

ローリングシャッター効果を用いた 音声情報復元とその評価

星野 翔^{1,a)} 嶋野 裕一郎¹ 崎山 一男¹

概要: 通常のカメラに組み込まれたイメージセンサのフレームレートよりも高い時間周波数の光信号を読み取るべく、ローリングシャッター効果を用いた研究が提案されている。しかし、取得可能な光信号の周波数限界を評価する研究はみられない。そこで本論文では、カメラセンサのローリングシャッター効果を用いて、取得可能な光信号の時間周波数限界を調査する。具体的には、ローリングシャッターを有するスマートフォンを用いて光に重畳した音声データを復元した。その結果、約 2000Hz 程度までの信号を取得できることが確認できた。

キーワード: ローリングシャッター, 可視光通信, 音声復元

Speech Information Recovery Using a Rolling Shutter Effect and Its Evaluation

KAKERU HOSHINO^{1,a)} YUICHIRO SHIMANO¹ KAZUO SAKIYAMA¹

Abstract: Research using the rolling shutter effect has been proposed in order to read an optical signal with a time frequency higher than the frame rate of the image sensor embedded in a standard camera. However, no studies have been conducted to evaluate the frequency limit of the acquirable optical signal. Therefore, in this paper, we investigate the time-frequency limit of the optical signal that can be acquired by using the rolling shutter effect of the camera sensor. Specifically, we try to recover the voice data superimposed on the light using a smartphone with a rolling shutter. As a result, we confirmed that the acquirable signals could be up to about 2000 Hz.

Keywords: Rolling shutter, Visible Light Communication, Audio reconstitutes

1. はじめに

1.1 研究の背景

現在、交通信号機や標識などの光源に、高寿命かつ高効率であるという利点から LED デバイスが広く利用されている。このような LED デバイスを可視光通信に利用することで、地下などの電波の届きにくい場所での通信が簡単に行えるようになる。文献 [1]-[3] ではインフラとしての LED を可視光通信に利用する手法について検討している。

通常、可視光通信では受信デバイスとして高感度の PIN

フォトダイオードが用いられるが、カメラを用いて可視光通信を行う方法も存在する [1]。カメラを用いた可視光通信において大きな問題となるのは、通信のデータレートがカメラのフレームレートによって制限されてしまう点である。そこで、イメージセンサーのローリングシャッター効果を用いてデータレートを高める研究が行われている [2]。ローリングシャッター効果は様々な用途で使用されている。文献 [4] では車両用 LED テールライトを送信機、ローリングシャッターセンサー搭載カメラを受信機とした可視光通信システムが提案されている。このシステムは、可視光通信における問題点である、光干渉による画像内のノイズに耐性をもつ。文献 [5] では様々な周波数のパルス波を LED

¹ 電気通信大学大学院
The University of Electro-Communications
^{a)} hoshino@uec.ac.jp



図 1 シャッターの違いによるカメラ撮像イメージ
Fig. 1 Image of difference by shutter

デバイスを入力し、iPhone11 で撮影する実験から、ローリングシャッター効果を用いて LED デバイスの物理指紋を抽出可能か検証している。

文献 [6] では本論文と同様に、カメラを用いて音声データの抽出を行なっている。この文献では、閉鎖された空間に音源と振動が伝わりやすい物体を設置し、ハイスピードカメラを用いて物体の変位を観測することで音声の復元に成功している。本論文では、ローリングシャッターを有するスマートフォンを用いて、光に重畳した音声データを復元する。

1.2 研究の目的

先行研究 [5] では、ローリングシャッター効果を用いて測定できる周波数限界を調査した。本研究では、ローリングシャッター効果を用いて取得可能な光信号の時間周波数限界を調査する。具体的には、ローリングシャッターを有するスマートフォンを用いて光に重畳した音声データを復元した。また、復元した音声を定量的に評価することで、ローリングシャッター効果を用いて得られる信号の精度を評価する。

2. 関連技術

2.1 ローリングシャッター

カメラに使用されているシャッターとして、一般的に CCD イメージセンサーに搭載されているグローバルシャッターと CMOS イメージセンサーに搭載されているローリングシャッターの 2 種類が採用されている。具体的な違いとしては、前者では全ての画素に対して同じ時刻に露光を行うことでシャッターを切るのに対して、後者では端から順次スキャンライン毎に露光を行うことでシャッターを切る。したがって、ローリングシャッターでの撮影はスキャンライン毎に異なる時刻を撮影していることになるため、1つの画像内で複数の時刻が存在する。

図 1 にグローバルシャッターとローリングシャッターの両方式で左から右に動く車を撮影した場合のイメージ図を示す。グローバルシャッターでは全ての画素が同時に露光されるため左図のような画像になる。他方、ローリングシャッターでは図 1 の水平方向スライスのように、スキャンライン毎に露光が行われるためスキャンラインと垂直な方向に時間差が生じた右図のような画像になる。このように、肉眼で見る状態と異なる画像を撮影してしまう現象を

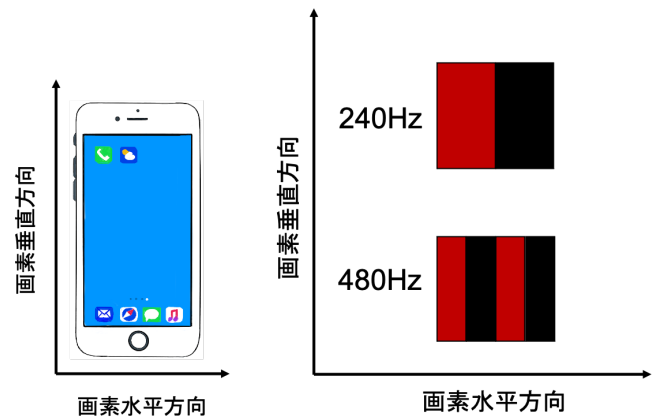


図 2 ローリングシャッターによる矩形波を重畳した光撮影時のイメージ

Fig. 2 Image when shooting a light source with a rolling shutter

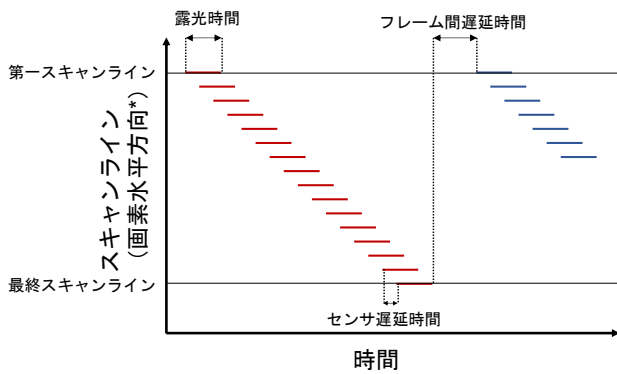
ローリングシャッター効果、正式にはフォーカルプレーン歪みという [7].

肉眼で蛍光灯や LED 照明などの光のちらつきが視認できることがある。これは、本来であれば 1 秒間に 100~120 回点滅しているため肉眼では確認できない点滅が、経年劣化などの要因によって点滅回数が肉眼で視認できる 1 秒間に 70 回以下程度まで減少することによって起こる。カメラによる撮影では肉眼では視認が難しい高速の点滅でも撮影でき、これはフリッカー現象と呼ばれる。つまり、ローリングシャッター効果は光が高速に点滅している場合にも起こり、光の点滅は画像内に明暗の線が出るという形で現れる。

ローリングシャッターセンサ搭載カメラの FPS (Frame Per Second) が 240 FPS であった場合に、光の点滅の回数を変化させることによって撮影されるイメージ図を図 2 に示す。これ以降左図のスマートフォンでの縦方向を画素垂直方向、横方向を画素水平方向と定義する。図 2 の右図上段は撮影対象の光が 240 Hz の信号の場合、右図下段は 480 Hz の場合を示している。信号が矩形波であり、信号の位相とカメラの撮影タイミングが一致している場合、図のような半分が明、半分が暗の画像が 1 秒間に 240 枚撮影される。信号が矩形波でない場合、この明暗の明るさや明と暗の切り替わりの間隔であることがわかるような変化が起きる。仮に解像度 1920×1080 で撮影を行い、1920 ラインの明暗が明確に撮影出来る場合、最大で $\frac{1920}{2} \cdot 240 = 230,400$ Hz の信号が読み取れることになる。

2.2 可視光通信

光を用いた無線通信の中で、可視光 LED を用いた通信を可視光通信と呼ぶ。可視光通信の特徴として、LED を任意の回数と間隔で点滅させることができ、高寿命、高効率、調光可能であるということが知られている。先行研究では、ローリングシャッター効果を可視光通信に応用する



*カメラによってスキャンラインの方向は異なる

図 3 ローリングシャッターにおける遅延時間
Fig. 3 Delay time on rolling shutter

研究が行われている [1],[4],[9]. 既存の可視光通信のアプリケーションのビットレートとしてはスマートフォンやタブレットに合わせた 8 bps 程度である. 一方で, ローリングシャッター効果を用いた可視光通信を行うことによってより高いビットレートでの通信を行う研究が進められている [1]. また, 可視光通信での環境光による影響を軽減するためにローリングシャッター効果を用いる研究も行われている [9].

3. 実験

3.1 実験の目的

以前行った実験 [5] では, スマートフォンでの光信号撮影によってローリングシャッターセンサ搭載カメラが測定可能な周波数限界を調査した. この時に用いた信号は一定の周波数のパルス波であり, 徐々に信号の周波数を高くして, その度に撮影をした. 今回の実験では音声データを光に重畳することによって, 一定の間隔の点滅ではなく, 変化し続ける光信号を読み取った場合のローリングシャッターカメラの時間周波数限界を調査する.

3.2 実験器具

本実験では音声データをファンクションジェネレータ (GWINSTEK AFG-3032) を用いて平行照明型 LED (Pi PHOTONICS, INC. HL4249) に入力し, 投光を行った. また, 撮影はスマートフォン (Redmi Note 10 Pro) で行った. カメラのフレームレートは 60 FPS であり, 画素数は 1080×1920 pixels, 露光時間 (シャッタースピード) は $\frac{1}{4000}$ s である. 使用したスマートフォンは図 3 のように画素垂直方向にスキャンラインが並んでおり, カメラセンサの左端から順にローリングシャッター方式で露光を行っている.

3.3 手法

3.3.1 ローリングシャッターセンサの遅延時間測定

本実験で測定, 計算を行った項目を図 3 に示す. 本実験で指すローリングシャッターセンサの遅延時間とは図 3

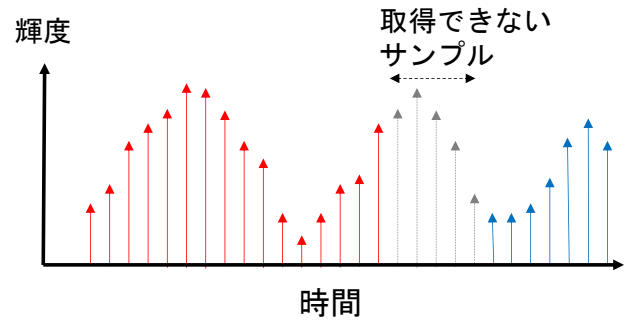


図 4 音声データと輝度データの変換イメージ
Fig. 4 Image when obtaining luminance

のセンサ遅延時間のことである. 図 3 からわかるように, ローリングシャッター方式ではフレーム毎の最終スキャンラインの露光が終わった後に処理時間やリセット時間に伴うフレーム間遅延時間が存在する. これによって図 4 の灰色部分のように取得できないサンプルが発生してしまう. フレーム間遅延時間を直接測定することは難しい. そのため, センサ遅延時間を実験により測定し, 1 フレーム全体としてかかる時間から設定した露光時間と測定したセンサ遅延時間の和を引くことにより求める. この時, 1 フレーム全体に対するフレーム間遅延時間の割合を計算すれば, 1 フレーム全体に対する取得できないサンプルの割合を導出できる.

3.3.2 音声復元

音声の復元は, 以下の手順で行った.

- (1) ファンクションジェネレータから平行照明型 LED に音声信号を入力
- (2) 平行照明型 LED によって投光された光をスマートフォンで動画撮影
- (3) 撮影された動画をフレーム毎の画像に変換
- (4) 画像から露光列と垂直な方向に等間隔で RGB 値を取得し, それぞれで輝度を計算
- (5) 輝度を元に音声ファイルを作成

音声復元時のプログラムでは, フレーム間遅延時間による取得できないサンプルを考慮し, 欠落したサンプルの部分に 0 を埋めることで元データのサンプル数と同数になるように調節した.

3.4 結果・評価

ローリングシャッターセンサの遅延時間を測定するために撮影した画像を図 5 に示す. ただし図 5 は輝線が確認しやすい様に色彩等に加工が施してある. 60 FPS で 600 Hz パルス波の点滅を撮影した. 理想的な撮影ができた場合図 2 の通り明暗 10 本ずつの輝線が撮影できる. しかし, 図 5 からは明線が 8 本しか確認できず, 理想的な結果は得られなかった.

音声復元に用いた輝度取得画像のうち 1 枚を図 6 に示す. 動画 1 秒につき図 6 のような画像を 60 枚取得できる.

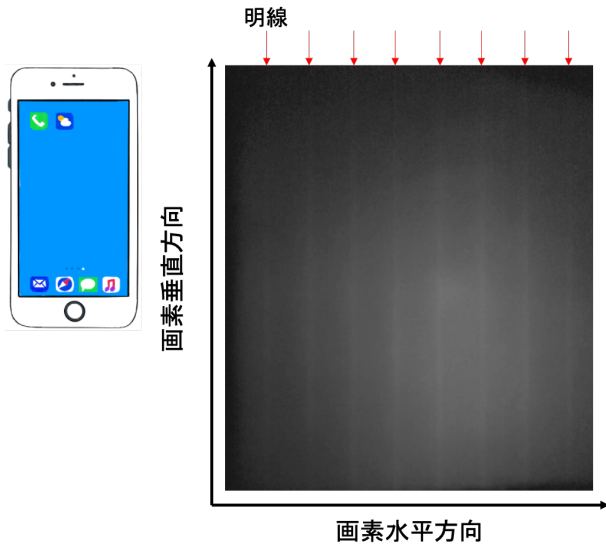


図 5 スマートフォン (Redmi Note 10 Pro) による 600 Hz パルス波の撮影

Fig. 5 600hz pulse wave shooting by Redmi Note 10 Pro

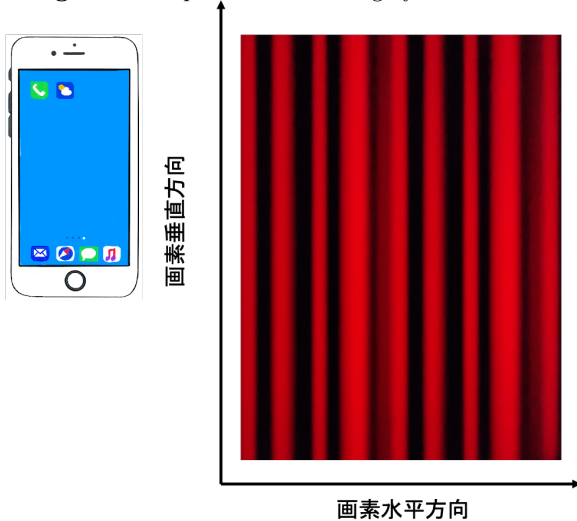


図 6 スマートフォン (Redmi Note 10 Pro) による輝度取得画像

Fig. 6 Luminance obtained image by Redmi Note 10 Pro

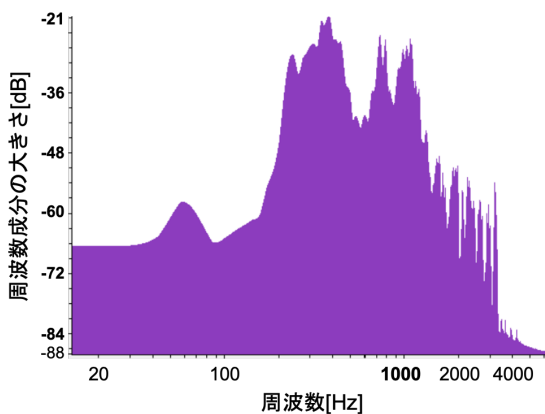


図 7 元の音声データの周波数スペクトル

Fig. 7 Frequency spectrum of the original audio data

それぞれの画像 1 枚ずつからスキャンラインに垂直な方向で輝度を取得し、音声データとした。元データには人の声

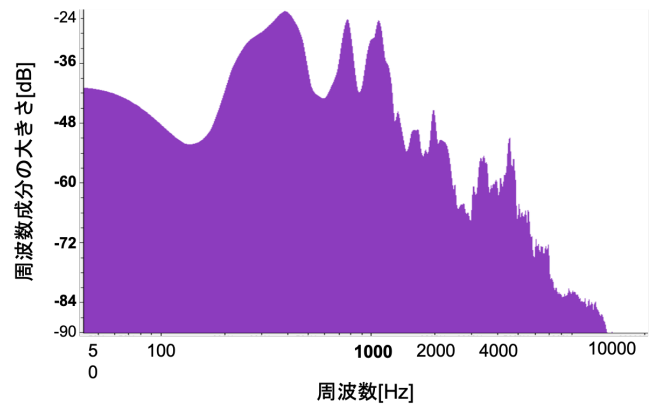


図 8 復元した音声データの周波数スペクトル

Fig. 8 Frequency spectrum of the recovered audio data

表 1 Redmi Note 10 Pro における測定・計算結果

Table 1 Measurement and calculation results on Redmi Note 10 Pro

	Redmi Note 10 Pro
輝線の間隔	124 pixel
ローリングシャッターセンサの遅延時間	$1.344 \cdot 10^{-5}$ s
1 フレームで露光を行なっている時間	$1.477 \cdot 10^{-2}$ s
欠落するサンプルの割合	11%
明瞭度 [8]	0.6458

を含む音声を用いたが、復元したデータではその言葉の発音が明確に確認できた。

音声信号では、ヒトが実際に音声を聞いた場合の主観的評価と SNR (Signal to Noise Ratio) による客観的評価との差が生じやすい。例えば、極端に SNR が低い場合に、ヒトはその発音を聞き取れることがある。したがって、SNR による評価ではなく Taal による知覚に基づく明瞭度測定 [8] を行った。

- (1) 元の音声データと取得した音声データをそれぞれ離散フーリエ変換 (DFT) によって 1/3 オクターブ毎の周波数に分解
- (2) 周波数毎に分解したデータをさらに一定時間毎のデータに分解
- (3) 正規化及びクリッピングによって雑音や飛び値を少なくする
- (4) 分解されたデータそれぞれに対し、元データと復元データの相関係数を計算
- (5) 得られた相関係数の平均値を導出

最終的に得られた相関係数の平均値が Taal らによる明瞭度の結果となる。つまり、明瞭度の値は $-1 \sim 1$ の範囲をとる。

評価を行うにあたって、元データと取得したデータとの比較を行うが、その開始時間を同期させるためのトリガーを作成することが難しかった。したがって、相互相関関数を用いて開始時間の同期を行った。これによって 1 ms 以下の精度で開始時間を同期し、評価を行った。結果を表 1

に示す。

この結果に加えて、視覚的な評価をするため元データと復元データの周波数分布を取得した。図 7 が元データの周波数分布、図 8 が復元したデータの周波数分布である。

3.5 考察

ローリングシャッター遅延時間の測定のために撮影した図 5 では、明線が 8 本、暗線が 9 本確認できる。理論上合計 20 本の輝線が確認できるはずなので、 $\frac{17}{20}$ のデータが取得可能だと予測できる。結果として、撮影した動画から取得できたサンプルは 89 % であり、画像 1 枚からローリングシャッターの性能を予測できることがわかった。

周波数分布を比較すると、2000 Hz 程度までは似たような周波数にピークがあることが読み取れる。しかし、3000 Hz から 4000 Hz 程度以降はほとんどデータが取得できていない。データが取得できなかった理由はカメラのシャッタースピードが $\frac{1}{4000}$ s であることが原因である考えられる。つまり、取得したいデータの周波数よりもシャッタースピードの周波数の方が高い必要があると予測できる。現在の一般的なカメラにおけるシャッタースピードの限界は $\frac{1}{30000}$ s 程度であることから、最大 15000 Hz 程度の周波数の信号であれば本実験と同様に取得できるという仮説が立てられる。

本論文で行ったローリングシャッターセンサの遅延時間の測定を iPhone11 でも同様に行った。その結果は付録に添付する。スマートフォン (Redmi Note 10 Pro) での撮影では欠落するサンプルが 11 % であったのに対し、iPhone11 では 73 % となった。2.1 章にて記載した通り、ローリングシャッター効果は被写体を撮影する時に 1 つの画像内で時間差が生じることにより、本来取得したくない画像が取得される。そのため、ローリングシャッターセンサの遅延時間を短くするために研究が進められ、ローリングシャッター効果を起こりにくくする工夫が製品に適用されている。しかし、2 種類のスマートフォンでの実験結果から分かるようにローリングシャッター効果による影響が大きい方が、異なる時間の情報が取得可能と言える。また、明瞭度による評価からもローリングシャッター効果による影響が大きい方が、復元した信号が元の信号に近いことがわかる。

4. まとめと今後の展望

本論文では、ローリングシャッター効果を用いて観測できる信号の精度を調査した。そのために、音声信号を重畳した光をカメラで撮影し、音声を復元する実験を行った。また、その結果に対して 2 つの方法で評価を行った。Taal らによる明瞭度測定から元データと復元したデータに正の相関があることがわかり、重畳された音声データが復元できていることが確認できた。また、周波数分布を比較することで 2000 Hz 以下であればピークがほとんど一致する

表 A-1 iPhone11 における測定・計算結果

Table A-1 Measurement and calculation results on iPhone11

	iPhone11
輝線の間隔	450 pixel
ローリングシャッターセンサの遅延時間	3.700×10^{-6} s
1 フレームで露光を行なっている時間	4.476×10^{-3} s
欠落するサンプルの割合	73%
明瞭度 [8]	0.4350

ことが確認できた。しかし、この結果は実際に信号を送信する手段とする場合の精度がわかりにくい。したがって今後、スマートフォンやタブレットなどの音声認識機能を用いて復元した音声信号の精度を調査する。加えて、シャッタースピードを現在の $\frac{1}{4000}$ よりも短い時間に設定し、取得できる信号の限界周波数を改めて測定する。

謝辞 本研究は JSPS 科研費 JP18H05289 の助成を受けたものです。

参考文献

- [1] 中川正雄. “ユビキタス可視光通信,” 電子情報通信学会論文誌 B 88.2, pp.351-359, 2005.
- [2] 中川正雄. “可視光通信,” 映像情報メディア学会誌, Vol.60 No12, pp.1908-1913, 2006.
- [3] 鈴木修司. “可視光通信の現状と展望-ユビキタスで安全な ICT インフラを目指して,” 情報処理 50.5, pp.418-425, 2009.
- [4] Ji, Peng, Tsai Hshin-Mu, Wang Chao, Liu Fuqiang, “Vehicular visible light communications with LED taillight and rolling shutter camera,” 2014 IEEE 79th Vehicular Technology Conference (VTC Spring). IEEE, 2014.
- [5] 星野翔, 崎山一男, “ローリングシャッター効果を用いた LED デバイスの物理指紋抽出に関する基礎的実験.” 暗号と情報セキュリティシンポジウム, 2021.
- [6] Davis, A., Rubinstein, M., Wadhwa, N., Mysore, G. J., Durand, F., Freeman, W. T. The visual microphone: Passive recovery of sound from video. 2014.
- [7] 蚊野浩. “デジタルカメラのしくみと画像処理,” 画像電子学会誌 41.3, pp.288-295, 2012.
- [8] Taal, C. H., Hendriks, R. C., Heusdens, R., Jensen, J. “An algorithm for intelligibility prediction of time-frequency weighted noisy speech.” IEEE Transactions on Audio, Speech, and Language Processing 19.7 (2011)2125-2136.
- [9] 堺井 大地, 川上 朋也, 柴田 直樹, 伊藤 実, “ローリングシャッター効果を用いた調光可能な可視光通信システムの設計と実装,” 第 25 回マルチメディア通信と分散処理ワークショップ論文集 2017, pp.165-168, 2017.

付 録

A.1 iPhone11 を用いた実験

本論文では、スマートフォン (Redmi Note 10 Pro) での実験を記述した。端末毎のローリングシャッターの遅延時間の違いによって取得できる情報に差があることがわかっており、iPhone11 で結果 600Hz のパルス波を撮影した

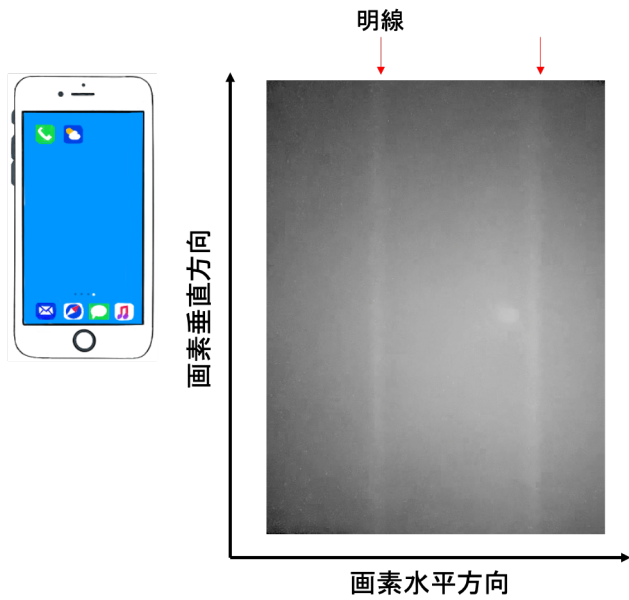


図 A.1 iPhone11 による
600 Hz パルス波の撮影

Fig. A.1 600 Hz pulse wave shooting by iPhone11

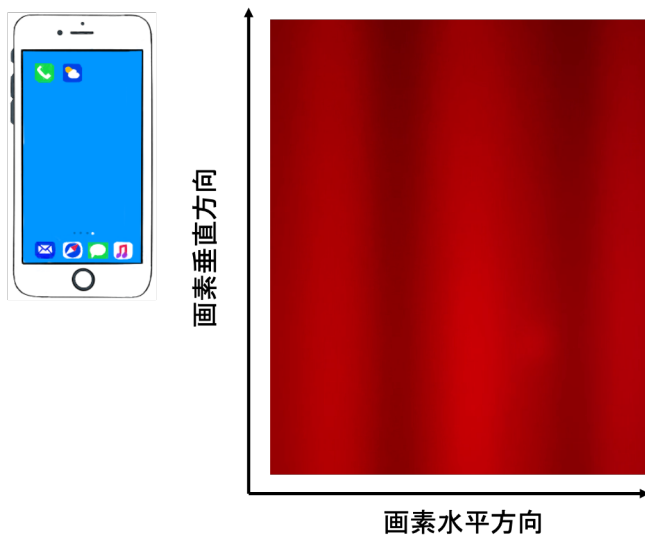


図 A.2 iPhone11 による
輝度取得画像

Fig. A.2 Luminance obtained image by iPhone11

結果を図 A.1 に示す．この時の設定はフレームレートはを 60 FPS, 画素数を 1080×1920 pixels, 露光時間（シャッタースピード）を $\frac{1}{2100}$ s とした．測定結果を表 A.1 に示す．また，iPhone11 を用いた音声復元実験での撮影結果を図 A.2 に示す．図 6 と比べると画像内での輝度の差が明確でないことが読み取れる．