

# ポイ捨て対策向け画像認識と音響発信を用いた注意喚起システム の実現可能性調査

山本浩登<sup>1</sup> 山田隆亮<sup>1</sup>

**概要:** 監視カメラの映像からポイ捨てを検知すると、容疑者に向けて指向性スピーカから音を流し、アクティブに不法行為を抑制する注意喚起システムを提案する。リアルタイム検知を行う基本機能を試作して実現可能性調査を行う。音と画像を処理する性能評価結果に基づいて運用設計と開発課題を明確化した。

**キーワード:** SDGs, ポイ捨て, 指向性スピーカ, YOLO, 実現可能性調査, 注意喚起システム

## Feasibility Study of Reminder System Using Video Recognition and Sound Transmission to Help Prevent Littering

HIROTO YAMAMOTO<sup>†1</sup> TAKAAKI YAMADA<sup>†1</sup>

**Abstract:** Reminder system using video recognition and sound transmission to help prevent littering as active countermeasure is proposed. A prototype system with real-time object-detection is developed for its feasibility study. By setting a use case of the prototype, operation design and technical problems are clarified based on the results of sound tests and video tests.

**Keywords:** SDGs, littering, directional speaker, YOLO, feasibility study, reminder system

### 1. はじめに

経済と社会と環境にまたがって世界規模で深刻な問題に対して全世界が協力して行動していくために、17の持続可能な開発目標 (SDGs: Sustainable Development Goals) が国連で採択されている。SDGs とは、2030年までに持続可能でよりよい世界を目指す国際目標であり、17のゴール・169のターゲットから構成され、地球上の誰一人取り残さないことを誓っている [1]。更に SDGs は発展途上国のみならず、先進国自身が取り組む普遍的なものであり、日本でも積極的に取り組みが行われている。

廃棄物対策は教育、衛生、消費、資源などの目標の複数に関わる複雑な問題である [2, 3]。特に個人での不法投棄 (ポイ捨て) は身近な問題である。従来は、地方自治体の清掃事業、近隣住民の自助努力と相互監視、学校教育、ボランティア活動などの社会システムが機能して、一部の不心得者のポイ捨ては社会的にさほど目立たなかった。昨今のモノ余りの時代において、路上飲みによる置き捨てや、道端や公共の場所に対するポイ捨て事例 [4] が後を絶たず、旧来の社会システムでは対処しきれなくなっている。

文化、風習などの多様な個性を認め合うダイバーシティが進み、衛生等での共通の価値観を共有してきた地域コミュニティは衰退の方向にある。地域コミュニティの衰退によって引き起こされることとして、行政からの情報伝達など連絡調整機能の低下、犯罪や事故に対する地域防犯機能

の低下などが挙げられる。これにより、ルールを守らずゴミの分別を行わない人や不法投棄者の増加が起き、住民だけの清掃作業がより難しいものとなっていく。

ゴミの不法投棄への対策として、看板等を用いて倫理を諭すなどのパッシブな対策と不法投棄を見つけた際に声掛けを行うなどのアクティブな対策がある。パッシブな対策には費用対効果のよい注意喚起効果がある一方で、犯人を特定しにくい点や、看板等を不法投棄者が見つけることが出来ないと効果を発揮しない点などが挙げられる [15, 16]。逮捕や拘束までの強制手段をもたないアクティブ対策では、呼びかけの発話等を通して行動抑制を促す自動制御技術の導入が考えられる。例えば、指向性スピーカを用いて歩きスマホを検知すると歩行者に注意を促す先行研究がある [17]。音響発信するシステムを街頭に常備するには、騒音対策 [9] での周辺住民の理解協力が欠かせない。

監視カメラの映像からポイ捨てを検知すると、容疑者に向けて指向性スピーカから音を流し、アクティブに不法行為を抑制する注意喚起システムを提案し、実現可能性調査のひとつとして、利用可能な要素技術の評価を行ってシステム運用方式に検討を加えたので報告する。

### 2. 関連する研究

#### (1) ゴミを捨てる心理

文献 [5] によると、ゴミを捨てる行動には「集め捨て」と

<sup>1</sup> 大阪工業大学 情報科学部 情報システム学科  
Osaka Institute of Technology

「散らし捨て」がある。前者は回収を目的として設置された容器へ人がゴミを自発的に投入することであり、「散らし捨て」は任意の場所にゴミを投棄ないし放置することをいう[5]。「散らし捨て」の中には、置いたまま放置する「置き捨て」とゴミを見えないように物陰に捨てられる「隠し捨て」に分類している。公園や観光地など人が集まる場所での、ゴミが捨てられる場所を観察すると散らし捨てた際にゴミが目立ちにくい場所に捨てられていることが多い。さらに、散らし捨てが既にされていると、散らし捨てをする抵抗は小さくなり「ゴミがゴミを呼ぶ」などの心理学上の研究が進められている。

## (2) ロボット警察

監視カメラの映像は捜査の際に重要な証拠となることも多いものの、プライバシー問題や設置場所など様々な問題が発生する。地域の治安維持を行うために警察官の代わりとして町をパトロールし、違法行為や不適切行為を見つけると取り締まりや警告を行うロボット警察の研究が進められている。ロボット犬が公営住宅での逮捕劇に試験的に導入された先行研究は、遠隔操作によって足場の悪い場所を移動したり障害物をよけたりしながら周囲を撮影し、情報を取得できるが、現在は利用停止している [5]。文献[7]によると、遠隔操作するロボットに警察業務の一部を仲介させる研究が進められていて、音声やそのボリュームが重要な役割を果たすとの報告がなされている

## (3) YOLO

不正行為を自動認識する要素技術の一例として、YOLO (You Only Look Once) は、画像全体を CNN (Convolutional Neural Network) に与えて複数のオブジェクトをリアルタイム検出する[8]。一般的な物体検出方法では物体の境界設定と物体検出を2段階にして行うことと対比して、YOLOはこれらを同時に行うことにより高速な物体検出を行う。具体的には、7×7のグリッドで分割した領域ごとに物体矩形と物体カテゴリを出力する。入力画像を畳み込みおよびプーリング処理を通し特徴マップを生成する。得られた特徴マップは領域特徴となる構造となり、この特徴マップを全結合層に入力する。全結合層を通して得られた値は各グリッド位置における物体カテゴリのスコアと二つの物体鑄型の位置、大きさ、信頼度となる。

## 3. 画像認識と音響発信を用いた注意喚起方式

### 3.1 注意喚起システムの構成

提案するシステムはビデオ識別とオーディオ出力の2つのサブシステムから構成する。提案システムのシステム構成を図1に示す。

周囲の映像を監視カメラで取得し、映像情報をオブジェクト検知機能へ出力する。オブジェクト検知機能では、入力された映像から映像内に存在しているオブジェクトの検

出を行う。検出されたオブジェクトの座標情報はゴミ判定機能へ出力する。ゴミ判定機能では、オブジェクトがゴミかどうかの判定を行う。オブジェクトがゴミであれば、座標情報を基にゴミが放置されているかどうかを判断し放置されていればメッセージ制御機能へ放置判定を出力し、雲台制御機能へゴミ座標を出力する。雲台制御機能へゴミ座標が入力されると、座標を基に雲台の制御情報を出力する。雲台では入力された制御情報を基に、雲台の向きを変更する。メッセージ制御機能にゴミの放置判定が入力されると、音の素材を選択する。選択された音素材は、スピーカに出力される。スピーカから不法投棄の容疑者に対して、入力された注意喚起の音素材を出力し不適切行為の注意を促す。

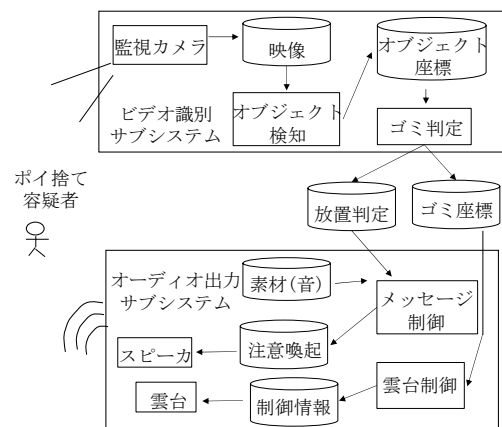


図1 提案方式のシステム構成

Figure 1 System design with the proposed method.

### 3.2 ビデオ識別サブシステム

ビデオ識別サブシステムは監視カメラを備え、入力されたカメラ画像から物体を認識するオブジェクト検知機能と、認識したオブジェクトがごみかどうかを判定するゴミ判定機能から構成する。オブジェクト検知機能では高速な物体検出を行う。オブジェクト検知機能の処理フローを図2に示す。

- Step1. 監視カメラから周囲の映像情報を取得する。
- Step2. オブジェクト検知機能に映像を出力する。
- Step3. 機械学習済みの学習モデルデータを読み込む。
- Step4. 入力された映像のフレーム取得を行う。
- Step5. オブジェクト検出を行う。
- Step6. 検出した座標を取得し、ゴミ判定機能へ出力する。

### 3.3 オーディオ出力サブシステム

オーディオ出力サブシステムは、スピーカとメッセージ制御機能と雲台と雲台制御機能から構成する。容疑の疑わしさを示すゴミの放置判定の情報に基づいて、メッセージ制御機能は、注意喚起に用いる音を選択する。注意喚起の音は音声メッセージに加えて優しいメロディ音や警告音などである。

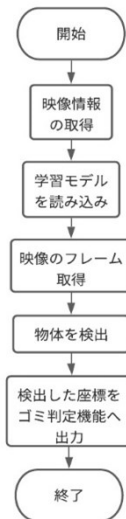


図 2 ビデオ識別サブシステムの処理フロー

Figure 2 Process flow in the video recognition subsystem.

## 4. 画像実験と音響実験による実現可能性調査

### 4.1 実験目的

オブジェクト検知やゴミ判定の性能によっては、不適切行為者を見逃したり、逆に適切行為者に対して警告を発したりするような機械判定誤りの懸念がある。また、機械判定の正誤に関わらず、頻繁な音響発信が社会環境にとっての新たな騒音源となつてはならない。画像処理と音響処理で利用可能な要素技術の評価を行い、提案する注意喚起方式の運用方法や実現可能性について検討を加えることを実験目的とする。

### 4.2 画像実験

#### 4.2.1 プロトタイプ実装

機械学習済みで提供されている YOLO を用いてリアルタイムオブジェクト検知の基本機能を作成した。コア数 6・1.80GHz の CPU、8GB メモリ、64bit-OS を備えた PC 環境を用い、confidence を指定し、写真や動画フレームから逐次オブジェクト検知ができる。図 3 にプロトタイプの実行画面例を示す。動画入力からペットボトルが検出された際に矩形で検出された物体を囲み、検出された物体名と confidence を矩形上部に表示する。

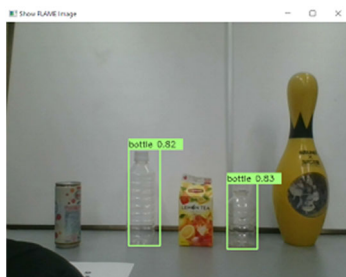


図 3 実行画面例

Figure 3 Display image of real-time object detection.

### 4.2.2 画像実験の方法

生物 (犬、猫、ハト、スズメ、ヒト)、仮想のごみ (くまのぬいぐるみ、犬型の玩具、ペットボトル、ビン、缶) の 10 点の写真を使用する。図 4 に例示するように、4032x3024 (犬型の玩具、図 4 右上)、3264x2448 (猫、図 4 右下) などの複数解像度を用いる。また、それぞれの縮小画像(100×100)に対しても実験する。confidence が 85%以上で検出できたものを検出と扱う。

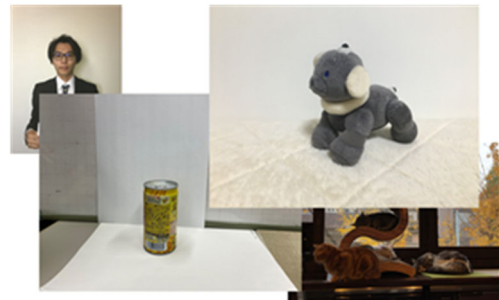


図 4 画像実験での検査画像例

Figure 4 Samples in 10 test images for object detection.

### 4.2.3 画像実験の結果

画像実験の結果を表 1 に示す。表 1 に示す通り、機械はゴミと判定しなかったが実はゴミだった場合の false negative は 29%(=2/(2+5))だった。また、機械はゴミと判定したが実はゴミでなかった場合の false positive が 0%(=0/(3+0))だった。誤検出を起こしたのは犬型の玩具と缶の画像であった。またゴミと判定したペットボトル、ビンの検出では confidence が 85%であった。一方、縮小画像(100×100)に対してはいずれもオブジェクト検知に失敗した。

表 1 オブジェクト検知の実験結果

Table 1 Results of object detection test

		検出結果	
		ごみ	非ごみ
検出対象	ごみ 5	3	2
	非ごみ 5	0	5

### 4.2.3 画像実験結果についての考察

ペットボトル、ビンでは confidence が 85%と検知精度が低かったため、オブジェクトの一部でも隠れてしまうと上手くオブジェクト検知できないことが考えられる。

オブジェクト検知後のゴミ判定において、鳥のような生物が画面に入ってきててもゴミと考えるべきではない。一方、法的には非生物の投棄は不適切行為である。それでも路上に一時的にカバンを置いた歩行者がいた場合にカバンが非生物と検知できたとしても、投棄なのか合理的な行為なのかのリアルタイム判定は難しい。ゴミ判定処理には倫理的な判断を要し、運用方法についての継続研究を要する。

例えば、最初は優しく注意喚起を促して歩いて歩行者が気づかないようなら大きく警告するなどが考えられる。

### 4.3 音響実験

#### 4.3.1 音響実験の方法

使用するスピーカ及びマイクの仕様をそれぞれ表 2、表 3 に示す。指向性のスピーカは設定や聞く位置によって聴こえ方が大きく異なる。無指向性のマイクは人に聞こえる音響と似た状況で集音するため環境雑音も自然に拾う。

音源には表 4 に示す音声、メロディ、警告音を用いる。音声には Text to speech ソフト[10]を使用し、メッセージは“ホイ捨ては止めましょう”とする。メロディ音、警告音はそれぞれ文献[11]、文献[12]の音源を利用する。聴こえる音の計測には PC 標準搭載の録音ソフト、と音圧測定ソフト[13]と波形表示ソフト[14]を用いる。

図 5 に示すように、大学構内の車道沿いにて実験を行う。また、スピーカの高さを地面から 1.4m に設置し、マイクの位置を地面にする。図 6 に示すように、スピーカの前後左右にヒューリスティックに測定地点を置き、音圧レベルの測定を行う。

予備実験から周りの騒音にある程度負けずに音が聞き取れるスピーカ出力を 80dB と定めた。表 2 の設定 1、設定 2、設定 3 を変えて計測を行う。音源から聞こえる音が周りの騒音とある程度近い音量だと主観的に判断した場合には、補正後の音源音圧レベル(dB)を  $x$ 、音圧レベル測定値(dB)を  $y$  として、数(1)を用いて補正する。自動車通過時などで騒音の方が音源より大きいと判断した場合は録音しな

$$x = y - 10 \log 2 \dots \dots \dots (1)$$

被験者にスピーカから 15m 離れた点に立ってもらい、指向性を被験者へ向け 3 種類の音源を流しどの音源が聞き取りやすかったかアンケートを行う。表 5 に示す評点と評価語を用い、平均した値を評価点とする。

表 2 スピーカ仕様

Table 2 Speaker specification

音圧レベル	≦93dB / 0.5m
指向角	±12°
周波数範囲	400Hz～20KHz
電力	12W / 5W (最大/平均)
設定切り替え	設定1.点音源 設定2.指向性音源を水平方向に向けた場合 設定3.指向性音源を斜め下に向けた場合

表 3 マイク仕様

Table 3 Microphone specification

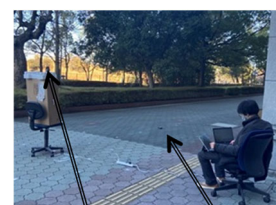
指向特性	無指向性(360° 集音)
集音範囲	0.8～3.0m
周波数特性(マイク)	150Hz～7500Hz
推奨温度	10℃～40℃

表 4 音源特性

Table 4 Characteristic of sound sources

呼称	ピーク周波数(Hz)	FFT 波形例
音声	220	
メロディ	180	
警告音	700	

FFT (Fast Fourier Transformation) 波形例では、横軸を周波数(20Hz～10000Hz)の対数軸、縦軸をスペクトル強度 0dB～100dB とする。



スピーカ マイク

図 5 実験風景

Figure 5 Experiment overview.

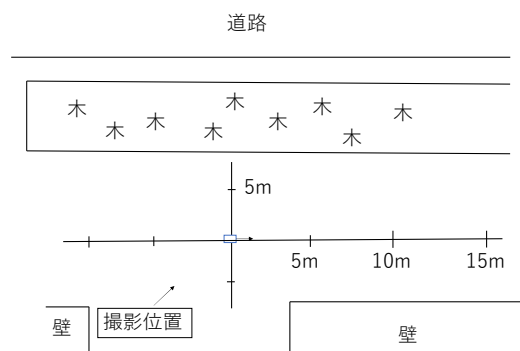


図 6 実験エリアの俯瞰図

Figure 6 Bird view of experiment area.

表 5 評価語

Table 5 Evaluation words for questionnaire

評点	評価語
5	非常に聞き取りやすい
4	聞き取りやすい
3	どちらともいえない
2	聞き取りにくい
1	非常に聞き取りにくい

#### 4.3.2 音響実験の結果

実験日は気温 5℃、湿度 71%、天候は晴れであった。マイクの推奨温度を下回る。警告音を音源として、スピーカ

設定における設定 1, 設定 2, 設定 3 の 3 通りを変えて比較実験を行った。各測定地点の音量計測結果を含む等圧線図を図 7、図 8、図 9 にそれぞれ示す。なお、図中の等圧線は計測結果から目視で主観的に補間して作成した。

図 7 に示すように、設定 1 (点音源) では、スピーカ近くは音圧が高く (75.9dB)、遠くなるほど周りの騒音に紛れていく。図 8、図 9 に示すように、設定 2 (水平方向) および設定 3 (斜め下) では、左右方向及び後方では前方と比べて音圧が低い。加えて、設定 3 では、マイクがスピーカに正対する位置となる 15m 前方で音圧最大値 (77.2dB) が計測された。

主観アンケートに被験者 4 人の協力を得た。音の聴こえ方についての調査結果を表 6 に示すように、評点平均値は音声 が 2.0、メロディが 3.5、警告音が 4.0 だった。

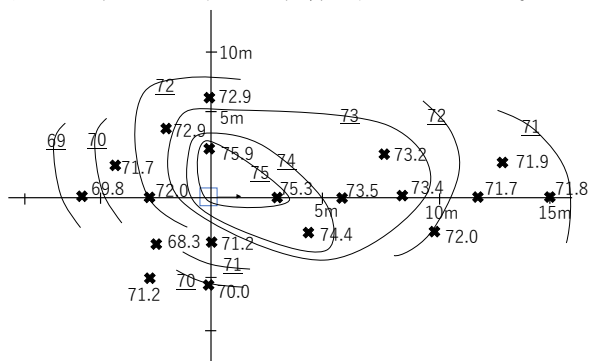


図 7 点音源を用いた場合 (dB)

Figure 7 Sound pressure level with point sound source (dB).

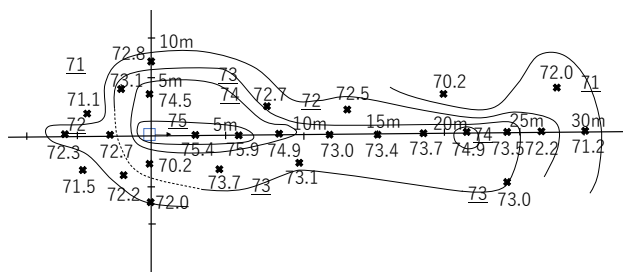


図 8 水平方向にスピーカを向けた場合 (dB)

Figure 8 Sound pressure level with horizontally-directional speaker (dB).

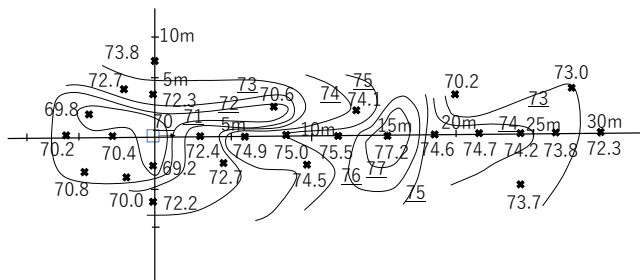


図 9 斜め下にスピーカを向けた場合 (dB)

Figure 9 Sound pressure level with diagonally-below directional speaker (dB).

表 6 音の聴こえ方についての調査結果

Table 6 Questionnaire survey results for hearing

音源	被験者				平均値
	A	B	C	D	
音声	2	2	3	1	2.0
メロディ	3	4	3	4	3.5
警告音	5	3	4	4	4.0

### 4.3.3 音響実験結果についての考察

図 8 と図 9 でスピーカの角度を変えた実験結果を比較すると、約 15m 先に 5m 範囲ほどのスポット音圧が生じる状況が読み取れる。表 2 に示す製品仕様での指向角が  $\pm 12^\circ$  なので、15m 前方では  $6.4m (=2 \times 15 \tan 12^\circ)$  が選択的に音圧の届く範囲の理論値である。実験に用いたスピーカはおよそ製品仕様通りの振舞いを示す。

### 4.4 提案システムの運用方式

4.2 節で述べたように、試作したオブジェクト検知の基本性能は false negative 29%、false positive 0% と一定の性能を示した。4.3 節で述べたように、指向性スピーカを用いた注意喚起は直進 15m の場所に 77.2dB の音圧が届いた。これらの評価結果に基づいて、街中で提案システムを運用する具体例として大学の通学路を取りあげてユースケースを整理する。提案システムの運用イメージを図 10 に示す。仮想設置場所の実写風景を図 11 に示す。

図 10 において、監視対象エリアの反対側の歩道の端にポールを立て提案システムを配置する。提案システムから発信された音が、約 15m 前方の容疑者に向けられる想定において、監視カメラの視野角は標準レンズ搭載時に  $46^\circ$  と想定すると、監視対象エリアの幅は  $13m (=2 \times 15 \times \tan(46^\circ/2))$  となる。監視カメラ画格を  $2048 \times 1024$  とするハイビジョン映像を撮影するとき、13m 幅の監視対象エリアにある 20cm のペットボトルは 31.5 画素 ( $=2048 \times 20/1300$ ) に相当する。

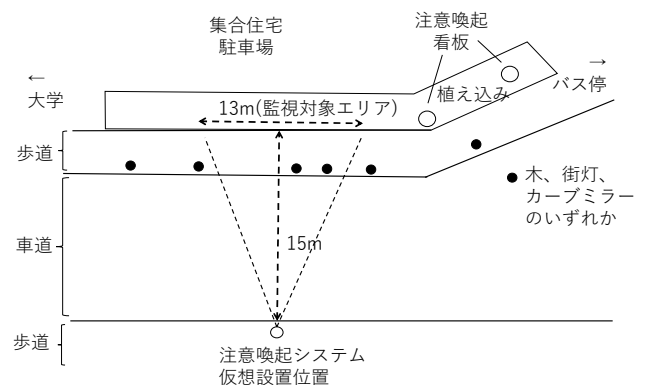


図 10 注意喚起システムの運用イメージ

Figure 10 Operation image of reminder system.





注意喚起システム 監視対象エリア  
仮想設置位置

図 11 仮想設置場所の実写風景

Figure 11 Photograph of the virtual installation location.

4.2.2 節で述べたようにオブジェクト検知機能は小さなオブジェクトに対する判定性能は弱い。現状のオブジェクト検知機能を用いる場合、監視カメラを複数化・高精細化してオブジェクトの長さが 300 画素以上となるよう運用するか、32 画素の長さのゴミを検知できるような機械学習をやりなおす必要がある。いずれにしても、小さい物体においては広い範囲での画像認識を行うことは難しいため、ゴミの不法投棄が集中する範囲で運用するとより効果的に使用できる。認識誤りを含んだ音響発信には騒音対策を要する。

一方、4.3.3 節で述べたように、指向角 $\pm 12^\circ$ の音響発信が選択的に届く幅は 15m 前方で 6.3m であり、監視対象エリアの幅 13m の半分に満たない。従って、(1) 近隣の騒音環境問題が深刻でなければ無指向性マイクを使う構成、(2) 2 台以上の指向性マイクを容疑者の認識位置に応じて切り替え制御する構成、(3) 指向性マイクの雲台を容疑者の認識位置に向ける制御を行う構成 (図 1)、がシステム設計での選択肢となる。

図 11 の画像左側は閑静な住宅地であり、図 7 に示すように、スピーカ背面の近隣住宅側の方が容疑者に聞こえるより大きな音圧となる構成(1)の注意喚起システムの設置は考えにくい。構成(2)では容疑者を検知した位置によっては十分な音響発信時間が得られない場合が懸念される。構成(3)において、容疑者がポイ捨てを行う場合に通常の歩行速度 4km/h の半分(2km/h)で歩くと想定し、指向性スピーカが容疑者に正対するよう雲台が制御されると、容疑者は指向性音響範囲 6.3m の半分の距離を歩くのに約 6 秒(=6.3m/2 ÷ 2km/h)要する。その間の音響発信は容疑者へのアクティブな注意喚起として不正行為抑制を期待しうる。

被験者へのインタビューでした中から注目した意見には、「音声は急に聞こえても文章として理解するまで時間がかかる」と「警告音は聞き取りやすいが急に流れると驚く」があった。設置場所や被害の酷さによって用いる音源を変えるべきだと考える。

## 5. おわりに

監視カメラの映像からポイ捨てを検知すると容疑者に向けてスピーカから音を流す注意喚起方式を提案した。既存技術 YOLO を用いて提案方式の一部をプロト開発し、利用可能な画像処理と音響処理の要素技術について評価実験を行った。オブジェクト検知機能はごみの識別において一定の性能を示す。指向性スピーカを用いた注意喚起の 80dB の音響発信は直進 15m の場所に 77.2dB の音圧で届く。評価実験結果に基づいて実環境でのケーススタディを行い、注意喚起システムの運用方式と課題を明確化した。ゴミ判定や雲台制御などのシステム構成要素が連携作用する前提において容疑者に 6 秒の注意喚起をアクティブに行える。提案方式にはある程度実現可能性がある。

今後の課題は次のものを含む。オブジェクト検知性能の向上を要する。また、ゴミ判定と雲台制御の機能開発を含むシステム実装を進め、提案システム全体としての評価を行うことが課題である。

## 参考文献

- [1] UN: "SDGs", <https://sdgs.un.org/goals> (1965).
- [2] 外務省: "持続可能な開発目標(SDGs)達成に向けて日本が果たす役割" (2021).
- [3] 大阪市環境局: "SDGs を意識した施策例(家庭系ごみ)" (2019)
- [4] "相次ぐ不法投棄 コロナが変えた生活影響か", 日テレ NEWS24 (online) (2020-10-28).
- [5] 橋本俊哉: "「ゴミ捨て行動」の心理と誘導方策", 農業土木学会誌, Vol. 70, No. 2, pp.101-104 (2002).
- [6] "米 NY 市警、ロボット犬の運用中止 反発の声強く", AFPBB News (online) 最終閲覧日(2021-11-25)
- [7] Fareed Boedbar, Roya Salehzadeh, Christian Cousin, Darrin J.Griffin, Nader Jalili, "Analyzing Human-Robot Trust in Police Work Using a Teleoperated Communicative Robot", 30th IEEE Int'l Conf, on Robot & Human Interactive Communication (RO-MAN), pp. 919-924 (2021),
- [8] 藤吉弘互,山下隆義, "深層学習による画像認識", 日本ロボット学会誌, Vol. 35, No. 3, pp180-185 (2017).
- [9] 山下充康,子安勝: "線状音源に対する障壁の遮音効果:模型実験による検討", 日本音響学会誌, Vol. 35, No. 4, pp. 207-203 (1973)
- [10] VOICEVOX, <https://voicevox.hiroshiba.jp/>, 最終閲覧日(2022-1-5)
- [11] 甘茶: "3時のおやつ", 甘茶の音楽工房, <https://amachamusic.chagasi.com/>, 最終閲覧日(2022-1-5)
- [12] OtoLogic, <https://otologic.jp/>, 最終閲覧日(2022-1-5)
- [13] MP3Gain, <http://mp3gain.sourceforge.net/>, 最終閲覧日(2022-1-5)
- [14] WaveSpectra, <https://efu.jp.net/soft/ws/ws.html>, 最終閲覧日(2022-1-5)
- [15] 森康浩,中俣友子: "河川でのポイ捨てを減らそう!", 日本環境心理学会第 13 回大会, Vol. 8, No. 1, pp. 26-34 (2020)
- [16] 中俣友子,阿部恒之: "ゴミのポイ捨てに対する監視カメラ・先行ゴミ・景観・看板の効果", 心理学研究, Vol. 87, No. 3, pp. 219-228 (2016).
- [17] 田縁正明,松村真宏: "指向性スピーカを用いた歩きスマホ防止策「おしゃべりスマホ」", エンタテインメントコンピューティングシンポジウム論文集, 208-213 (2016)