ステレオ録音に基づく移動音源モデルによる走行車両検出と走行方向推定*

☆遠藤純基, 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

交通管制システムや渋滞緩和のために、車道のある地点を単位時間に通過する車両の台数を示す交通量 [1] の情報は重要である。従来の交通量の観測手法には、人手による手法や、ループ型センサ、超音波センサ、赤外線センサ、画像処理技術を用いた定点観測手法などがある [1]。しかし、人手による観測は人件費がかかってしまい、また自動計測のためにはセンサの設置や管理に高いコストを要することが多いため、簡便な交通量の計測方法が求められている。

低コストで簡便な交通量の計測法として、音を用いる手法が考えられる。音響センシングは、単純にマイクロホンを設置するだけなので設置コストを低く抑えることができることと、録音を開始すれば計測は自動で行えるという利点がある。交通量モニタリングへの応用としても、従来から、単一チャネルの短時間パワー [2] やマイクロホンアレイによる時間差分析 [3] などによって車両の検出が試みられてきた。ただし、単一チャネルの場合は車両の走行方向の情報は得られず、また、マイクロホンアレイを用いるためには同時サンプリングが必要となり簡便性が損なわれる問題があった。

そこで我々は、身近なデバイスであるスマートフォンや安価な IC レコーダー等の録音機器に着目し、これらの機器を利用した音響センシングによる交通量モニタリングの手法の検討を行ってきた。これまでの検討では、同期していない複数の独立したモノラル録音機器を Fig. 1 のように両側 2 車線道路の路肩に配置し、パワーエンベロープのピークおよび録音機器間のピークの大きさの違いを分析することにより、交通車両数と交通車両の走行車線の推定を試みてきた [5]。しかし、複数の録音機器のパワーエンベロープのピークを比較するためには、事前処理として同期補償 [8] を正確に行う必要がある。正確な同期補償のためには、既知の同期用信号を複雑な手順で鳴らさなければならず、雑音抑圧のために車両走行音区間と雑音区間を手動で切り出す必要があるため手動の計測手順が多い。また、各録音機器のパワーエンベロープのピークを統合して車線推定を行うため、それぞれの録音機器で S/N 比の違いなどからピークが正しく検出されない場合、車線推定が正しく行われない問題があった。

本研究ではステレオ録音機器の使用を前提に、単一録音機器のチャネル間位相差の変化を利用して単一進行方向の走行音を強調したパワーエンベロープを分析することによる、装置間の同期が不要で音源移動を推定する計測手順の少ない車両検出手法を提案する。なお、本稿ではステレオマイクを従来法と同様に Fig. 1 のように設置した両側 2 車線道路における交通量を扱う。

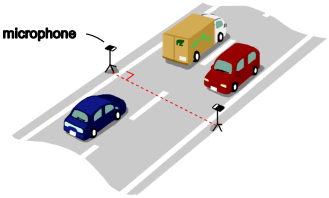

Fig. 1 Traffic monitoring with ad-hoc microphone-array.

2 従来手法と問題設定

2.1 従来手法の概要

非同期マイクロホンアレーによる交通量推定の従来手法 [5] は各機器で録音した信号の同期 [6, 7, 8] をとる必要があり、単一音源区間を用いてサンプリング周波数ミスマッチと録音開始時刻オフセットの補償を行う [8]。次に、各チャ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の推定を行う。また、チャネル間のエネルギー比を用いて交通車両の走行車線の推定を行う。ここでは本稿の提案手法と関連が深い、エネルギーエンベロープの分析とピーク検出について述べる。

2.2 音響パワーのピーク検出による交通車両数推定

定速走行する車両の走行音は、マイクロホンの正面を通過した時に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。ここで、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ウィーナーフィルタによる雑音抑圧とガウス窓型 FIR フィルタによる信号の平滑化を行う。なお以下では短時間フーリエ変換（STFT）領域の複素信号を扱う。

i 番目のマイクで観測した時間フレーム t における信号を $X_i(\omega, t)$ とすると、ウィーナーフィルタ $W(\omega)$ により雑音抑圧された信号は

$$\hat{X}_i(\omega, t) = W(\omega) X_i(\omega, t) \tag{1}$$

のように得られる。ウィーナーフィルタ $W(\omega)$ は、手動で切り出した車両走行音区間と雑音区間それぞれの観測パワースペクトル $S_V(\omega)$,$S_N(\omega)$ を用いて以下のように設計する。

$$W(\omega) = \frac{S_V(\omega)}{S_V(\omega) + S_N(\omega)} \tag{2}$$

---

*Vehicle detection and traveling direction estimation by moving sound source model based on stereo recording. by Junki ENDO(University of Tsukuba),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)

ここで、ω は周波数インデックスを表し、$\omega = 0, ..., \frac{L}{2}$ の値をとる。なお、L は短時間フーリエ変換のフレーム長を表す。

次に、エネルギーエンベロープのピークを計測するために、信号エネルギーの時系列を以下のように求める。

$$Y_i(t) = \frac{1}{U+1} \sum_{\omega=0}^{U} |\hat{X}_i(\omega, t)|^2 \tag{3}$$

ここで、U は帯域上限であり、この手法ではナイキストレート $\frac{L}{2}$ に設定している。$Y_i(t)$ のピークを検出することによってエネルギーエンベロープのピークを計測できると考えられるが、エネルギーの細かな変動により車両通過時刻に対応する真のピークのみを計測することができない。そこで、信号エネルギーの時系列を平滑化して細かな変動を軽減する。平滑化にはガウス窓型 FIR フィルタ

$$g(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{m^2}{2\sigma^2}\right) \tag{4}$$

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タを表す。平滑化された信号は

$$\hat{Y}_i(t) = \sum_{m=\frac{-G+1}{2}}^{\frac{G+1}{2}} Y_i(t-m) g(m) \tag{5}$$

のように得られる。ここで G はガウス窓の窓幅を表す奇数の値とする。平滑化処理によりエネルギーの変動は大幅に緩和できるが、雑音に起因する微小な擬似ピークが残ってしまう。そのためピーク検出に閾値 h を設け、h を上回るピークのみを計測する。計測したピーク時刻を車両が通過した時刻として、車両数を計測する。

2.3 チャネル間のパワー比による交通車両の車線推定

道路を挟んだ両側で録音した 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、各チャネルでエネルギーエンベロープの大きさが異なる。そのため、チャネル間のエネルギー比を用いて走行車線の推定を行う。**2.2** 節で求めた各チャネルごとのピーク時刻の系列を $s_i(e)$ とする ($i = 1, 2$、$e = 0, 1, ..., E-1$)。ここで、i はチャネル番号、E は検出したピークの総数を表す。また、ピーク時刻 $s_i(e)$ に対応するピークの値を $p_i(e)$ とする。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。なお本稿では、チャネル 1 側の車線を右、チャネル 2 側の車線を左と定義する。

2.4 従来手法の問題点

車両の両側に配置した二つのマイクロホンによる観測信号を用いた従来手法では、各録音機器間の同期補償を正確に行うために、複雑な計測手順を踏む必要がある。また、ウィーナーフィルタの設計には、雑音と車両走行音の単一音源区間を手動で探す必要

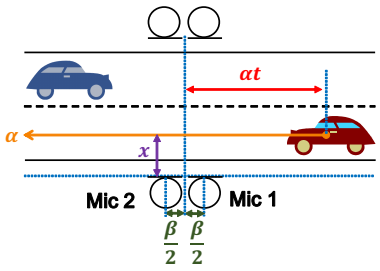

Fig. 2 Configuration of microphone setting.

がある。そこで、本研究では同期補償が不要なステレオ録音に着目し、**2.2** 節で設計したウィーナーフィルタによる雑音抑圧手法を改善して、移動音源モデルを用いて走行音を進行方向毎に強調するマスクを設計することで、雑音の事前学習と同期補償が不要である簡便な車両検出の手法を検討する。

3 提案法：ステレオ情報を用いた移動音源強調

3.1 提案法の概要

本手法では、Fig. 1 のように両側 2 車線道路の両側にステレオマイクを設置し、各車線に近い方の路肩に置いたステレオマイクを用いて車両数を推定する。そのために、各車線の進行方向に移動する走行音のみを強調する時間周波数マスクを構築する。まず、一つの車両がある時刻に想定する方向と速度でステレオマイクの前を通過する場合の時間差推移の曲線モデルを求める。次に、観測信号の時間差を時間周波数スロット毎にマッチングをとる。1 つのスロットに対して、その前後数フレームにおける時間差推移と曲線モデルのマッチングを評価し、モデルとの距離が閾値を上回る成分を抑圧することにより、特定方向に移動する走行音らしい時間差変化をもつ成分のみを取り出す。対象とする方向の走行音のみを取り出すマスク処理が走行方向推定の役割を担っており、取り出された時間周波数成分のみを扱うことで走行方向毎に車両数を計測する。

3.2 移動音源の時間差推移モデルの設計

ステレオマイクに近い方の車線の車両走行音の時間差推移モデルを作成する。車両がマイクの正面を通過した時の時間フレームを $t = 0$ frame とした場合の、車両走行音の波面の到来のマイク 1 に対するマイク 2 の遅延 $M[t]$ s は、Fig. 2 に示すように車道の中央とマイクとの距離を x m、音速を c m/s、車両の移動速度を α m/frame、ステレオマイク内のマイク間の距離を β m とすると、

$$M[t] = \frac{\sqrt{x^2 + \left(\alpha t - \frac{\beta}{2}\right)^2} - \sqrt{x^2 + \left(\alpha t + \frac{\beta}{2}\right)^2}}{c} \tag{6}$$

となる。ただし、車両の移動速度 α はマイク 1 からマイク 2 の方向に進行する場合に正の値を持つものとする。このモデル $M[t]$ の例を Fig. 3 に示す。車両がステレオマイクの正面を通過する N フレームの長さをもつ時間差推移モデル $M[t]$ を作成した。ここで α は車両速度によって変動するため α の値によって

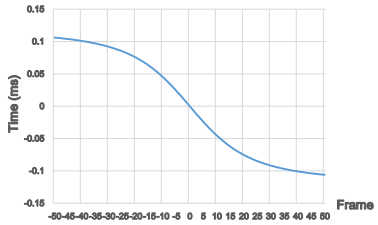

Fig. 3　Moving sound source model.

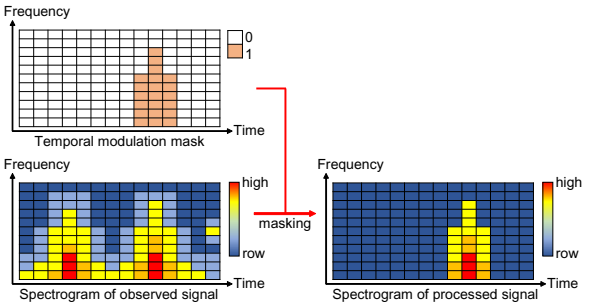

Fig. 4　Target sound emphasis processing by temporal modulation mask.

$M[t]$ は異なる値を持つ。そこで、時速 V km/h の車両速度に対応した α を求め、α の値から複数の時間差推移モデル $M_V[t]$ を作成する。

3.3　時間差推移モデルを用いた非目的方向の走行音抑圧

3.2 節で求めた時間差モデルを用いて、非目的方向の走行音を抑圧する手法について述べる。

まず、ステレオマイクで収録し、フーリエ変換された二つの観測信号から得られる位相 $\theta_1(\omega,t)$、$\theta_2(\omega,t)$ を用いて、二つの観測信号の間の位相差 $\Delta\theta(\omega,t)$ を求める。

$$\Delta\theta(\omega,t) = \theta_1(\omega,t) - \theta_2(\omega,t) \tag{7}$$

ただし検出のための信号分析には素子間隔に対して空間的エイリアシングが起こらない周波数ビンのみを使用する。このような帯域では、到来時間差に対応する位相差は $-\pi$ から π の範囲の値になるため、位相差 $\Delta\phi(\omega,t)$ をこの範囲に補正する。

$$\Delta\phi(\omega,t) = \begin{cases} \Delta\theta(\omega,t)+2\pi & (\Delta\theta(\omega,t) < -\pi\ ) \\ \Delta\theta(\omega,t)-2\pi & (\Delta\theta(\omega,t) > \pi\ ) \end{cases} \tag{8}$$

信号分析を空間的エイリアシングが起こらない低域に制限することは、車両の走行音が 2 kHz 以下の低域のみにパワーを持つため、数センチ程度の素子間隔であれば問題にならない。

次に、求めた位相差 $\Delta\phi(\omega,t)$ から時間差 $\Delta T(\omega,t)$ を次のように求める 。

$$\Delta T(\omega,t) = \frac{\Delta\phi(\omega,t)}{2\pi f} \tag{9}$$

ここで、f は周波数インデックス ω における中心周波数を表す。続いて、モデルとマッチングの評価に用いるための、遅延 $\Delta T(\omega,t)$ の前後の $(N+1)$ フレームの系列とモデル $M_V[i], i=-N/2,\ldots,N/2$ の系列とのユークリッド距離 $R_V(\omega,t)$ を、想定される複数の速度 V のそれぞれについて求める。

$$R_V(\omega,t) = \sqrt{\sum_{i=-\frac{N}{2}}^{\frac{N}{2}} \left(\Delta T(\omega,(t+i)) - M_V[i]\right)^2} \tag{10}$$

得られたユークリッド距離 $R_V(\omega,t)$ を用いて、想定される方向と速度の移動音源のみを通過するバイナリマスクを設計する。ここで各車両の速度と通過時刻はどちらも未知であるため、まず各時刻 t に対象の車線で 1 台の車両が通過すると仮定した場合の、その車両の速度を推定する。各時刻 t において、ユークリッド距離 $R_V(\omega,t)$ が最小になる帯域 ω が最も多くなる想定速度 V を選択することにより、時刻 t に通過していると仮定した車両の速度 V を推定する。次に、時刻 t に速度 V で通過する走行音のモデルによくマッチし、走行音が支配的であると思われる時間周波数成分のみを強調するバイナリマスクを設計する。時刻 t における推定速度 V に対応するユークリッド距離 $R_V(\omega,t)$ を $R(\omega,t)$ とし、閾値 γ を設定して、$R_V(\omega,t) < \gamma$ となる観測信号の時間周波数成分だけを通すバイナリマスクを設計し、観測信号に適用する。バイナリマスクを適用した観測信号に対して **2.2** 節と同様に平滑化処理後、ピーク検出を行う。時間差推移モデルとマッチする周波数成分だけを取り出し、それ以外の周波数成分は除去することにより、対象とする方向の走行音のみの車両数推定が行える。設計した時間周波数マスクによる目的方向の走行音強調の例を Fig. 4 に示す。

Fig. 5　A picture of the recording location.

Table 1　Experimental conditions.

道幅	7.1 m
録音機器間距離	9.5 m
ステレオマイク内のマイク間隔 β	0.04 m
車線の中央とマイクとの距離 x	3.0 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 L	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	18
平滑化フィルタ長 G	6σ+1
時間差推移モデルのフレーム数 N	30 samples
マッチング閾値 γ	0.00033
ピーク検出閾値 h	0.001
提案手法の使用周波数帯域	[0, 4] kHz
録音機器	SANYO ICR-PS603RM
ビデオカメラ	JVC GZ-HM670

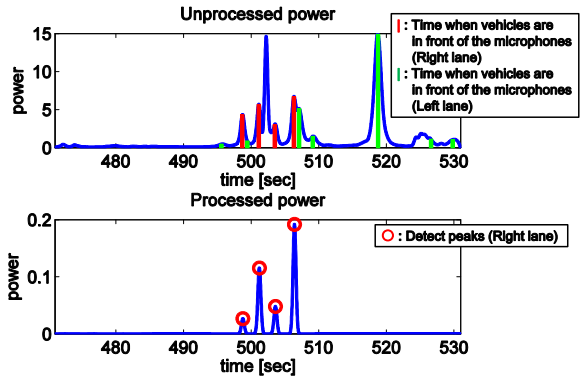

Fig. 6　Experimental result (Right Lane).

4　実環境下での実験

4.1　実験条件

Fig. 5 に示すように、筑波大学近辺の片側 1 車線道路において、対向に設置した二つのステレオマイクを用いて車両の走行方向毎に車両数の推定を行った。録音と同時に車両の走行の様子を動画で撮影し、動画をもとに人手で車両走行方向と通過時刻のラベリングを行った。動画中で車両がマイクロホンの正面を通過した時刻を車両通過の正解時刻とし、それぞれの走行方向毎に検出したピークの時刻と正解時刻との差が 0.5 秒以下の時に正確にピーク検出できたものとした。時間差推移モデルの速度パターン V は実験環境における走行車両の平均的な速度から $V = 30, 35, ..., 60$ の 7 パターンを用いた。ガウス窓型 FIR フィルタの標準偏差パラメタ σ、ピーク検出の閾値 h、時間差推移モデルのフレーム数 N とマッチング評価の閾値 γ は実験的に決定した。その他の実験条件は Table 1 に示す。

4.2　評価尺度

実験結果の正確性の評価のため、以下で表される F 値を尺度として用いた。

$$\text{precision} = \frac{N_c}{N_e} \tag{11}$$

$$\text{recall} = \frac{N_c}{N_r} \tag{12}$$

$$\text{F-measure} = \frac{\text{precision} \cdot \text{recall}}{\frac{1}{2}(\text{precision} + \text{recall})} \tag{13}$$

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。

4.3　実験結果と考察

提案手法により、ある 60 秒の区間において右車線の車両走行音以外の音を抑圧した結果を Fig. 6 に示す。Fig. 6 より、車両数が走行方向毎に正しく推定されていることがわかる。

従来手法および本手法で得られた推定車両数と実際の交通車両数をそれぞれ Table 2、Table 3 に示す。さらに、Table 2、Table 3 の値を用いて算出された F 値を Table 4 に示す。想定の方向と速度で移動する車両音源に起因するピークのみを強調することにより、車両検出精度が向上して F 値の 1 からの差が 31% 減少した。従って、提案手法はステレオマイクの位相差分析を利用することにより、従来手法よりも簡便かつ高精度な車両推定が行えることを確認した。

Table 2　Numbers of vehicles in the ground truth and in the detection in Left Lane (left) and Right Lane (right) by conventional method. [5] (T = True, F = False)

Ground truth (Left Lane)			
		T	F
Detected vehicles	T	36	4
	F	2	-

Ground truth (Right Lane)			
		T	F
Detected vehicles	T	26	0
	F	0	-

Table 3　Numbers of vehicles in the ground truth and in the detection in Left Lane (left) and Right Lane (right) by proposed method.

Ground truth (Left Lane)			
		T	F
Detected vehicles	T	36	0
	F	2	-

Ground truth (Right Lane)			
		T	F
Detected vehicles	T	25	1
	F	1	-

Table 4　Calculated F-measure

	F-measure
Conventional method [5]	0.9538
Proposed method	0.9682

5　おわりに

非同期マイクロホンアレーによる交通量推定の従来手法は非同期マイクロホンの同期補償と雑音の事前学習による観測手順の複雑化が問題であった。そこで本研究ではステレオ録音機器の使用を前提に、同期補償の不要な単一録音機器のチャネル間の処理で想定の移動音源のみを強調する時間周波数マスクを設計することにより、雑音の事前学習が不要であり、手動の計測手順が少ない車両検出の手法を提案した。実験結果から、従来手法よりも高精度かつ簡便に車両検出が行えることが確認できた。

謝辞　本研究は科学研究費補助金基盤研究 (B) (25280069) の助成を受けたものである。

参考文献

[1] 飯田 他, “交通工学,” オーム社, 2008.
[2] Sobreira *et al.*, *Proc. IOA*, pp. 6221–6226, 2008.
[3] 嶋田 他, 信学技報 SIS2009-71, pp. 125–128, 2010.
[4] Liu, *Proc. IWAENC*, 2008.
[5] 豊田 他, 音講論 (秋), pp. 643–646,2014
[6] Markovich-Golan *et al.*, *Proc. IWAENC*, 2012.
[7] Miyabe *et al.*, *Proc. ICASSP*, pp. 674–678, 2013.
[8] Sakanashi *et al.*, *Proc. APSIPA*, 2013.