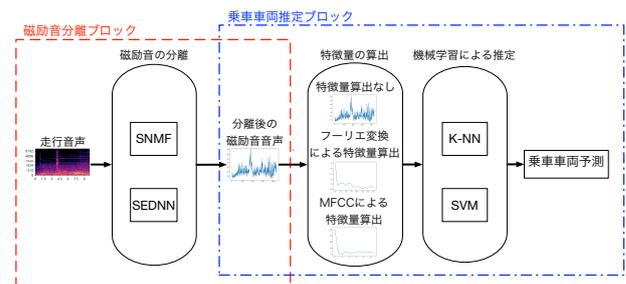


図 2 磁励音を用いた乗車車両推定システムの構成

取得し、その音声を短い時間窓に区切り機械学習によってジョイント音かそうでないかを判別する。取得した走行音声全体に対してジョイント音の合計回数を算出し、乗車車両位置の推定を行う。この手法では、車両や鉄道駅内に設備を設置する必要はなく、自己の乗車位置情報を保持する端末や他の端末を必要としない。しかし、近年ジョイント音の抑制や保守コストの低減を目的にロングレールを採用する路線が増えており、そのような路線ではこの手法は適用できない。また、全ての周波数を対象に特徴量を導出していることから、雑音がある場合の検出が困難であるという課題がある。

3. 磁励音を用いた乗車車両推定システム

図 1 に、磁励音を用いた乗車車両推定システムの概要を示す。提案システムは、主電動機から発生する磁励音を抽出し、磁励音の時間的な音量変化を機械学習することで乗車車両の推定を行う。

一般に、主電動機はある一定の車両間隔で搭載され、主電動機を搭載する車両のみが磁励音を発する。観測者の前方を電車が横切るときに観測される磁励音の音量は時間的に変化することから、この変化に基づいて主電動機搭載車両の通過を検出し、乗車車両を推定する。

実際には車両型式 (JR 東日本 E235 系など) によって主電動機を搭載する車両位置が異なることから、位置情報から推定される路線情報と磁励音を用いて車両型式を推定した上で、乗車車両を推定する。車両型式の推定は本稿の範囲外とし、本稿では車両型式が既知である前提で乗車車両を推定する。

図 2 に、提案システムのシステム構成を示す。提案システムは、磁励音分離ブロックと乗車車両推定ブロックの 2 つのブロックから構成される。磁励音分離ブロックでは、主電動機から発せられる磁励音の発生過程が音声の発生過程と類似していることに着目し、ジョイント音や周囲の人々が発する雑音が含まれた音に対して既存の音源分離手法を応用することで磁励音のみを分離する。乗車車両推定ブロックでは分離された磁励音から各種特徴量を算出して機械学習により乗車車両を推定する。

以降では各ブロックについて説明を行う。

4. 磁励音分離ブロック

磁励音分離ブロックでは音声と磁励音の類似性に着目し、既存の音源分離手法を用いて磁励音を分離する。具体的には、半教師あり非負値行列因子分解 (Semi-supervised Non-negative Matrix Factorization: SNMF) [12] 及び Deep Neural Network (DNN) を用いた音声強調手法 (Speech Enhancement Based on Deep Neural Networks: SEDNN) [13] の 2 種類の音源分離手法の適用を検討する。提案システムにおいてどちらの手法が有効であるかは評価において検証する。

4.1 磁励音

磁励音は、交流誘導電動機である主電動機を正弦波ではなくパルス列で駆動するために発生する。近年の電車は主電動機として交流誘導電動機を採用しており、可変電圧・可変周波数型 (Variable Voltage Variable Frequency: VVVF) 方式のインバータを用いて主電動機を駆動している。インバータはパルス幅変調 (Pulse Width Modulation: PWM) 方式 [14] を用いて擬似的に正弦波を生成するため駆動電流はパルス列となり、独特な磁励音が発生する。

磁励音の発生過程は音声の生成過程と類似している。実際に観測される磁励音は駆動電流のパルス列そのものではなく、高調波成分を多く含む擬似正弦波が複数の音響的な伝達特性を経て音色が付加された上で観測されることが報告されている [15]。この過程は、全体の振動によって作られた高調波成分を多く含む音に声道の形状によって定まる伝達特性で音色が付加される音声の発生過程と類似していると言える。

本研究では磁励音と音声の発生過程の類似性に着目し、音声に対して適用可能な既存の音源分離手法を適用する。音声ではスペクトル包絡が音色情報を表すと考えることから、本研究でも磁励音の特徴的な音色表現がスペクトルの包絡構造に現れると考えて磁励音を分離する。

4.2 SNMF を用いた磁励音の分離

図 3 に、SNMF を用いた磁励音分離の概要を示す。SNMF を用いた磁励音分離は学習フェーズと分離フェーズで構成される。

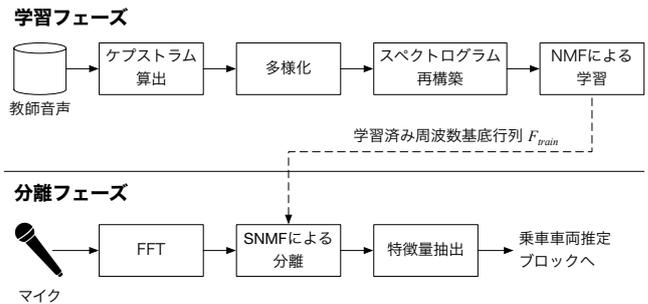


図 3 SNMF における処理概要

学習フェーズでは磁励音の教師音声を用いて非負値行列因子分解 (NMF) により磁励音の分離に必要な周波数基底行列 F_{train} を学習する。まず、教師音声を固定幅のウィンドウに区切った上で各ウィンドウでケプストラムを算出する。種々のインバータ・電動機の磁励音に対応するため、種々の磁励音のケプストラムの相関係数を計算し、相関係数の高いものだけ、すなわち種々の磁励音に共通する成分を抽出してスペクトログラムを再構築する。最後に、共通成分を含むスペクトログラムを用いて NMF による学習を行う。

NMF による学習では、観測信号の振幅スペクトログラム $Y = [y_1, \dots, y_N] \in \mathbb{R}^{M \times N}$ に対し教師基底行列 F_{train} を用いて、全ての行列の要素が非負であるという制約のもとで以下の行列分解を行うことで F_{train} を学習する。

$$Y = F_{train}G \quad (1)$$

ここで、 $G = [g_1, \dots, g_N] \in \mathbb{R}^{K \times N}$ は教師基底行列に対する結合係数行列である。

分離フェーズでは学習により獲得された F_{train} を用いて周辺環境ノイズを含む実際の音声から磁励音を分離した上で、乗車車両推定に必要な特徴量を抽出する。まず、マイクから得られた音声を固定幅のウィンドウに区切り、高速フーリエ変換 (Fast Fourier Transform: FFT) によって振幅スペクトログラム Y を算出する。次に、学習フェーズで獲得した教師基底行列 F_{train} を用いて以下の行列分解を全ての行列の要素が非負であるという制約のもとで行い、分離された磁励音の結合係数行列 G を得る。

$$Y \simeq HU + F_{train}G \quad (2)$$

ここで、 $H = [h_1, \dots, h_K] \in \mathbb{R}^{M \times K}$ 、 $U = [u_1, \dots, u_N] \in \mathbb{R}^{K \times N}$ はそれぞれ雑音に関する基底行列、分離された磁励音の結合係数行列である。最後に、磁励音の結合係数行列 G の周波数方向の総和を計算した結果を特徴量として乗車車両推定ブロックに入力する。

実際には学習フェーズにおいて磁励音とは無関係な成分が含まれていると分離性能が低下するため、磁励音とは無関係な成分を除去する前処理を施した。具体的には、教師音声からケプストラムを算出した後にクラスタリングを用いて磁励音と無関係な成分を除去した。スペクトログラム

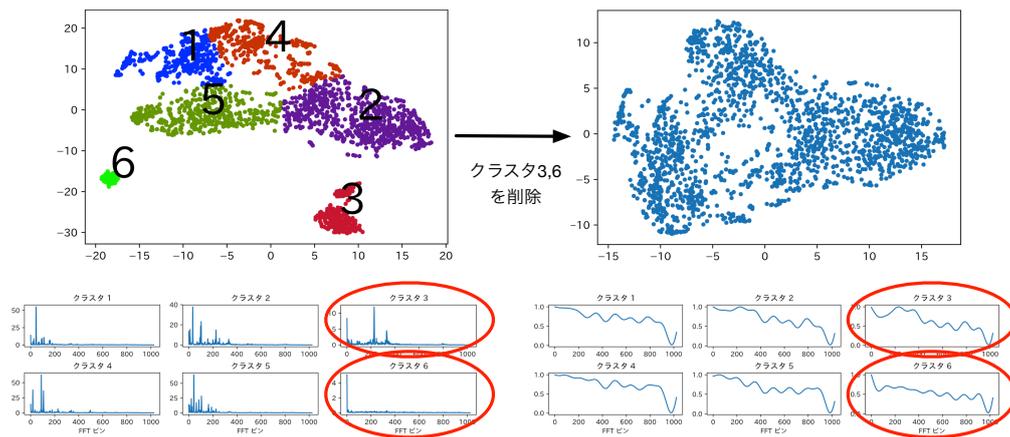


図 4 磁励音と無関係な成分除去の例

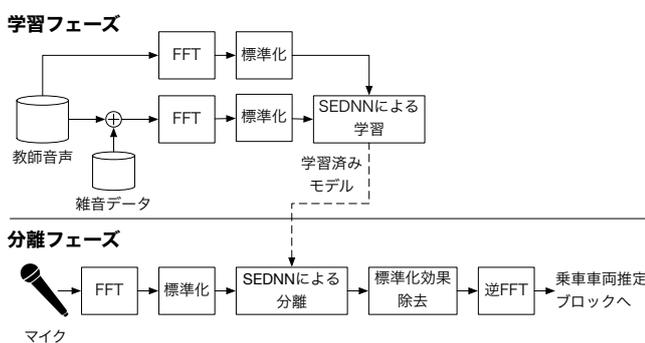


図 5 SEDNN を用いた磁励音分離の概要

を t -SNE (t -distributed Stochastic Neighbor Embedding) を用いて 2次元に次元削減した上で k 平均法を用いてクラスタリングし、手動で不要成分を特定した。

図 4 は磁励音と無関係な成分除去の例を示している。図左上に示すように $k = 6$ として k 平均法によりクラスタリングし、図左下・右下に示すように各クラスタごとのスペクトラム・ケプストラムを確認した。図 4 のクラスタ 3 は基本周波数よりも高い周波数成分にピークを持ち、クラスタ 6 はスペクトラムにおいて直流成分のみを持つことから、磁励音とは無関係な成分と考えられる。このクラスタ 3,6 を削除することで、磁励音と無関係な成分の除去を行った。

本稿では、教師音声として小型電動機を使ってエミュレートされた磁励音を用いたため、分離フェーズにおける SNMF による分離の代わりに直交化罰則条件付き SNMF (Penalized SNMF: PSNMF) [16] を用いる場合も評価する。PSNMF は教師基底行列 F_{train} と雑音の基底行列 H の関係が無相関となるように罰則条件を設けることで分離性能を向上させる手法である。エミュレートされた磁励音と実際の磁励音は完全には一致しないことから、学習時に得られる教師基底行列では実際の磁励音を完全に表現することができず、磁励音の一部が雑音として分離されてしまう可能性がある。PSNMF により、磁励音の教師基底に類似する雑音の基底成分を取得させないことで、この問題を回避する。

4.3 SEDNN を用いた磁励音の分離

図 5 に、SEDNN を用いた磁励音分離の概要を示す。SNMF を用いる場合と同様に SEDNN を用いた磁励音分離も学習フェーズと分離フェーズで構成される。

学習フェーズでは、教師音声を用いて SEDNN [13,17,18] により磁励音を分離するニューラルネットワークモデルを学習する。教師音声を固定幅のウィンドウに区切った上で FFT により対数振幅スペクトラムを算出する。対数振幅スペクトラムデータは標準化して SEDNN のニューラルネットワークに入力し、学習を行う。このとき、教師用音声に雑音データを重畳した対数振幅スペクトラムを入力として学習を行うことで、雑音を含む音声から磁励音を抽出するモデルを学習する。雑音音声は教師音声に対して $-5, 0, 5, 10, 15, 20$ [dB] の 6 レベルで混ぜ合わせたデータを作成して学習を行った。

分離フェーズでは、学習したニューラルネットワークモデルを用いて磁励音を分離する。学習フェーズと同様に、マイクから得られた音声を FFT により対数振幅スペクトラムに変換した上で、標準化処理を施してからニューラルネットワークに入力する。ニューラルネットワークが出力する磁励音は標準化の影響を受けているため、標準化効果除去処理を施し、逆 FFT により時間領域の磁励音に戻した上で乗車車両推定ブロックに入力する。

5. 乗車車両推定ブロック

乗車車両推定ブロックでは、分離後の磁励音音声から各種特徴量を算出し、機械学習を用いて乗車車両の推定を行う。特徴量には、分離後の音声をそのまま用いる場合に加え、フーリエ変換を用いるものとメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficients: MFCC) を用いるものの 2 つの特徴量を採用する。また、機械学習手法には、 k 近傍法 (k -Nearest Neighbor algorithm: k -NN) と Support Vector Machine (SVM) の 2 つの手法を用いる。これら複数の手法について車両推定精度を比較し、より高精度に乗車車両の推定が可能な手法を構築する。

表 1 各音声と手法に対する SDR[dB](SNR=5 時)

	A	B	C	D	E
SNMF($\beta=0$)	5.2	3.22	3.69	1.12	0.66
PSNMF($\beta=0$)	5.05	0.6	3.03	1.4	1.01
SNMF($\beta=1$)	6.31	6.15	3.63	1.43	-0.2
PSNMF($\beta=1$)	6.26	6.19	3.47	1.48	4.89
SNMF($\beta=2$)	7.01	6.74	5.71	2.87	3.28
PSNMF($\beta=2$)	7.08	6.88	6.05	2.79	3.26
SEDNN	8.27	9.17	7.34	6.36	6.62

表 2 各音声と手法に対する SIR[dB](SNR=5 時)

	A	B	C	D	E
SNMF($\beta=0$)	8.77	7.79	11.34	7.68	8.23
PSNMF($\beta=0$)	9.56	9.92	10.73	7.82	8.95
SNMF($\beta=1$)	6.56	6.51	12.58	6.74	6.18
PSNMF($\beta=0$)	6.5	6.52	12.18	6.99	5.69
SNMF($\beta=2$)	7.5	7.29	11.46	6.84	7.01
PSNMF($\beta=0$)	7.57	7.41	11.39	6.81	7.14
SEDNN	9.83	11.8	12.35	11.72	12.38

表 3 各音声と手法に対する SAR[dB](SNR=5 時)

	A	B	C	D	E
SNMF($\beta=0$)	8.26	5.76	4.81	2.89	2.1
PSNMF($\beta=0$)	7.41	1.56	4.19	3.18	2.29
SNMF($\beta=1$)	19.8	18.05	4.46	3.77	1.88
PSNMF($\beta=1$)	19.9	18.43	4.35	3.7	1.36
SNMF($\beta=2$)	17.5	16.78	7.36	5.92	6.47
PSNMF($\beta=2$)	17.49	16.98	7.86	5.81	6.31
SEDNN	13.9	12.88	9.24	8.14	8.21

フーリエ変換による特徴量の算出では、分離後の磁励音音声に対し振幅スペクトルを求め、低周波成分のみを取り出す。磁励音音声の音量変化の概形が乗車車両の推定に重要であると考えられるため、低周波成分のみを取り出す事で、この性質をよく表現する特徴量となると考えられる。

MFCC を用いる場合は、分離後の磁励音音声全体に対し MFCC を算出し、特徴量とする。MFCC は、周波数領域におけるスペクトル形状に関する特徴量であることから、本研究においても、フーリエ変換を用いた特徴量の場合とは異なり、スペクトル形状に特徴を持つ場合は良い特徴量となると考えられる。

この算出した特徴量を入力とし、k-NN と SVM を用いて乗車車両の予測を行う。特に、単純なアルゴリズムである k-NN と比較的良好な精度を示す傾向にある SVM を用いて識別精度を比較することで、適用する機械学習手法の違いによる推定精度への影響について検討を行う。

6. 評価

6.1 分離性能の評価

6.1.1 評価手法

4.2 節で述べた SNMF を用いる手法と 4.3 節で述べた SEDNN を用いる手法に対して、磁励音の分離性能の評価を行う。SNMF を用いる手法では、文献 [12,16] において β -divergence による擬距離を用いた目的関数が定義されており、特に、 $\beta=0,1,2$ の場合について議論が行われている。このことから、本研究においても、SNMF と PSNMF について $\beta=0,1,2$ と設定した場合の合計 6 手法を対象に分離性能の比較を行う。

これら合計 7 手法の磁励音教師音声には、[19] に掲載されている異なる 3 種類の車両形式を実機再現した音声を用いた。磁励音教師音声のスペクトログラムを図 6 に示す。また、SEDNN の学習時に必要な雑音の教師音声には、[13] 内で提案されたものと同様に [20] で提供される 100 個のデータと、Google Audioset [21] の駅構内で観測されそうな 18 種類のカテゴリから 85 個のデータを取得し、合計 185 個用いた。

分離性能の評価には [22] で提案されている、Source to Distortion Ratio (SDR), Source to Interferences Ratio (SIR), Sources to Artifacts Ratio (SAR) という 3 つの評価指標を用いる。これらの指標による分離性能の評価を行うためには、目的音と雑音が混合したデータ、目的音のみ

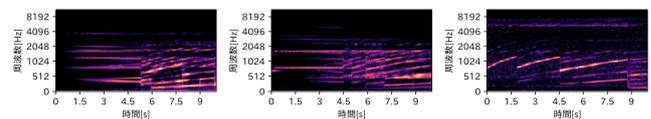


図 6 学習に用いる磁励音音声

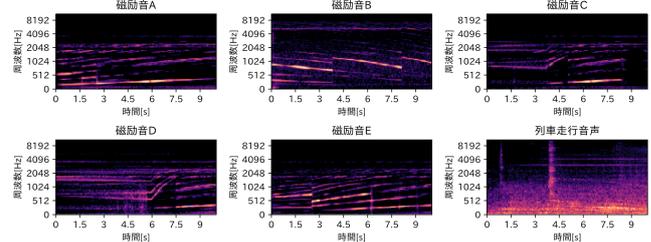


図 7 分離性能の検証に用いる磁励音音声と電車走行音声

のデータ、雑音のみのデータの 3 つが必要となる。

実際には、電車走行時の音声に対し目的音のみ、および雑音のみの完全な真値データを得ることは不可能である。そのため本研究では、磁励音発生の原因となる装置を搭載しない電車の走行時音声と、[19] に掲載されている磁励音のみの音声を任意の信号/ノイズ比 (Signal-to-Noise Ratio: SNR) で合成することにより模擬的な雑音重畳音声を作成する。磁励音のみの音声には、学習時に用いたものと同車両形式の磁励音を実機再現した 2 つの音声と、異なる車両形式の磁励音を実機再現した 3 つの音声の合計 5 つの音声を用いた。これら各磁励音音声と電車走行音声のスペクトログラムを図 7 に示す。以降、これらの磁励音は A-E の名称で示す。ここで、A および B は学習時と同一車両形式の磁励音音声であり、C, D, E は異なる車両形式に対する磁励音音声である。この模擬的な雑音重畳音声に対し、それぞれ SDR, SIR, SAR の算出を行う。

6.1.2 評価結果

表 1, 表 2, 表 3 に、実際の電車走行時の音声に最も近い

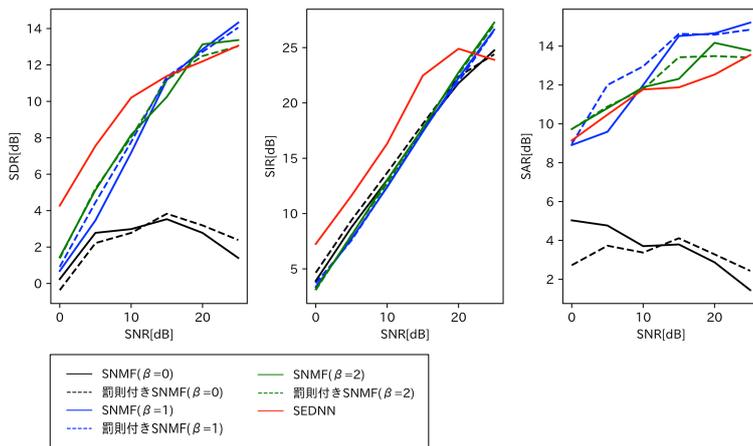


図 8 全ての SNR に対する平均的な分離性能



図 9 録音環境

と感じられた SNR=5[dB] の場合の全音声に対する SDR, SIR, SAR の性能を示す。

SNMF による手法では、 $\beta = 2$ のパラメータの時に最も安定した分離性能を示していることがわかる。また、SNMF における罰則項の追加による分離性能の向上はほとんど見られず、SDR, SIR, SAR 全てにおいて分離性能が降下するものも存在した。車両形式の違いにおける SIR の有意な差は見られず、学習に用いなかった車両形式である磁励音 C に対してもかなり高い精度で分離可能であった。一方、SDR と SAR に対しては学習時に用いたものと同じ車両形式の音声の方が比較的高い性能を示していることから、SDR において若干の違いが生じることがわかった。

SEDNN による手法では、磁励音 C を除く全ての磁励音音声に対して SDR および SIR が高い性能を示している。特に SIR については SNMF による手法と比べても平均的に 2-3[dB] ほど高い性能を示していることから、より目的音と雑音の分離が可能であることがわかった。また、SAR について、学習時と同じ車両形式に対しては SNMF を用いた手法より低い性能となっているが、異なる車両形式に対しては、SNMF ほど性能が低下せず良い性能が得られた。

次に、SNR を 0-25[dB] に設定した時の分離性能の平均をとったものを図 8 に示す。

SNMF による手法では、SDR と SAR のグラフから、 $\beta = 0$ の時に著しく性能が落ちることがわかった。一方、 $\beta = 1$ および $\beta = 2$ の場合は、どの SNR でも SDR, SIR の両方において安定した性能を示しており、特に $\beta = 2$ は SNR=20[dB] までの SDR についてわずかに優位な結果を示している。また、SIR については、どの手法も全 SNR に対し平均的に 2-3[dB] 程度の分離性能があることがわかった。SAR については $\beta = 1$ の時の罰則付き SNMF が良い結果を示すことがわかった。

SEDNN による手法では、SDR について低 SNR 時に平均的に 5-7[dB] 程度の分離性能がある一方で、SNR=10[dB] あたりを境に SNMF による手法に比べて性能が低くなっている。また、SAR については全体的に低い性能を示し

ており、特に、SNR=20[dB] の地点からは $\beta = 1$ および $\beta = 2$ の SNMF による手法と比べて大きく劣っている。

これらの結果から、学習時と同一の車両形式だけに限らず、異なる車両形式の磁励音に対しても分離可能であることがわかった。今回 SNMF の学習時には磁励音の音色が共通する成分だけを残したため、より一般的な磁励音の特徴を得られたことがこの結果につながったと考える。また、SEDNN による手法についても、異なる車両形式に対して分離可能であり、学習の過程で磁励音共通の特徴を自動的に得られている可能性がある。

さらに、SNMF は $\beta = 2$ の時に安定して分離が行えること、SEDNN と SNMF では条件次第でどちらも性能の良し悪しが存在することがわかった。

次節からの乗車車両推定精度の評価では、本項における評価でより優れた分離性能を示した SNMF($\beta = 2$) と SEDNN の 2 つの手法を用いて磁励音の分離を行いそれぞれに対して推定精度を比較する。

6.2 乗車車両推定精度の評価

6.2.1 評価手法

提案する乗車車両推定手法の性能評価を行うため、JR 九州福吉駅および JR 九州周船寺駅において音声データを取得し、その音声データを用いて実証評価を行った。図 9 に実験時の環境を示す。データ取得者はプラットフォーム上の無作為に選んだ場所にスマートフォンを持って立ち、電車が前方を通過してから完全に停止するまでの走行音声を録音した。今回対象とした電車は JR 九州 305 系電車 (6 両編成) であり、1 車両につき 12 個、全車両合計で 72 個の音声データを取得した。

取得したそれぞれの音声について、6.1.2 項で良い分離性能を示した。SNMF($\beta = 2$) および SEDNN による手法によって磁励音部分を分離した後、特徴量の算出を行う。この特徴量算出後のデータ全てに対し Leave-one-out cross validation 法を用いて k-NN および SVM による乗車車両の推定を行う。

表 4 各識別手法の推定精度 (SNMF 適用時)

	正解率 [%]	平均推定誤差 [車両]
no feature + k-NN	71	0.4
FT + k-NN	78	0.33
MFCC + k-NN	60	0.63
no feature + SVM	75	0.28
FT + SVM	78	0.28
MFCC + SVM	53	0.63

表 5 各識別手法の推定精度 (SEDNN 適用時)

	正解率 [%]	平均推定誤差 [車両]
no feature + k-NN	17	2.5
FT + k-NN	13	2
MFCC + k-NN	67	0.38
no feature + SVM	24	1.01
FT + SVM	8	1.81
MFCC + SVM	70	0.36

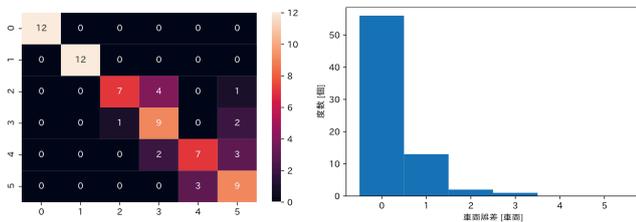


図 10 FT+SVM の場合の混合行列と車両推定誤差の分布 (SNMF 適用時)

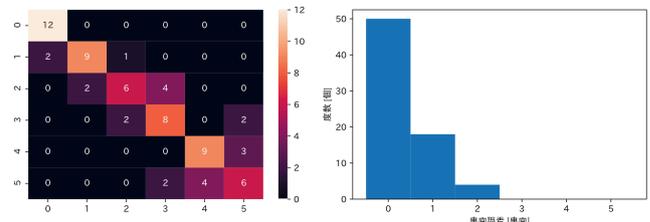


図 12 MFCC+SVM の場合の混合行列と車両推定誤差の分布 (SEDNN 適用時)

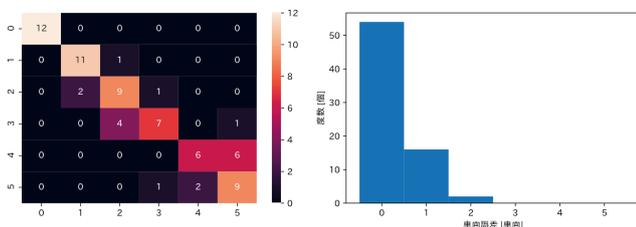


図 11 no feature+SVM の場合の混合行列と車両推定誤差の分布 (SNMF 適用時)

評価指標には、以下の式で示される正解率および平均推定誤差を用いた。この時、 $tr(i)$ は i 番目のデータに対し予測値が正しい場合は 1 を返す関数、 y_i は i 番目のデータに対する真値、 P_i は i 番目のデータに対する推定値、 N はデータの総数である。

$$tr(i) = \begin{cases} 1 & (y_i = P_i) \\ 0 & (otherwise) \end{cases} \quad (3)$$

$$\text{正解率} = \frac{\sum_i^N tr(i)}{N} \quad (4)$$

$$\text{平均推定誤差} = \frac{\sum_i^N |y_i - P_i|}{N} \quad (5)$$

6.2.2 評価結果

表 4 に、SNMF($\beta = 2$) による磁励音の分離を行った場合の、各手法の組み合わせによる正解率と平均推定誤差を示す。ここで特徴量を算出しなかった場合を「no feature」、フーリエ変換を用いた特徴量の算出を「FT」と表記している。最も良い性能を示した組み合わせは、FT と SVM を用いた場合であり、正解率 78%、平均推定誤差 0.28 車両で推定可能であった。この組み合わせの場合の混合行列と車両推定誤差の分布を図 10 に示す。正解しなかった場合の推定誤差について見ると、2 車両の誤差があったものは 2 個、3 車両の誤差があったものは 1 個存在した。また、2 番目に良い性能を示した組み合わせは、特徴量を算出せ

ず SVM を用いたものであった。この組み合わせについての混合行列と車両推定誤差の分布を図 11 に示す。最も良かった組み合わせと比べ、正解率は 3% 減少するが、平均推定誤差は変わらない。特に、車両推定誤差については、ほぼ 1 車両の誤りのみであり、僅かに 1 つだけが 2 車両の誤りであった。このことから、FT による特徴量の算出によって正解率の向上は見込めるものの、僅かに車両推定誤差が増幅される可能性があることがわかった。また、特徴量による違いは正解率と平均推定誤差の両方の性能に大きく関わっているが、単純なアルゴリズムである k-NN でも高精度に推定できており、機械学習手法の違いによる性能の差はあまり見られなかった。

次に、SEDNN による磁励音の分離を行った場合の、各手法の組み合わせによる正解率と平均推定誤差について表 5 に示す。この場合、最も良い性能を示した組み合わせは、MFCC と SVM を用いた場合であり、正解率 70%、平均推定誤差 0.36 車両で推定可能であった。この組み合わせの場合の混合行列と車両推定誤差の分布を図 12 に示す。上述の SNMF+FT+SVM の組み合わせによる手法に比べ、正解率は 8 ポイント減少した。また、車両の推定誤差では、1 車両の誤差以外に僅かに 2 つだけ 2 車両の誤差が発生した。SEDNN による分離を行った場合には、特徴量の算出に MFCC を用いる組み合わせのみが良い結果を示すことがわかった。SNMF による分離を行った場合と同様に、特徴量の算出の方法によって正解率と平均推定誤差の精度は大きく変化する傾向にあるが、機械学習手法の違いによる精度の差はあまり見られなかった。

これらの結果から、SEDNN に比べ磁励音の分離性能が劣る傾向にあった SNMF を用いた場合に、より高い推定性能を示すことがわかった。この理由の 1 つとして、SNMF では分離後音声に対して時間領域への復元を行っておらず、磁励音に対する基底結合係数行列 G の情報をそのまま扱ったことが有効に働いたと考えられる。この行列は、スペク

トログラムに対しては各周波数基底の時間的変化を表すということが一般的に知られているが、周波数方向で総和をとった場合には、ある一定時間間隔ごとの磁励音の振幅変化を表現したものとなる。これによって、結果的に音量変化を強調するような効果が得られたと考える。また、音声への復元なしに、この行列 G をそのまま利用する場合、音声への復元処理や次元削減処理、MFCC などの複雑な特徴量の算出処理が必要なくなるため、計算効率やその処理時間の面においても大きな利点になることが考えられる。

7. おわりに

本研究では、車両ごとの混雑度推定手法を構築することを目的とし、マイクロフォンを用いた乗車車両推定手法について検討を行った。特に、観測音声から磁励音のみを分離した上で、磁励音の変化を機械学習することで外的雑音に対して頑健な手法を実現した。複数の手法について車両推定精度を比較し、SNMF による磁励音の分離後、フーリエ変換を行なって低周波成分を抽出したものを特徴量として SVM を適用することで、正解率 78%、平均推定誤差 0.28 車両の精度で推定可能であることを確認した。

今後の課題として、学習データの種類増加による分離性能の向上について検討を行う必要がある。また、車両形式や車両数によって主電動機の搭載される位置が変わる場合があるため、電車走行音声から車両形式の推定や車両数の推定を行う手法に対しても今後検討を行う必要がある。本研究で提案した分離手法では、車両形式によらず磁励音を分離・強調可能であったことから、このような推定手法について検討する際にも、まず磁励音を分離・強調することで、より良い推定精度が得られると考える。

謝辞 本研究の一部は、JSPS 科研費 JP15H05708 の助成を受けたものである。

参考文献

- [1] 国土交通省: 三大都市圏の主要区間の混雑率, <https://www.mlit.go.jp/common/001245347.pdf>.
- [2] 篠原菊紀, 蹴揚秀男: 交通の健康学的影響に関する研究, *PRI review*, Vol. 国土交通政策研究 第 55 号, No. 16, pp. 2-9 (2005).
- [3] 国土交通省: 遅延の「見える化」を開始, <http://www.mlit.go.jp/common/001215328.pdf>.
- [4] 西村友洋, 樋口雄大, 山口弘純, 東野輝夫: スマートフォンを活用した屋内環境における混雑センシング, *情報処理学会論文誌*, Vol. 55, No. 12, pp. 2511-2523 (2014).
- [5] 西村友洋, 樋口雄大, 山口弘純, 東野輝夫: スマートフォンを活用した屋内混雑センシングの実装と評価, *情報処理学会研究報告マルチメディア通信と分散処理研究会報告*, Vol. 2014, No. 9, pp. 1-8 (2014).
- [6] 前川勇樹, 内山 彰, 山口弘純, 東野輝夫: 鉄道における Bluetooth RSSI 特性を用いた乗車車両および混雑の推定手法, *情報処理学会論文誌*, Vol. 55, No. 6, pp. 1614-1624 (2014).
- [7] 佐藤孝嗣, 石田繁巳, 田頭茂明, 福田 晃: マイクロフォンを用いた鉄道乗車位置推定手法の設計と評価, *研究報告高度交通システムとスマートコミュニティ (ITS) 2019-ITS-76*, No. 5, pp. 1-7 (2019).
- [8] Nirjon, S., Liu, J., DeJean, G., Priyantha, B., Jin, Y. and Hart, T.: COIN-GPS: indoor localization from direct GPS receiving, *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pp. 301-314 (2014).
- [9] 株式会社 NTT ドコモ: 報道発表資料, https://www.nttdocomo.co.jp/info/news_release/2014/03/04_00.html (2014).
- [10] Sato, K., Ishida, S., Kajimura, J., Uchino, M., Tagashira, S. and Fukuda, A.: Initial Evaluation of Acoustic Train Detection System, *ITS Asia-Pacific Forum*, pp. 1092-1103 (2018).
- [11] Sato, K., Ishida, S., Kajimura, J., Tagashira, S. and Fukuda, A.: Proposal of Acoustic Train Detection System for Crowdsensing, *Intelligent Transport Systems for Everyone's Mobility* (Mine, T., Fukuda, A. and Ishida, S., eds.), Springer, pp. 427-444 (2019).
- [12] 八木浩介, 猿渡 洋, 鹿野清宏, 近藤多伸, 高橋 祐: 教師音として抽出対象信号情報を用いた非負値行列因子分解による目的楽器音信号の抽出, *電子情報通信学会技術研究報告. EA, 応用音響*, Vol. 111, No. 136, pp. 21-26 (2011).
- [13] Xu, Y., Du, J., Dai, L.-R. and Lee, C.-H.: A Regression Approach to Speech Enhancement Based on Deep Neural Networks, Vol. 23, No. 1 (2015).
- [14] 岸本隆博: 電車制御時の発生騒音, 騒音制御, Vol. 29, No. 4, pp. 287-289 (2005).
- [15] Varghese, S.: Electromagnetic Noise from Permanent Magnet Motors, Master's thesis, KTH, Aeronautical and Vehicle Engineering (2013).
- [16] Yagi, K., Takahashi, Y., Saruwatari, H., Shikano, K. and Kondo, K.: Music signal separation by orthogonality and maximum-distance constrained nonnegative matrix factorization with target signal information, *AES 2011: The 45th International Conference on Applications of Time-Frequency Processing in Audio, 1-4 March, 2012, Helsinki, Finland.*, pp. 142-147 (2011).
- [17] Xu, Y., Du, J., Dai, L.-R. and Lee, C.-H.: An Experimental Study on Speech Enhancement Based on Deep Neural Networks, *Signal Processing Letters, IEEE*, Vol. 21, pp. 65-68 (online), DOI: 10.1109/LSP.2013.2291240 (2014).
- [18] Xu, Y., Du, J., Huang, Z., Dai, L.-R. and Lee, C.-H.: Multi-Objective Learning and Mask-Based Post-Processing for Deep Neural Network Based Speech Enhancement, *Interspeech2015*, (online), DOI: 10.1109/LSP.2013.2291240 (2015).
- [19] 床下電子音楽 - VVVF Synth Sound : <http://vvvf.webcrow.jp/products/> (2019).
- [20] Hu, G.: 100 nonspeech environmental sounds, <http://www.cse.ohio-state.edu/pnl/corpus/HuCorpus.html> (2004).
- [21] Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M. and Ritter, M.: Audio Set: An ontology and human-labeled dataset for audio events, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 776-780 (2017).
- [22] Vincent, E., Gribonval, R. and Févotte, C.: Performance measurement in blind audio source separation, *IEEE Trans. Audio, Speech, Language Process.*, Vol. 14, No. 4, pp. 1462-1469 (2006).