Event-based Visual Microphone: Non-contact Vibration Sensing using Event Camera*

○丹羽遼吾，伏見龍樹，山本健太（筑波大学），△落合陽一（筑波大学，ピクシーダストテクノロジーズ株式会社）

1　はじめに

量子スケールから宇宙スケールまで，物体の振動はエンジニアリング構造の欠陥や隣室の会話など，様々な情報を符号化することができる．そのため，様々な形状の振動を測定するために，幅広い接触および非接触の測定方法が開発されてきた．

例えば，レーザードップラー振動計（LDV）は光の周波数シフトを利用して振動を測定することができ，非常に高精度で微細な振動まで計測可能である．しかし装置が非常に高価でセットアップに時間を要することが課題である．また，Sheinin ら[1]は計測したい対象にレーザーを照射し，その反射模様を撮影することで高音質に音を記録できる方法を開発した．

しかし LDV と Sheining らの方法は高精度で振動を計測できるが，レーザーを使用するため安全性に課題がある．Davis ら[2]は高速カメラで振動している物体を撮影し，振動を復元する方法を提案した．Davis らの方法はレーザーを使用しないため安全だが，高速カメラで撮影するため非常に明るい照明が必要である．

そこで，私たちは”Event-based Visual Microphone”という低コストかつ非接触での振動測定手法を提案する.イベントカメラは，通常の RGB カメラとは異なり，明るさの変化と時刻のみを記録する特殊なカメラである．高速カメラと異なり撮影速度と解像度の間にトレードオフもなく，撮影時に非常に明るい照明も必要ない．また振動部分には明るさの変化が伴うため，イベントカメラは振動している物体のみを高速（最大 1 MHz）で記録可能である．

2　手法

私たちは Dorn ら[3]が開発したアプローチを用いて，イベントカメラで記録したデータから振動を抽出する．提案システムのセットアップは図 1 の通りである．イベントカメラは Silky EvCam HD を使用し，焦点距離 8mm のレンズ（LM8HC-SW）を使用して撮影を行った．

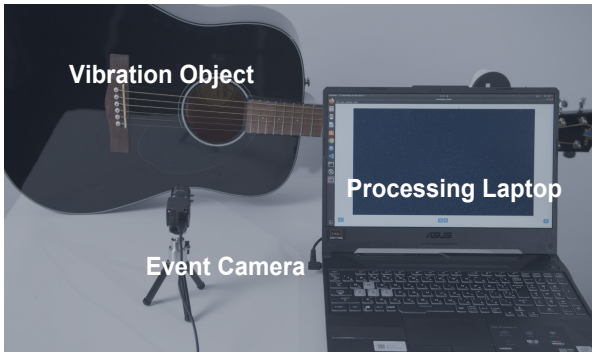

図 1. 提案システムのセットアップ

2.1 イベントカメラ

イベントカメラとは，生物の目を模倣して作成されたカメラであり，物体の明るさ変化のみを記録するカメラである[4]．イベントカメラは物体の明るさが変化した位置と時刻しか記録できないが，最大で 1 MHz の速度で記録できる．また省電力で作動するためロボット分野などで主に注目されているカメラである．イベントは以下の式

$$\Delta \log\{I(x,y,t)\} > \theta_{on}$$

$$\Delta \log\{I(x,y,t)\} < -\theta_{off}$$

で表されるように，あるピクセルにおける明るさ$I(x,y,t)$の対数変化が閾値を超えて増加すると$e(x,y,t)=1$と記録され，閾値を超えて減少すると$e(x,y,t)=-1$として記録される．

音源などの振動している物体のエッジは明るさ変化を伴うため，イベントカメラで記録できる．また，従来手法のハイスピードカメラ

* “Event-based Visual Microphone”, by Ryogo Niwa, Fushimi, Tatsuki, Kenta Yamamoto and Yoichi Ochiai (University of Tsukuba).

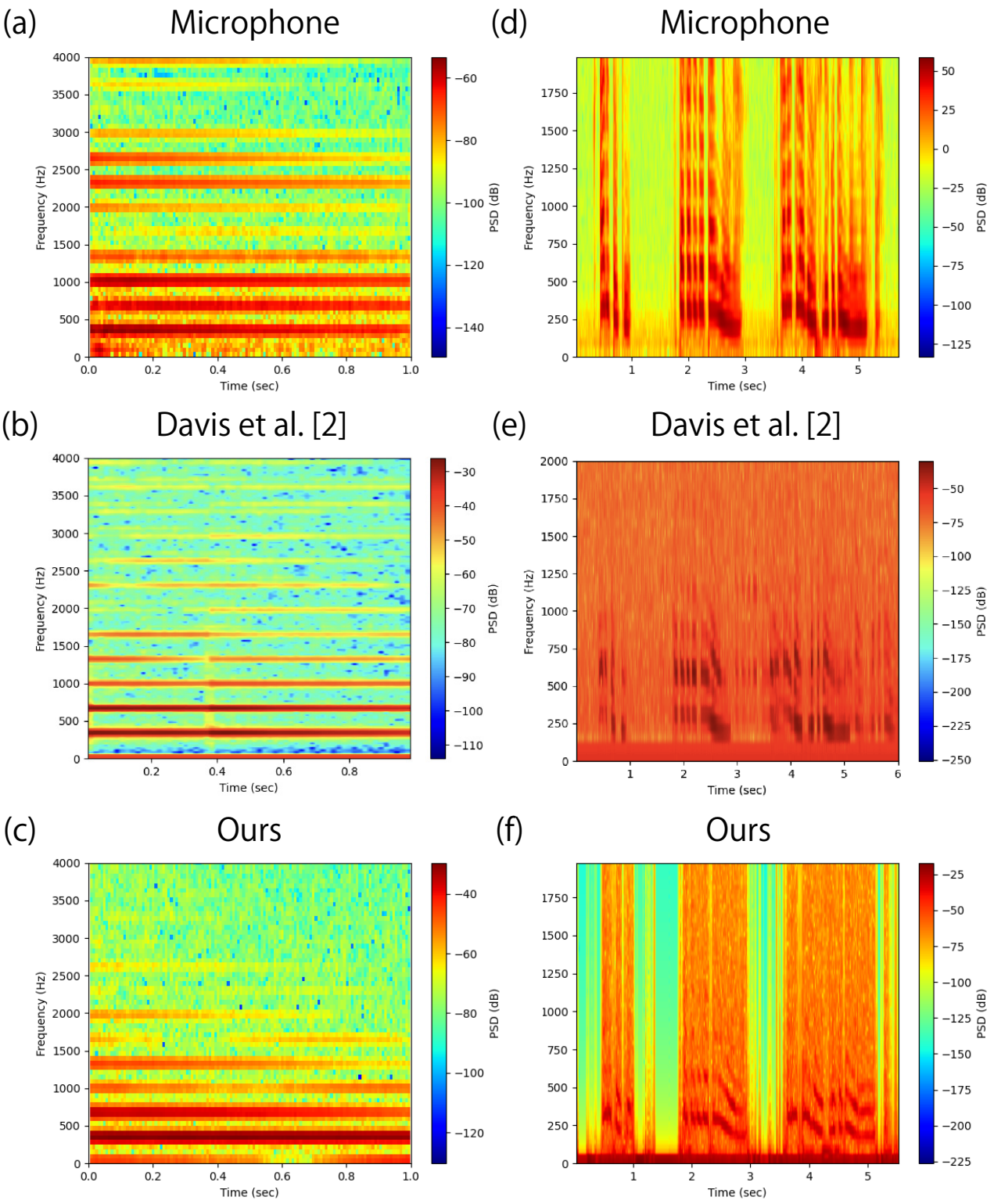

図 2. (a) (d)はそれぞれマイクで収録した音であり，(b) (d)は Davis ら[2]の手法で復元した音 (c) (f)は提案手法で復元した音を表す．(a) (b) (c)はギターの 6 弦を撮影し音を復元した．(d) (e) (f) はスピーカーにくっつけた棒を撮影し人の音声を復元した．

と比べて振動していない背景は記録されないことから，高いデータ効率で記録できる．

2.2 実装

イベントカメラはイベントのみを記録し，フレームという概念を持たない．従来の画像処理技術をそのまま適用することは困難であるため Dorn ら[3]と同様に，擬似フレームを．正のイベントデータからのみ作成する．擬似フレームは以下の式で計算される

$$S(x,y,t) = \sum_{t-\Delta t}^{t} e(x,y,t)$$

なお, 経験的にこのΔtは単一のイベントデータのみを含むほど小さくて良いことがわかっている[3]．この単一のイベントデータのみから作成された擬似フレームを以降$\hat{S}$と呼ぶ．$\hat{S}$は，イベントが発生したピクセルを中心としてn ×nピクセルを表す二次元配列で, 中心の値のみが 1 で，他は 0 の値が入っている．
イベントはその位置において物体のエッジやパターンが振動していることを示すため，擬似フレームの局所的な位相の時間変化は，エッジやパターンの動きを表す．そのため局所位相を算出するために，まず局所反応

$$R = \hat{S} * G_{\omega_0}$$

を$\hat{S}$に対する単一スケールの複素ガボールフィルタを用いて計算する．複素ガボールフィルタはパラメーターとして振動の方向θを入力する．この計算を行うことで，局所位相の値が計算され，イベント発生近傍のθ方向のピクセルに値が割り当てられる．
ノイズの影響を減らすために，局所反応Rに対して振幅重み付きのガウスブラーを計算する．Aは局所振幅$A = |R|$，ϕは局所位相$\phi = \arg(R)$である．K_ρはガウスカーネルであり，次の式で表される．

$$K_\rho = e^{\left(-\frac{x^2+y^2}{\rho^2}\right)}$$

この空間的にぼかされた位相信号$\hat{\phi}$は以下のように表現できます．

$$\hat{\phi} = \frac{A\phi * K_\rho}{A * K_\rho}$$

次に，各ピクセルの位相信号の値を時系列順に並べ替え，線形補間を行うことで，すべてのピクセルの振動を復元できる．

2.3 ノイズ処理

より高精度で振動を再構築したい場合，主成分分析を用いることができます．まず，記録した時刻においてイベントデータ数が最も高いNピクセルから振動データを抽出し，行列$\delta \in R^{N\times T}$を作成します．ここでTは復元データのサンプル数を表します．私たちはδに対して特異値分解を行い

$$\delta = U\Sigma V^*$$

$\eta = U_r * \delta$を計算して振動データを数個の主要な成分に線形射影する．

$$\eta = U_r^*\delta$$

ここで，U_rはUの最初のr列からなる行列，つまり$[u_1, \dots, u_r] \in R^{N\times r}$となる．$\eta = [\eta_1, \dots, \eta_r] \in R^{r\times T}$のi行目は，$\delta$のi番目の主成分に対応し，より精密な信号を提供する．

3 応用

イベントカメラは，オブジェクトが振動するときの輪郭またはパターンだけを記録します．したがって，物体が音源である場合，その物体または付属する物体が発生させる音を再構築することができる．例えば，ギターの弦や

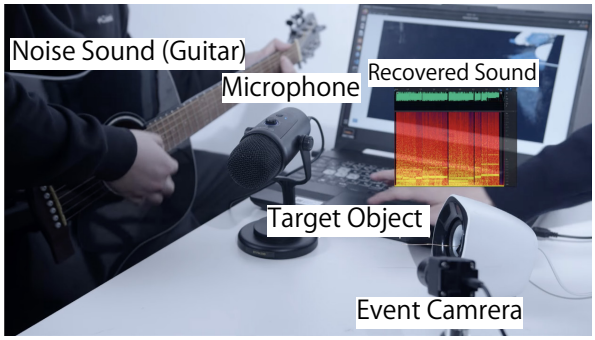

図 3. 音源分離タスク時のセットアップ

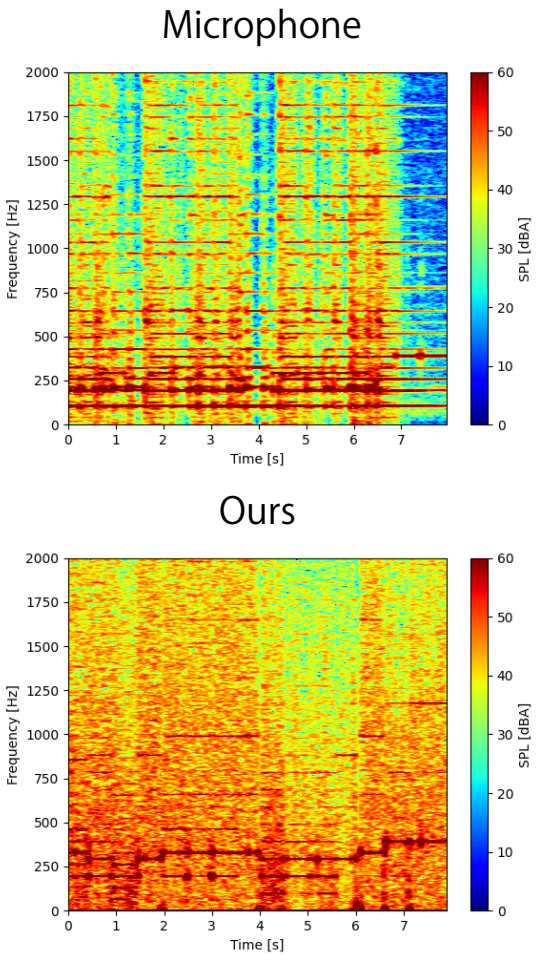

図 4. （上図）マイクで記録した音
（下図）提案手法で記録した音

スピーカーに取り付けられた棒を撮影することで音を復元可能となる．

3.1 音源分離

提案手法では，雑音の多い環境でも記録したい音だけを復元することが可能である．図３のようにギターとスピーカーの２つの音源を用意した．提案手法ではスピーカーにくっつけた棒を撮影しスピーカーの音のみを復元できることを示した（図４参照）．

3.2 遠方の音源

望遠レンズを使用すると，数十メートル離れた音源も撮影でき，音を復元できる．私たちは 10m 程度遠方のギターの第３弦を TAMRON AF180mm F3.5 Di レンズを使用して撮影し音を復元できることを示した．図 5 は，マイクで録音された音と提案手法で再構築された音との比較である．

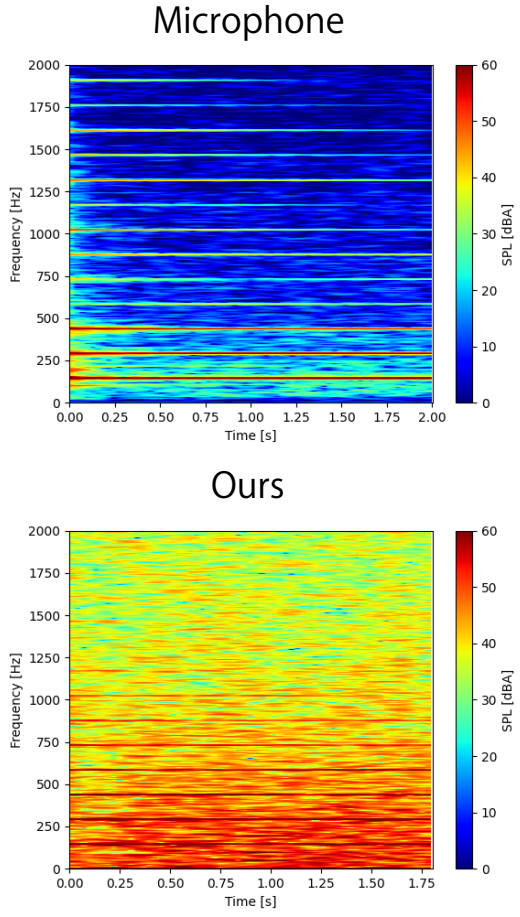

図 5 （上図）遠方にあるギターを通常のマイクで記録した音
（下図）提案手法で記録した音

4 議論

提案手法で記録した音は，図 2 (b) (c) (e) (f)からわかるように，1500 Hz 以上の高周波数のパワーがマイクと比べて小さいことがわかる．これは Davis ら[2]の方法も提案手法も，カメラを用いて振動を記録しているため，振幅があまりに小さい振動はカメラに記録されないことが原因であり，多くの場合高周波数の振動は振幅が小さいためである．そのため将来的には，機械学習モデルを用いて高周波数の振動を補完するなどのアプローチを行うことで, より高音質な記録が可能となると考える．音声に対しては Super speech resolution[5]が有効であると考える．

5 おわりに

非接触測定装置はレーザードップラービブロメータ（LDV）や高速カメラなどが，様々な振動を測定するために開発されてきた．LDV は高価であり，高速カメラは解像度とサンプリング周波数の間で大きなトレードオフがある．そこで，私たちは明るさの変化だけを記録するイベントカメラを用いて振動を測定する非接触でシンプルかつ低価格な手法を提案した．

この研究では，イベントカメラを使用して聞こえる音の再構築を実証した. この研究は，非常に明るい照明を必要とせず，コスト効率の良い方法で人間の声などの可聴音を再構築できる振動を測定し，再構築する課題に取り組んだ．この方法は，科学者やエンジニアが振動や音を測定し再構築するためのコスト効率の良い非接触手法を開発するための新たな道を提供する．

謝辞

本研究は CREST AIP チャレンジを含む JST CREST(JPMJCR19F2)の支援のもと実施されました．

参考文献

[1] Mark sheinin et al., IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16303-16312, 2022

[2] Abe Davis et al., ACM Transactions on Graphics, 33(4), 79:1-79:10, 2014

[3] Charles Dorn et al., Journal of Engineering Mechanics , 144(7), 2018

[4] M. Mahowald and C. Mead, Scientific American, 264(5), 76-83, 1991

[5] Haoche Liu et al., Interspeech, 4227-4231, 2022