

分散型マイクロホンアレイを用いた交通車両検出とその車線推定の検討*

☆豊田卓矢 (筑波大), 小野順貴 (NII/総研大), 宮部滋樹, 山田武志, 牧野昭二 (筑波大)

1 はじめに

交通量調査は渋滞緩和や道路整備などのために重要であり、定点観測、パーソントリップ調査、アンケート調査、プローブ調査などの様々な調査方法がある [1]。このうち定点観測は、特定の道路の特定の場所を通過する車両数を調査するものであり、人手で数え上げる方法の他に、センサで計測した情報の処理による自動化が取り組まれている。定点観測に用いられるセンシング方式としては、ループ型センサ、超音波センサ、赤外線センサ、地震計、カメラによって撮影された動画に対する画像処理などが代表的なものとして挙げられる。一般に、こうしたセンシングシステムにおいては、設置に要するコストが少なく、また、昼夜の別や天候などの環境変動に対して頑健な方式が望まれている。

我々は、低コストで簡便な交通量モニタリングのための手法として、音響センシングに着目している。車両の走行音をマイクロホンで観測する音響センシングは、振動センサを埋め込むタイプのもの [2] を除くと、設置のためのコストを比較的安く抑えることができる利点がある。従来研究としては、車種分類を目的として、事前学習した車両走行音を用いて Time Delay Neural Network を適用する手法 [3] や、音響特徴量と振動特徴量を組み合わせた Support Vector Machine [4] などが試みられている。車両検出を目的としたものとしては、例えば単一チャンネルの短時間パワーによる車両検出 [5] は、単一マイクロホンを用いているために設置コストの面で有利であるが、車両の走行方向などを認識することができない。一方、大口径マイクロホンアレイによる時間差分を用いた接近車両検出 [6] なども行われているが、装置規模が大型になるという問題がある。

我々は、多チャンネル観測を容易に実現するため、Fig. 1 に示すような非同期分散型マイクロホンアレイ [7, 8] を用いた交通量モニタリング [9] を検討している。本手法では、まず各機器で録音した信号を同期 [10, 11, 12] するために、単一音源区間を用いてサンプリング周波数ミスマッチや録音開始時刻オフセットの補償を行う [12]。次に、各チャンネルでエネルギーエンベロープを求め、ピーク検出を行うことで交通車両数を推定する。また、多チャンネル情報を利用した、車線位置の推定のため、チャンネル間のエネルギー比を用いた手法を検討し、実際の公道で計測したデータを用いた実験により、時間差を用いた前報告での手法 [9] と比較を行う。

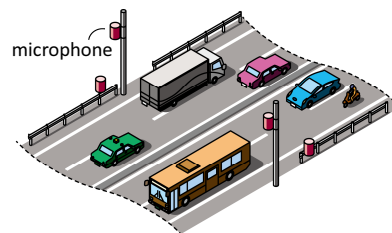


Fig. 1 Traffic monitoring with ad-hoc microphone array

2 チャネル同期

非同期マイクロホンアレイの枠組みでは、非同期録音された多チャンネル信号でアレイ信号処理を行うために、その同期化がまず重要となる。そこで、始めに二組の単一音源区間を用いた同期手法 [12] について述べる。

いま、同時刻における2つのマイクロホンの連続信号 $x_1(t)$ 、 $x_2(t)$ (t は連続時刻) が異なる A/D 変換器でサンプリングされて離散信号 $x_1[n_1]$ 、 $x_2[n_2]$ (n_1 、 n_2 は離散時刻) が得られたとする。ここで、 $x_1[n_1]$ のサンプリング周波数は f_s 、 $x_2[n_2]$ のサンプリング周波数は未知の無次元パラメタ ϵ により表される $(1+\epsilon)f_s$ であるとする。ただし、二つの A/D 変換器は共通の公称サンプリング周波数を持ち、 $|\epsilon| \ll 1$ であると仮定する。このとき

$$x_1[n_1] = x_1\left(\frac{n_1}{f_s}\right) \quad (1)$$

$$x_2[n_2] = x_2\left(\frac{n_2}{(1+\epsilon)f_s} + \Delta T_{21}\right) \quad (2)$$

となる。ここで、 ΔT_{21} は $x_1[n_1]$ に対する $x_2[n_2]$ の録音開始時刻の遅れを表す。また、チャンネル1とチャンネル2で同一の連続時刻 t を参照する離散時刻 n_1 、 n_2 は

$$n_1 = t f_s \quad (3)$$

$$n_2 = (1+\epsilon)(n_1 - D_{21}) \quad (4)$$

$$D_{21} = \Delta T_{21} f_s \quad (5)$$

と表される。ここで、 D_{21} はチャンネル2で録音を開始したときのチャンネル1での離散時刻を表す。もし二組のある時刻 $\{n_{A1}, n_{A2}\}$ と $\{n_{B1}, n_{B2}\}$ がそれぞれ同一の連続時刻に対応していれば、サンプリング周波数ミスマッチ ϵ と録音開始時刻オフセット D_{21} をそれぞれ求めることができる。これらの変数は

$$n_{A2} = (1+\epsilon)(n_{A1} - D_{21}) \quad (6)$$

* Vehicle detection and its lane estimation with ad-hoc microphone array. by Takuya TOYODA (University of Tsukuba), Nobutaka ONO (National Institute of Informatics / The Graduate University for Advanced Studies), Shigeki MIYABE, Takeshi YAMADA, Shoji MAKINO (University of Tsukuba)

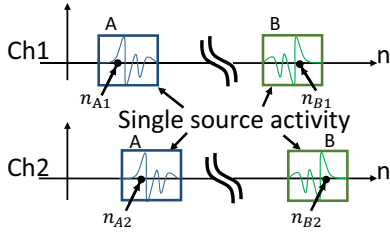


Fig. 2 Two pairs of single source activities

$$n_{B2} = (1 + \epsilon)(n_{B1} - D_{21}) \quad (7)$$

を満たし、 ϵ と D_{21} は

$$\epsilon = \frac{n_{B2} - n_{A2}}{n_{B1} - n_{A1}} - 1 \quad (8)$$

$$D_{21} = \frac{n_{A1}n_{B2} - n_{A2}n_{B1}}{n_{B2} - n_{A2}} \quad (9)$$

と表せる。このように、二組の対応する時刻 n_{Ai} と n_{Bi} ($i=1,2$)を用いることで、 ϵ と D_{21} が求まる。このような時刻のペアは、両チャンネルで同一の単一音源が観測される時刻を同一とみなした時刻の対応によって求める。まずFig. 2に示すように、同一の単一音源が録音される2つの短い時間区間A、Bそれぞれの波形を各チャンネルから手で切り出し、相互相関の最大化によって切り出した区間内の単一音源の時間差を求める。この時間差をもとにして切り出した区間の時刻の対応を求め、同一時刻に対応する離散時刻のペア $\{n_{A1}, n_{A2}\}$ と $\{n_{B1}, n_{B2}\}$ を推定する。実際には単一音源録音は到来時間差の影響により厳密に同一時刻とはならないため、音源とマイクロホンの位置関係をもとにして適切に補償する必要がある。

サンプリング周波数ミスマッチ ϵ と録音開始時刻のオフセット D_{21} が得られたのちには、短い分析フレーム内ではサンプリング周波数ミスマッチの影響が無視できるものと仮定して、フレームの切り出し時刻を補償する形で同期を補償したフレーム分析を行う。まずチャンネル1については通常のフレーム分析を行い、チャンネル2については、チャンネル1の分析フレームの先頭時刻 n_1 と式(4)の対応関係にある n_2 を先頭としたフレームの切り出しを行う。整数の n_1 に対して n_2 は一般に非整数となるため、まず n_2 を整数に丸めた時刻でフレームの切り出しを行い、丸めによる時刻の小さな誤差は全域通過の線形位相フィルタを用いてフレーム毎に補償する。このようにして同期を補償した分析フレームのオーバーラップ処理によって、同期を補償した長時間の波形を得る。

3 音響イベント検出

定速走行している車両がマイクロホンに最も接近したときに車両走行音のエネルギーエンベロープが極大になると期待できるため、車両走行音のエネルギーエンベロープのピークを検出することで交通車両数を計測することを試みる。周囲の雑音や観測値

の変動の影響によりエネルギーエンベロープは多数の偽のピークを持つため、ピーク検出の前処理として雑音抑圧と信号の平滑化を行う。

3.1 ウィーナーフィルタによる雑音抑圧

ここでは車両走行音と雑音のパワースペクトルの性質に着目し、線形時不変ウィーナーフィルタの適用を検討した。

車両走行音と雑音のエネルギーの定常性を仮定し、SNRが高い帯域を強調するウィーナーフィルタ $W(f)$ を以下のように設計する。

$$W(f) = \frac{S_V(f)}{S_V(f) + S_N(f)} \quad (10)$$

ここで、 $S_V(f)$ 、 $S_N(f)$ はそれぞれ車両走行音のパワースペクトル、雑音のパワースペクトルを表し、 f は周波数インデックスを表す。そして、 i 番目のマイクロホンで観測した信号を $X_i(f, k)$ とすると(k はフレーム番号を表す)、雑音抑圧された信号は

$$\hat{X}_i(f, k) = W(f)X_i(f, k) \quad (11)$$

のように得られる。

実際には、まず車両走行音と雑音が観測された時刻を、車両走行音の録音とともに車両の走行を撮影した動画をもとに見つけ、それぞれ前後数秒を短時間フーリエ変換(STFT)する。そして、周波数帯域ごとの絶対値2乗のフレーム平均を求め、走行音のものを $S_V(f)$ 、雑音のものを $S_N(f)$ とする。ここで、 $f = 0, \dots, F-1$ であり、 F はSTFTのフレーム長を表す。

3.2 ガウス窓型FIRフィルタによる平滑化とエネルギーエンベロープのピーク検出

エネルギーエンベロープのピークの数を実測するために、以下のように求まる信号エネルギーの時系列について考える。

$$Y_i(k) = \frac{1}{F} \sum_{f=0}^{F-1} |\hat{X}_i(f, k)|^2 \quad (12)$$

エネルギーエンベロープのピーク計測のシンプルな方法としては、 $Y_i(k)$ のピークを検出することが考えられるが、エネルギーの細かな変動によって車両通過時刻に対応するピークのみを実測することができない。したがってこの細かな変動を軽減するために、信号エネルギーの時系列を平滑化しエネルギーエンベロープを算出する。平滑化にはガウス窓型FIRフィルタ

$$g(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{m^2}{2\sigma^2}\right) \quad (13)$$

を用いる。ここで、 m 、 σ はそれぞれフレームインデックス、ガウス窓の標準偏差パラメータを表す。平滑化された信号は

$$\hat{Y}_i(k) = \sum_{m=-\frac{G}{2}}^{\frac{G}{2}} Y_i(k-m)g(m) \quad (14)$$

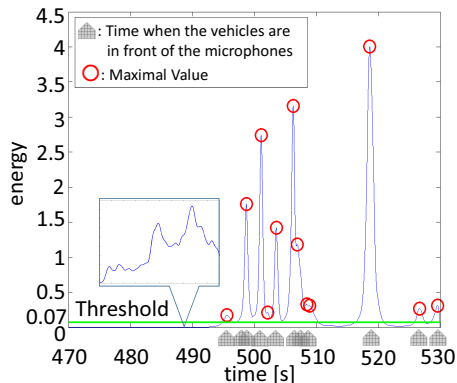


Fig. 3 Detected peaks with larger power than threshold

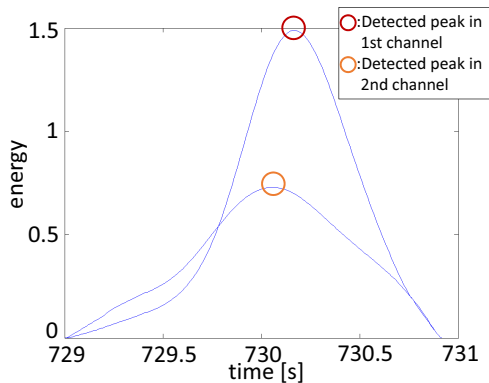


Fig. 4 Energy difference of peaks between channels

のように得られる。ここで G はガウス窓の窓幅を表す。このような平滑化処理によってエネルギーの変動は大幅に緩和されるものの、ごく小さい値の疑似的なピークが残ってしまう。そこで、このようなピークを無視するために、閾値を下回るピークを除外する。もし、 $\hat{Y}_i(k)$ が極大値でかつ閾値 h よりも大きな値ならばこれを $p_i(e)$ とし、 $s_i(e)$ をピークの時刻とする。ここで、 $p_i(e)$ 、 $(e = 0, \dots, E-1)$ はチャンネル i で検出したピークの値、 E は検出したピークの総数を表す。閾値処理によるピーク検出の実際の例を Fig. 3 に示す。Fig. 3 より、車両通過時にのみピークを検出できていることが分かる。

4 チャンネル間のピークのエネルギー比による車線推定

道路を挟んだ両側で録音した 2 チャンネルの信号を観測することにより、交通車両の走行車線を推定する方法について述べる。道路を挟んだ 2 つのマイクロホンで観測される車両走行音は、走行車線と各マイクロホンまでの距離の違いから、Fig. 4 に示すようにエネルギーがチャンネル間で異なったものとなる。ここではチャンネル間のエネルギー比を用いて走行車線の推定を行う。

まず、3.2 節で求めた各チャンネルのピーク時刻の系列 $s_i(e)$ の、チャンネル間の対応付けを行う。ピークの時刻差 $|s_1(e) - s_2(e)|$ が閾値 d を下回っている場合は、エネルギーのピークが同じ走行車両によって生

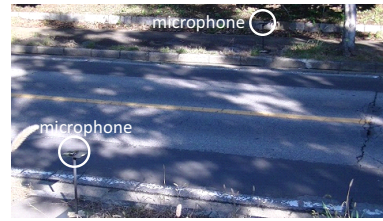


Fig. 5 A picture of the recording location

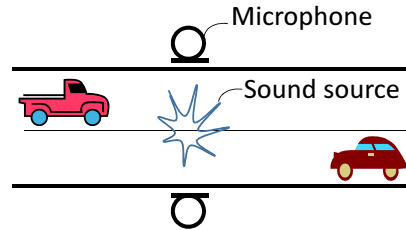


Fig. 6 Generating a single source activity for synchronization

Table 1 Experimental conditions

道幅	7.1 m
マイク間距離	9.5 m
録音時間	660 s
サンプリング周波数 f_s	48 kHz
STFT のフレーム長 F	2048 samples
STFT のフレームシフト幅	512 samples
平滑化フィルタのパラメタ σ	14
平滑化フィルタ長 G	$6\sigma+1$
閾値 h	0.07
ピーク対応付けの幅 d	38 frames
録音機器	SANYO ICR-PS603RM
ビデオカメラ	JVC GZ-HM670

成されたものであるとみなし、 $p_1(e)$ 、 $p_2(e)$ をそれぞれ $\hat{p}_1(e)$ 、 $\hat{p}_2(e)$ とする。ここで閾値 d は、走行車両の速度や道幅、マイク間距離、サンプリング周波数、STFT のフレーム長などを考慮して決定する。この条件を満たすピークが複数ある場合、時刻差が最小のものを対応付ける。本稿では、チャンネル 1 側の車線を右、チャンネル 2 側の車線を左と定義する。そして、 $\frac{\hat{p}_1(e)}{\hat{p}_2(e)} < 1$ のとき走行車線は左、 $\frac{\hat{p}_1(e)}{\hat{p}_2(e)} > 1$ のとき走行車線は右であると推定する。

5 実環境下での実験

5.1 実験条件

チャンネル間の時間差を用いた走行車線推定（前手法）による結果と、チャンネル間のピークのエネルギー比を用いた走行車線推定（提案手法）による結果とを比較した。Fig. 5 に示すように、筑波大学近辺の 2 車線道路において、対向に設置した二つのマイクロホンを用いて車両走行音の時間測定を行った。また、録音と同時に車両の走行を動画として撮影し、録画後に人手で車両の走行方向と通過時刻のラベリングを行った。動画をもとに車両走行音の時間波形の振幅が極大になる時刻を正解時刻と定めた。ここで、検出したピークの時刻と正解時刻との差が 0.5 秒以下のと

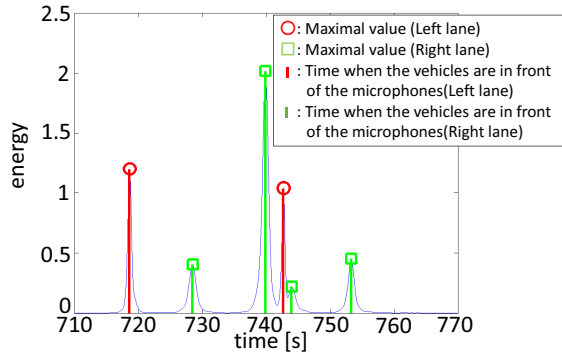


Fig. 7 Experimental result

Table 2 Numbers of vehicles in the ground truth and in the detection in the left traffic lane (left) and the right traffic lane (right) by previous method [9] (T = True, F = False)

Ground Truth				Ground Truth			
Detected vehicles		T	F	Detected vehicles		T	F
	T	35	5		T	22	5
F	3		F	4			

きに正確にピーク検出できたものとした。Fig. 6 に示すように道路の中央で手を叩き、生成された音響信号を単一音源として用いた。本実験ではバイクや自転車が走行していない区間の録音信号を使用した。推定結果より算出した F 値を用いて提案手法の評価を行った。ガウス窓型 FIR フィルタの標準偏差パラメタ σ と閾値 h は実験的に決定した。その他の実験条件は Table 1 に示す。

5.2 F 値の算出

正確性と網羅性の総合的な評価のため、以下で表される F 値を尺度として用いた。

$$\text{precision} = \frac{N_c}{N_e} \quad (15)$$

$$\text{recall} = \frac{N_c}{N_r} \quad (16)$$

$$\text{F-measure} = \frac{\text{precision} \cdot \text{recall}}{\frac{1}{2}(\text{precision} + \text{recall})} \quad (17)$$

ここで、 N_c は正確に推定された交通車両数、 N_e は推定された交通車両数、 N_r は実際に通過した交通車両数を表す。F 値は 0–1 の範囲の値をとり、高ければ性能がよいことを示す。

5.3 実験結果と考察

ある 60 秒の区間において提案手法により得られた結果を Fig. 7 に示す。Fig. 7 より音響イベント計測、走行車線推定ともに正しく行えていることが確認できる。前手法および提案手法によって得られた推定交通車両数と実際の走行車両数をそれぞれ Table 2、Table 3 に示す。そして、Table 2、Table 3 を用いて算出された F 値を Table 4 に示す。Table 4 より前手法に比べ提案手法の方が高い F 値が得られたこと

Table 3 Numbers of vehicles in the ground truth and in the detection in the left traffic lane (left) and the right traffic lane (right) by proposed method

Ground Truth				Ground Truth			
Detected vehicles		T	F	Detected vehicles		T	F
	T	36	4		T	26	0
F	2		F	0			

Table 4 Calculated F-measure

	F-measure
Previous method [9]	0.8702
Proposed method	0.9538

から、エネルギーのピーク探索に基づく車両検出と、エネルギー比を用いた走行車線推定の有効性を確認した。

6 おわりに

本稿では分散型マイクロホンアレイを用いた交通量モニタリングの走行車線推定の精度向上を図った。交通車両の走行車線をチャンネル間のピークのエネルギー比によって推定した。実験結果でチャンネル間の時間差を用いた前手法に比べ、提案手法の方がより高い F 値を得られたことから、提案手法の有効性が確認された。

参考文献

- [1] 飯田 他, “交通工学,” オーム社, 2008.
- [2] 荒井 他, “路面振動に着目した夏・冬タイヤの判別手法に関する基礎研究,” 信学技報 EA2013-105, pp. 11–14, 2014.
- [3] Nooralahiyan *et al.*, “A field trial of acoustic signature analysis for vehicle classification,” *Transpn Res-C*, vol. 5, no. 3/4, pp. 165–177, 1997.
- [4] Evans *et al.*, “Automated acoustic identification of vehicles,” *Proc. IOA*, pp. 238–245, 2008.
- [5] Sobreira *et al.*, “Automatic classification of traffic noise,” *Proc. IOA*, pp. 6221–6226, 2008.
- [6] 嶋田 他, “マイクロホンアレイを用いた音響センシングによる接近車両に関する一検討,” 信学技報 SIS2009-71, pp. 125–128, 2010.
- [7] Robledo-Arnuncio *et al.*, “On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation,” *Proc. WASPAA*, pp. 34–37, Oct. 2007.
- [8] Liu, “Sound source separation with distributed microphone arrays in the presence of clock synchronization errors,” *Proc. IWAENC*, 2008.
- [9] 豊田 他, “分散型マイクロホンアレイを用いた交通量モニタリング,” 信学総大 A-10-10, p. 151, 2014.
- [10] Markovich-Golan *et al.*, “Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming,” *Proc. IWAENC*, 2012.
- [11] Miyabe *et al.*, “Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain,” *Proc. ICASSP*, pp. 674–678, 2013.
- [12] Sakanashi *et al.*, “Speech enhancement with ad-hoc microphone array using single source activity,” *Proc. APSIPA*, 2013.