

視覚障害者の屋外移動時における周辺物体検出の可聴化に関する基礎検討とプロトタイピング

馬場 哲晃^{1,a)} 石曾根 奏子^{1,b)} 高野 衛¹ 安藤 大地¹ 渡邊 英徳² 釜江 常好³

概要：我々はこれまで視覚障害者の屋外活動移動を支援する為、深層学習を利用した物体検出及び、そのデータセット開発を行ってきた。本稿では開発したデータセットをもとに、検出した物体に対する可聴化手法を検討した。カメラから検出された物体に対してユーザに聴覚を通じて情報提示するためには、視覚においては空間、聴覚においては時間がそれぞれに適したマッピング情報であるため、変換モデルが必要となる。そこで可聴化に関していくつかの手法をプロトタイプし、本手法に適した可聴化手法をヒューリスティックに検討、改良した。本稿では設計した検出物体と出力波形のマッピングおよび、タイムラインベースなシーケンシャル可聴化手法に関して報告する。

キーワード：視覚障害、支援技術、アクセシビリティ、深層学習、物体検出、プロトタイピング、データセット、可聴化、ソニフィケーション

TETSUAKI BABA^{1,a)} KANAKO ISHISONE^{1,b)} MAMORU TAKANO¹ DAICHI ANDO¹ HIDENORI WATANAVE²
TSUNEYOSHI KAMAE³

1. 背景

我々は2018年度より、視覚障害者の屋外移動支援のためのデータセット設計および開発を行ってきた [1][2][3][4]。1年間の開発を通じ、図1に示すとおり認識結果となり、データセット数は少ないが試作段階においては実験可能な動作を実現した。登録したデータセットに関する詳細は本稿では省くが、現時点(2019年7月23日)にて、約4万程度のバウンディングボックスのラベリング作業を行った。今後も継続してデータセットを拡充していく一方で、検出システムを利用した情報提示に関してプロトタイピングが可能となったことから、本稿においてはユーザに対するフィードバック手法をプロトタイピングをもとに議論、検討を進めていく。図2に示すとおり、本研究では可能な限り特別なデバイスを必要とせず、容易にシステムを体験で

きることを優先し、プロトタイプ開発を進めていく。なお、本研究で対象とする視覚障害は原則として全盲を対象としている。画像情報を視覚障害をもつユーザにどのように提示するかに関しては、これまで多くの研究が報告されているが、特に二次元画像情報を提示するには提示手法の親和性から触覚が用いられることが多い。触覚提示は触図が広く普及しており、これらを即時提示に適応させる場合、機構が大掛かりになる問題がある。これに対して近年では電気刺激を用いた提示によって、アクチュエーティング部の軽量化が図られた事例がいくつか報告されている。梶本ら [5] は手掌部に対して、視覚情報をマトリクス状に配した512の電極提示により視覚情報の触覚化を実現している。またフォトトランジスタ駆動にスマートフォンディスプレイの発光を用いることで、低コストで高密度な電気触覚提示を実現している点が、本研究における既存デバイス活用の観点から興味深い。提示箇所については、額を利用したもの [6]、舌を利用 [7] 等のいくつかの事例が見られる。これによりユーザは視覚情報を即座に触覚情報に変換することは可能であるが、画像における明暗やエッジ、サーフェース等の単純な画像情報に限られている。また触覚提示には専用のデバイスが必要となる。本研究ではプロトタイプ体験価値を優先することから、システム自体をなるべ

¹ 首都大学東京
Tokyo Metropolitan University, Asahigaoka, Hino, Tokyo
191-0065, Japan

² 東京大学
The University of Tokyo

³ 東京大学/スタンフォード大学
The University of Tokyo/Stanford University

a) baba@tmu.ac.jp

b) kanako.ishisone@gmail.com

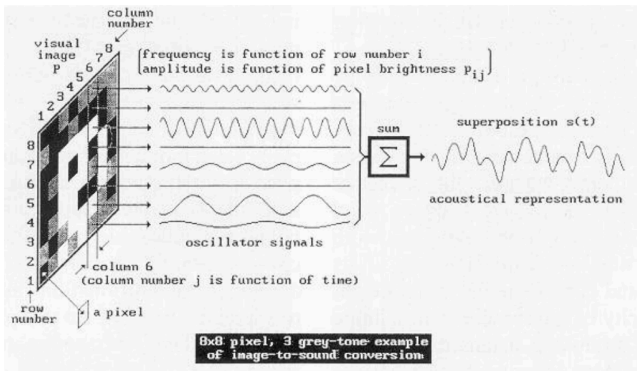


図 3 ピクセルデータから音情報へのマッピング。縦位置の高いピクセルほど高い周波数となり、画素値が大きい（明るい）ほど振幅が大きくマッピングされる。文献 [8] より引用

る物体に予め割り当てられた波形を生成する。また、当該物体の重心位置を基準に左右のボリュームコントロールも同時に行う。図 4 の場合、やや左よりから三角波が、ほぼ中心から三角波よりもやや周波数の高い正弦波が生成されている。検出物体と波形のマッピングに関してはこの時点では行わず、まずは全て同一の正弦波にて実装を行い、検討を開始した。

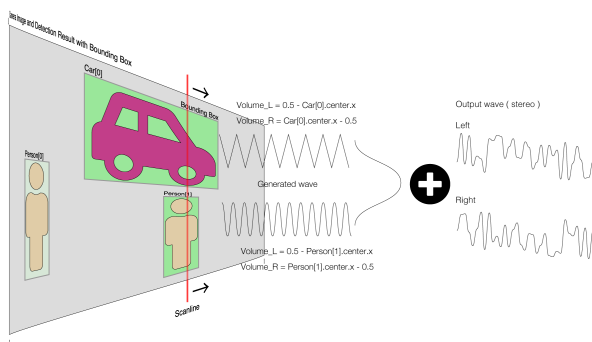


図 4 検出物体に対して任意波形を割り当て、走査線（図中の scan-line）が物体と重畳しているときに登録された任意波形を出力する。ただし物体の重心位置に応じて左右ボリュームが調整される。

3.1 物体と波形のマッピング

開発したプロトタイプを用いて、現代作曲家と可聴化を専門とする研究者とディスカッションを行い、暫定的に以下の指針を導き出し、マッピングすることとした。

- 点字ブロックや横断歩道等の歩行を誘導する物体に対してはホワイトノイズを利用し、ユーザに平面感を与える
- 動的障害物（車や自転車、人）に対しては注意を促す為、矩形波とする
- 段差や車止め、茂みやフェンス等の静的障害物に関しては正弦波を用いる
- 横断歩道の青、赤については実世界と同じ歩行誘導音

を利用する

以上の指針をもとに当該データセット 32 クラスに対して暫定的なマッピングを行った。

3.2 走査線パラメータ

走査線に関しては左-右、右-左、下-上、上-下、中心位置から 2 本の走査線がそれぞれ左右へ、中心位置から 2 本の走査線がそれぞれ上下へ移動するパターンを検討した。それぞれの場合を図示したものを図 5 に示す。今回のプロトタイプでは (1) が最も理解が容易であると判断し、(1) を採用したが、慣れによっては (5), (6) に関しても今後再度検討する可能性がある。なお走査線の速度は低速度とし、1 回の走査にかかる時間を 1,500[ms] としているが、走査線移動速度に関してはユーザ可変であることが望ましい。

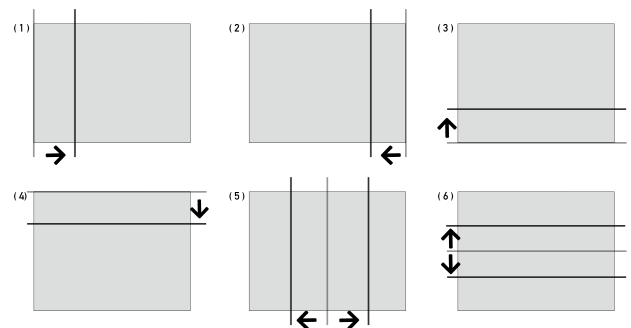


図 5 左-右、右-左、下-上、上-下、中心位置から 2 本の走査線がそれぞれ左右へ、中心位置から 2 本の走査線がそれぞれ上下へ移動するパターンをそれぞれ (1)-(6) とする。(5),(6) に関しては走査線が 2 本あり、それらが同時に矢印方向へ直進する。なお速度は定速度としている。

3.3 その他

物体が検出できていない場合、走査線位置情報を取得することができないため、ミュージックシーケンサの概念を踏まえクリック音を付与することにした。走査線の移動速度により BPM が変わることになるが、例えば一回の走査にかかる時間を 1,500[ms] とした場合、100[BPM]（四拍子）のクリックとなる。これによりユーザは物体が検出できなかった場合においても、走査線位置を把握することができるようになる。

図 6 に実際に動作している際のスクリーンショットを示す。ユーザは Loop モードをオンにすることで、繰り返し走査線を左から右へ自動的に移動させ、検出物体を可聴化できる。

4. まとめ

本稿では、物体検出を用いた、視覚障害者屋外歩行支援の為の可聴化手法に関してプロトタイプベースで検討

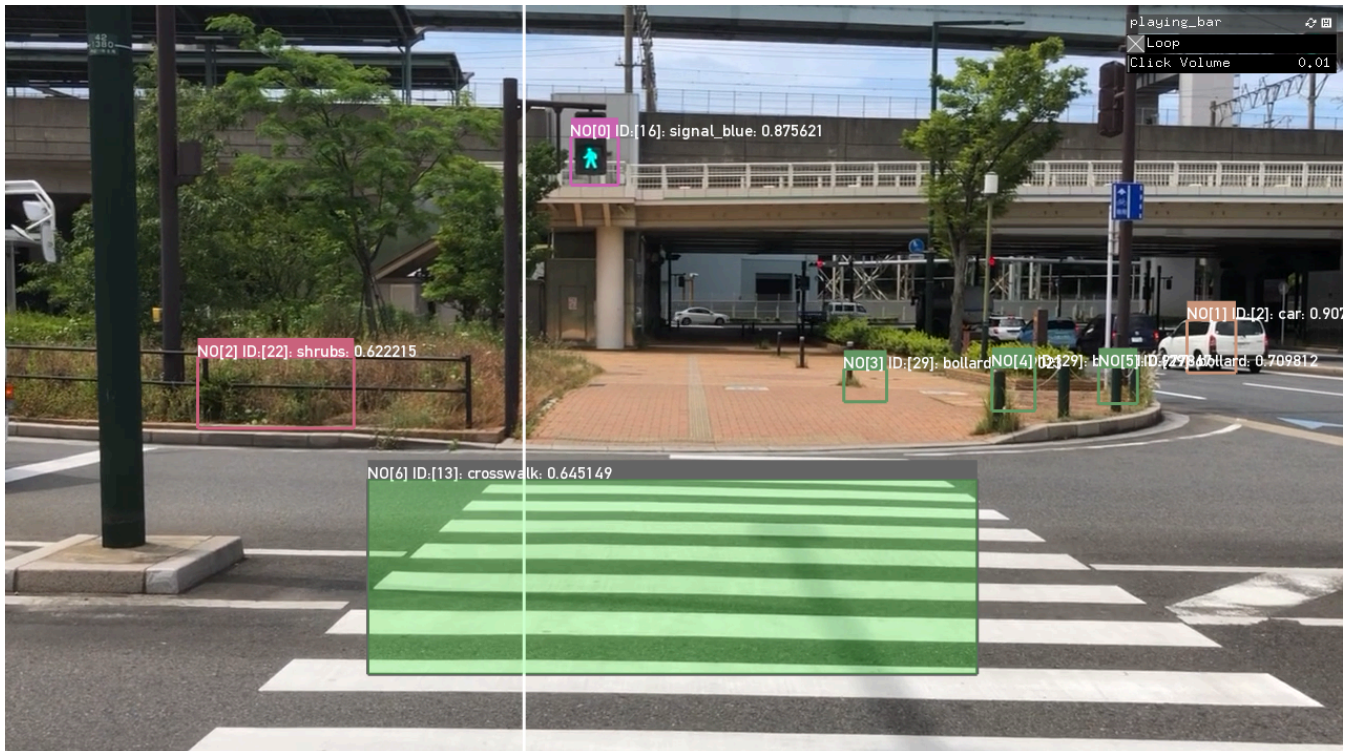


図 6 プロトタイプ動作時のスクリーンショット。縦方向に伸びる白い線が走査線。走査線と重畳している横断歩道が透明度をもった緑色として塗りつぶされている。この場面では横断歩道を示すホワイトノイズが中心定位より発音されている。また、signal.blue を認識しているため歩行者誘導音も合わせて出力されている。

を行い、ある程度の設計指針を示した。今後は空間把握精度のユーザテストや当事者からのヒアリングを進めていく。

謝辞 本研究は JSPS 科研費 JP18H03486 の助成を受けたものです。

参考文献

- [1] 馬場哲晃, 渡邊英徳, 釜江常好: 深層学習による物体検出を用いた視覚障害者の屋外活動支援システムにおけるデザイン指針の検討とプロトタイピング, 技術報告 8, 首都大学東京, 東京大学, 東京大学/スタンフォード大学 (2018).
- [2] 石曾根奏子, 馬場哲晃, 渡邊英徳, 釜江常好: 視覚障害者の屋外移動支援に向けた物体検出データセットの基礎検討とプロトタイピング, 技術報告 9, 首都大学東京, 首都大学東京, 東京大学, 東京大学/スタンフォード大学 (2018).
- [3] 石曾根奏子, 馬場哲晃, 渡邊英徳, 釜江常好: ユーザ参加型アノテーションにおける UI 及びデータオーグメンテーションのデザイン, 技術報告 1, 首都大学東京, 首都大学東京, 東京大学, 東京大学/スタンフォード大学 (2018).
- [4] 渡邊康太, 馬場哲晃, 田村賢哉, 渡邊英徳, 釜江常好: 広島平和記念公園における視覚障がい者の碑巡り支援の基礎検討, 技術報告 2, 首都大学東京, 首都大学東京, 東京大学, 東京大学, 東京大学/スタンフォード大学 (2018).
- [5] Kajimoto, H., Suzuki, M. and Kanno, Y.: HamsaTouch: Tactile Vision Substitution with Smartphone and Electro-tactile Display, *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI EA '14, New York, NY, USA, ACM, pp. 1273-1278 (online), DOI: 10.1145/2559206.2581164 (2014).
- [6] Kajimoto, H., Kanno, Y. and Tachi, S.: Forehead Retina

- System, *ACM SIGGRAPH 2006 Emerging Technologies*, SIGGRAPH '06, New York, NY, USA, ACM, (online), DOI: 10.1145/1179133.1179145 (2006).
- [7] Stronks, H. C., Mitchell, E. B., Nau, A. C. and Barnes, N.: Visual task performance in the blind with the BrainPort V100 Vision Aid, *Expert Review of Medical Devices*, Vol. 13, No. 10, pp. 919-931 (online), DOI: 10.1080/17434440.2016.1237287 (2016).
- [8] Meijer, P.: An experimental system for auditory image representations, *Biomedical Engineering, IEEE Transactions on*, Vol. 39, pp. 112 - 121 (online), DOI: 10.1109/10.121642 (1992).
- [9] Bologna, G., Deville, B., Pun, T. and Vinckenbosch, M.: Transforming 3D coloured pixels into musical instrument notes for vision substitution applications, *Eurasip Journal on Image and Video Processing*, Vol. 2007, No. 76204 (online), DOI: 10.1155/2007/76204 (2007).
- [10] Protasio Ribeiro, F., Florencio, D., Chou, P. A. and Zhang, Z.: Auditory Augmented Reality: Object Sonification for the Visually Impaired, *MMSP, IEEE*, (online), available from (<https://www.microsoft.com/en-us/research/publication/auditory-augmented-reality-object-sonification-for-the-visually-impaired/>) (2012).