

研究助成成果報告 研究助成 (S)

環境センサと融和した注意指向・取捨選択型
聴覚支援システムの研究開発

Attention-based sound-selective hearing support system in harmony with environment sensors

2198001



研究代表者	株式会社国際電気通信 基礎技術研究所	グループリーダー	石井カルロス寿憲
共同研究者	株式会社国際電気通信 基礎技術研究所	研究員	港 隆 史
	株式会社国際電気通信 基礎技術研究所	研究員	劉 超 然
	京都大学 情報学研究科	特定講師	Jani Even
	近畿大学 医学部 医学科	教授	土 井 勝 美
	聖マリアンナ医科大学 耳鼻咽喉科教室	研究員	加 藤 弓 子

[研究の目的]

世界各国で人口の1~2割が難聴・聴覚障害を持っていると言われており、さらに高齢化社会になって老齢性難聴者が増えれば、その割合は今後さらに高まると考えられる。従来の補聴器では不要な雑音まで増幅するなどの問題があり、健聴者並みに社会活動に参加するには不十分である。

本研究では、環境に設置されたセンサと連動して、周りの人が持っているスマートホン（以降、スマホ）も二次利用し、聴覚障害者が補聴器単体の知覚に比べて、健聴者並みの聴力が得られる方法の実現を目指す。人々と機械が融和・協調することによって当事者の聴力が向上され、社会参加の促進や、新たな生産的活動や創造的活動が可能になる。例えば、聴力が衰えたために他者と会話することに消極的になっている者が、聴覚能力拡張によって他者と関わる意欲を持つことができ、人間関係が疎遠になっ

てきたと言われている現代の社会問題やそれが引き起こす認知症の問題解決にも繋がる。更に、本研究の技術は健聴者も利用可能である。例えば、工事現場の騒音がうるさい環境下でも利用者が必要とする音や音声のみを聴くことが可能になり、聴覚を通じた異常・危険の察知やコミュニケーションを妨げない騒音性難聴の予防・改善を実現できる。

[研究の内容、成果]

最近の補聴器は空気中の振動から入ってくる音声信号だけでなく、Bluetooth 接続により自分のスマホを介して外部の音や音声信号を受け取る機能が備わっている。本研究では、これまで研究開発を進めてきた音環境知能技術を発展させ、利便性と実用性を重視しつつ、環境センサに加えて周りの人が持つスマホのマイク情報も二次利用とし、補聴器とも連動させて、周囲のさまざまな音の中から、利用者が聞きたい音

を聞こえやすくし、聞きたくない音を抑圧する技術（注意指向・取捨選択型聴覚支援）の実現に取り組んだ。図1にシステムの概要を示す。基本的な仕組みとして、マイクロホンアレイ（以降マイクアレイ）処理により、音の方向を検出したり、音を分離して強調したりして、LRF（レーザ距離センサ）により周囲の人の位置を推定し、誰がどこでいつ発話したのかを推定することができる。取捨選択された音源に対し、利用者が装着する姿勢センサにより、個々の音源の空間情報を再構築してイヤホンまたは補聴器に、バイノーラルで聞かせる仕組みとなっている。

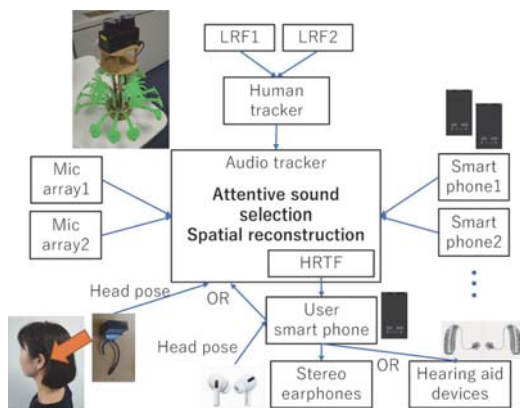


図1 開発した聴覚支援システムの概要図

以降、研究開発項目ごとの成果を記載し、開発した聴覚支援システムを聴力レベルの異なる難聴者によって検証した結果について報告する。**環境センサおよび姿勢センサのデバイスの開発**

環境に馴染むマイクアレイのデザインとして、植物型のアレイフレームを3Dプリンタで作製した。マイクには小型MEMSマイクを採用し、それを固定するフレームも新たに設計した。また人位置推定に用いるLRF（レーザ距離センサ）とアレイを一体型にしたフレームも作製した（図2）。利便性も考慮し、持ち運び可能な傘型のアレイフレームも試作した。

IMU（方位/慣性計測センサ）および小型マイコンを内蔵した姿勢計測デバイスを作製した。基板が平行となるように耳掛け型のケースを3Dプリンタでデザインし、樹脂の素材も何種

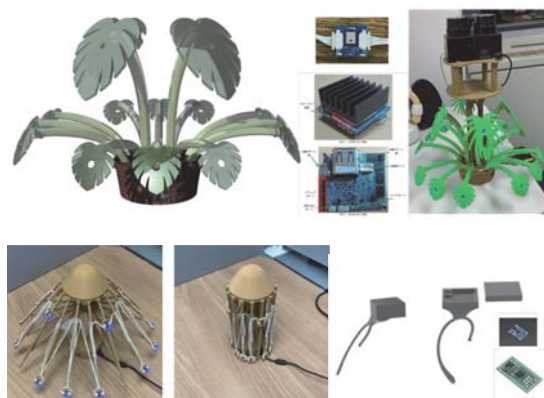


図2 開発した環境センサおよび姿勢センサのデバイス

類を試して、耳周りが痛くない柔らかい素材の装着デバイスを試作した。

スマホおよびイヤホン・補聴器との連携の開発

スマホのマイク音を聴覚支援システムに送信し、システムが提供する音をスマホで受信する機能について、Apple製およびAndroid OSのスマホに対応したアプリを開発した。スマホで受信したシステムからの音は、Bluetoothでスマホに接続されたイヤホンまたは補聴器でストリーミングされる。Bluetoothイヤホンにおいては、Apple製のAirPodsProに姿勢センサも搭載されており、スマホでその情報を読み取ってシステムに送る機能も実装した。

スマホのマイク音をターゲット（強調すべき音）にするか、アンチターゲット（抑圧すべき音）にするかの設定もアプリ上の画面操作で可能とした。ターゲットとして設定した場合、マイクアレイが提供できる音質を補う働きとなり、アンチターゲットとして設定した場合は、そのスマホの周辺の音は、聴覚支援システム上で抑圧される働きとなる。これにより、不要な雑音源をアンチターゲットとして設定することも可能になり、プライバシー保護として、明示的に特定エリアの音をシステムに提供しないことが可能になる。

スマホのマイク音の処理として、図3に示すようにアレイ処理の出力と同期し、チャンネル間の抑圧処理（Wienerフィルタ）にマイク音を導入した。

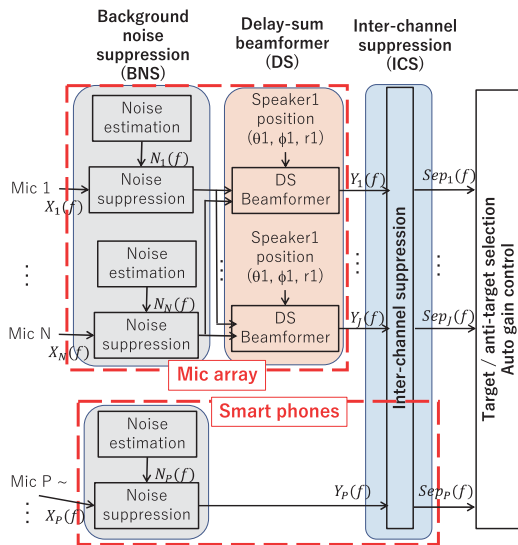


図3 アレイ処理とスマホマイクの処理の統合

遅延短縮のための対策

音を取得し、処理して利用者の装着デバイスで再生するまでの遅延を十分に短くするという、本技術の社会実装に向けての重要な課題に取り組んだ。オーディオキャプチャおよびオーディオプレイヤーの ASIO デバイスへの対応化 (PortAudio ライブラリーの導入) とオーディオデバイスの仮想 ASIO デバイス化によりキャプチャの遅延を 90 ms から 4 ms 程度に短縮した。また、音源分離処理において、図3の赤い枠に示すように、アレイ処理の背景雑音抑圧とビームフォーミングの並列化と、スマホのマイク処理の背景雑音抑圧とチャンネル間抑圧の処理の並列化により、20 ms 程度の短縮ができた。以前はプレーヤにも 90 ms の遅延が生じていたが、スマホを通して Bluetooth でイヤホンに音を送るため、この遅延は無線 LAN と Bluetooth 通信の遅延に置き換えられ、40 ms 程度に短縮される。総合的には、これまで 300 ms 以上かかっていた遅延を 100 ms 程度に短縮することができた。

注意対象の音量制御機能

対話中は、通常話し手の顔を見ながら聞くことを踏まえ、利用者が向いた方向の対象音源を強調する機能を実装した。姿勢センサから得られる利用者の顔の向きを用いて、利用者の前方

−90 度から 90 度の音源に対し、コサイン波に応じて真正面で最大 3 dB のゲインを加える注意対象機能を実装した。当初は最大ゲインを 6 dB で設定していたが、前方以外の対象音との音量差が大きすぎたため、3 dB に設定した。3 dB でも注意対象の音源は顕著に強調される感覚が与えられることは確認した。また、対象音の音量を上げると、分離音に残っている背景雑音の音調も増幅されてしまうため、対象音を分離する際には、増幅する分、雑音除去の度合いも強くするようにした。これにより、頭部を動かしても背景雑音の音量も同時に増幅されることを回避できた。

音源の自動選択機能

対象音を毎回選択する煩わしさと、人位置検出の追跡失敗により同じ音源に新しい ID が割り当てられてしまう問題も踏まえ、システムの利便性を向上させるため、利用者の周囲の指定範囲 (例えば 2 m) に新しく検出された人物は自動的に対象音として選択される機能を実装した。ただし、自動で選択された音源は、GUI インタフェースにより画面上でクリック (タッチパネルの場合はタッチ) することにより、利用者が手動で選択を解除したり、アンチターゲットとして再選択したりすることができる。

センサの自動キャリブレーション

現状では手動でセンサの位置や向きのキャリブレーションを行う必要があるが、利用者にとってのシステムの利便性を考慮すると、センサの自動キャリブレーションは必要な技術となる。本課題には、図4に示すような深層学習技

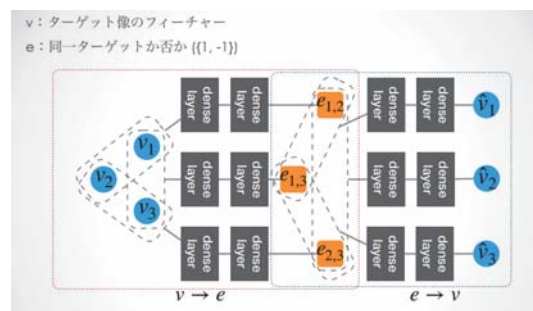


図4 深層学習によるセンサキャリブレーション法

術（graph neural network に基づく手法）を取り入れて、異なる種類のセンサ（アレイや距離画像センサ）を自動的にキャリブレーションする手法を提案した [1]。

スマホに適した音源方向推定手法の開発

スマホなどのモバイルデバイスの問題点として、電源供給と計算パワーが限られていることが挙げられる。その制約を考慮し、方向スキャンを要しない音源方向推定手法を提案した。従来法として、同じく方向スキャンを必要としない rootMUSIC が存在するが、計算した伝達関数しか用いることができないという制約がある。提案手法は、事前に計測した伝達関数を使用することが可能であるため、識別精度がより高い。具体的には、計測した各方向から伝達関数の位相情報のみを用いて Hash テーブルを構築し、観測した音声のマイク間位相差と照らし合わせることで音の到来方向を推定する手法である。検出された音の到来方向は、音源分離やスマホの位置推定にも利用される。

移動型センサによる 3 次元音環境地図生成の開発

移動ロボットを使って地図生成を行う研究はこれまで多く行われてきたが、本研究開発では、人が移動型センサを持って環境を歩きながらスキャンする課題に取り組んだ。センサを持ってゆっくり歩きながら周辺環境をスキャンした場合、地図生成に関するパラメータ調整により、十分な精度（10～20 cm の誤差）で 3 次元地図が生成可能であることを確認した。また、3 次元地図上に音情報を付加するソフトウェアを実装し、音情報は、画像深度とマイクアレイを備えた手持ち型センサ（Kinect Azure）を用いて、3 次元地図生成と同時に取得でき、20～30 cm の精度で 3 次元音環境地図が生成できた。上述はオンラインで音地図を生成した場合であり、オフライン処理で雑音成分の除去も考慮した処理を施すと音地図も 10～20 cm の精度で推定可能になった。図 5 に得られた地図の例を示す。これは利用者が新しい環境でシステムを利用した際に、雑音源を特定し、簡単に選択できるイ

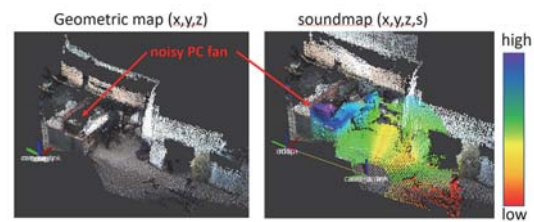


図5 3次元地図（左）と3次元音環境地図（右）の例

ンタフェースを実現するために利用される。

聴覚特性への適応

難聴者に対して、聴力に応じた周波数特性の補償が必要であるが、Bluetooth 機能で補聴器に直接システムから音をストリーミングすれば、ストリーミングされた音も補聴器でフィッティングされた特性が反映されるものが既に開発されている。（Oticon More 3, Starkey Livio, Resound, Signia, などが挙げられる。）

また、聴覚支援において空間情報を伝達する際の重要な課題として、視聴覚刺激による空間知覚の解明に取り組んだ。そのため、既存の回転刺激、視覚刺激に、聴覚刺激を追加制御するためのソフトウェアの拡張を実装した。この実験装置を用いて、前庭眼反射（Vestibulo-ocular reflex: VOR）の量的変化による空間知覚精度の評価指標を確立するための実験を行った。その結果、VOR の視覚刺激の影響から、視覚刺激の強度が強ければ、視角 20 度程度の「あいまいさ」があっても、おおよその方向としてとらえられる可能性が示唆され、聴覚支援システムが音像を合成する際に視角 20 度程度の範囲内で音像を提示できれば、大きな違和感がないことが期待される。なお、聴覚支援システムで使用している頭部伝達関数の分解能が 5 度であることから、音源の方向感覚は視覚情報とともに伝わることを示唆される。

開発した聴覚支援システムの検証

聴覚支援システムの効果を評価するための被験者実験を実施した。雑音環境を模擬するために、図 6 に示すように、複数のスピーカから複数の雑音（バブルノイズ 65 dBA, 妨害音声 65 dBA）を流し、そのうちの一つのスピーカか

ら対象となる言葉（対象音声 65 dBA）の聞き取りテストと、雑音環境下での 3 者対話実験を実施した。バブルノイズとは、多数人の声が混合した雑音で、特定の人の声を聞き取る際に最も妨害となる雑音である。また、バブルノイズを拡散性の（指向性のない）雑音として再現するため、2つの上向きのスピーカを被験者の両側の床に置いた。被験者には、この雑音が聞こえてくる方向が特定できないことを確認した。また、このスピーカは床上から 20 cm 程度の高さに位置するため、システムでは指向性のある音源として検出されないことも確認した。その他の音源（単語用のスピーカ、音楽用のスピーカ）は、LRF センサに見える 1.1 m 程度の高さに設置した。聞き取りテストの対象となるスピーカからは、FW03 のデータセットより、親密度が 55～70% の 4 拍単語リストから選定して流した。雑音の有無、システム利用の有無の条件ごとに同じ親密度で、異なる単語セットを使用した。三者対話実験では、話者は椅子に座って対話をする。雑音環境としては、騒がしい居酒屋で少し離れた席の人と会話をするような場面に相当する。

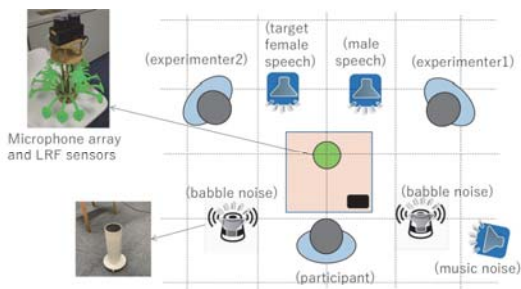


図6 実験に用いたスピーカや話者の位置関係（1 m grid）

スマホは、Apple iPhone SE を使用し、無線 LAN（wifi6）でシステムと接続し、bluetooth 通信により、イヤホン（AirPods Pro）もしくは補聴器（Oticon More 3, Starkey Livio）にシステムからの音をストリーミングした。AirPodsPro には姿勢センサも搭載されているため、スマホを通してシステムに姿勢情報を送信した。補聴器装用者の場合、システムと連携

した状態では、音が二重に聞こえないように補聴器のマイクの音はミュートに設定した。

実験は「雑音なし」（clean）, 「バブル雑音」（babble noise）, 「バブル + 音楽」（babble + music）の条件を「システム無し」と「システム利用」の条件で実施した。

実験参加者は 2 つの方法で募集した。一つはシルバー人材センターを通して老人性難聴の可能性が高い 65 歳以上の高齢者を募集し、28 名の応募があった。2 つ目は、共同研究機関の近畿大学病院より、補聴器を装着している中等度の難聴患者を募集し、3 名の応募があった。難聴患者の場合は、近大病院にて聴力に合わせた補聴器のフィッティングを事前に行ってから実験に参加してもらった。シルバー人材センターの被験者には、実験開始前に、オーディオメータ（リオン社 AA-M1C）を用いて、自動モードで聴力測定を行った。

聴力測定の結果を基に、被験者を 4 つのグループに分類した。

- ・「正常」：平均聴力レベルが 25 dB 以下（ 16 ± 4 dB）が 5 名。
- ・「軽度難聴」：平均聴力レベルが 25～40 dB（ 30 ± 10 dB）が 9 名。
- ・「中等度難聴」：平均聴力レベルが 40～70 dB（ 48 ± 9 dB）が 13 名。
- ・「中等度難聴の補聴器装用者」：Bluetooth 対応の補聴器を装着する中等度難聴者（ 48 ± 9 dB）が 4 名。

単語明瞭度

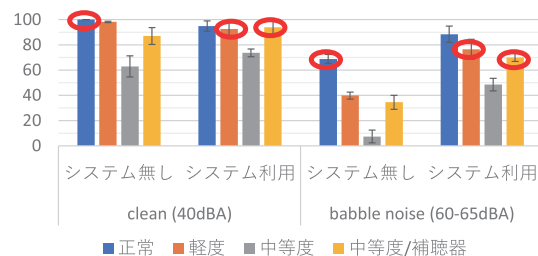


図7 システム無しとシステム利用の条件における明瞭度

聴力グループごとの単語明瞭度の結果を図7に表示する。静かな環境（clean）では、シス

テム無しでは、「正常」「軽度」の被験者はほぼ100%の正解率を示すのに対し、「中等度」では平均60%程度で、補聴器装用者「中等度/補聴器」は90%前後である。雑音下（babble noise）の条件では、システム無しでは、すべてのグループで正解率が下がり、「中等度」では10%前後で、ほとんど聞き取れていない状態である。

それに対し、雑音環境（babble noise）で、「システム無し」から「システム利用」の条件の変化に着目すると、単語明瞭度が「正常」：70% → 90%、「軽度」：40% → 80%、「中等度」：10% → 50%、「中等度/補聴器装用者」：35% → 70%、のように、すべてのグループで上昇し、システムの効果が示された。

音節単位で明瞭度を測定した結果（グラフは省略）も、「正常」：80% → 95%、「軽度」：55% → 85%、「中等度」：15% → 65%、「中等度/補聴器装用者」：55% → 85%のように、システム利用により正解率が上昇し、単語の一部のみ正しく聞き取れなかったことが分かる。

最後に、静かな環境（clean）でも、システム利用により、「中等度」のグループによる音節明瞭度は70%から85%まで上昇した。

三者対話実験における主観的評価においても、システム無しとシステム利用で、音楽の雑音を付加しても、印象が「聞き取りにくい」から「聞き取りやすい」に著しく変化した結果が得られた（図8）。会話中も、システム無しの条件では、前のめりになったり、聞き返しが多かったり、会話が全く噛み合わない場面も見ら

れたが、システム利用により、会話を円滑に楽しんでいる様子が見られた。

システム利用の聞き取りやすさ以外の印象評価においても、聴力に依らず、顔を向けた方向の音が際立って聞こえる機能は「良い」、声が聞こえる方向は「だいたい分かる」の評定が得られた。興味深い結果としては、被験者のうち、3名が片耳難聴であったが、うち2名は声が聞こえる方向はだいたい分かるかと答えた。これは、音の空間的情報を再構築しているからではなく、顔を向けた方向の対象音が際立って聞こえるから音の方向が特定できたものと考えられる。

実験を終えての感想として、中等度難聴の補聴器装用者のうち、現役の弁護士が1名いたが、「本システムを是非法廷で使いたいので早く実用化してほしい」、その他、「食事会でうるさいから補聴器を外してしまうがシステムがあれば友達と会話ができそう」、「自治会の会議で使いたい」等の感想もいただいた。

本実験の結果を踏まえて、システム利用により、軽度難聴者も、補聴器を装用した中等度難聴者も、健聴者の聴力に近づけることができたと言える（図7の赤丸）。補聴器を普段装用している中等度難聴者は、日常では補聴器単体として使用し、必要な場面（騒がしい環境や聞き逃しができないような場面）において、聴覚支援システムと連携して聴力を上げることが出来る。一方、軽度難聴者は普段補聴器を装用しない人が殆どであるが、聴覚支援システムと連動して市販の AirPodsPro のようなイヤホンを装用する場合でも十分に効果が得られることが分かった。（本研究の成果は論文にまとめて、投稿中である。）

[今後の研究の方向、課題]

聴覚支援システムをより快適に利用者に提供するためには、装着デバイスを可能な限り簡略化する必要がある。軽度難聴者の場合は装着デバイスとして、AirPodsPro のイヤホンのみで

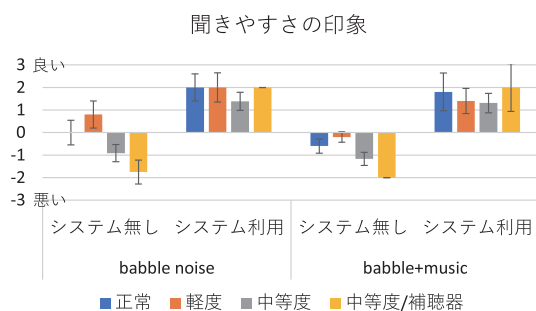


図8 システム無しとシステム利用の条件における印象

利用可能であるが、補聴器の機能を必要とする利用者は、姿勢センサも別途装着する必要がある。姿勢センサを搭載した補聴器が開発されれば、この問題は解決するが、メーカーとの共同研究開発が必要であり、課題として残される。

難聴者への支援に関しては、補聴器装用者を対象者としたが、高度難聴者で人工内耳を必要とする利用者においては、本システムを人工内耳と連携した研究開発が新たに必要となる。

なお、本プロジェクトで研究開発を進めた技術の一部は、聴覚支援への応用に限らず、ロボットインタラクションの分野にも活用している。騒音環境でも特定の人の発話区間の検出し、その人の声を音声認識部に提供する幅広い応用が可能である。実際、開発したシステムの一部は、対話ロボットの対話相手の声を抽出する前処理として活用している。車内対話エージェントの実現に向けて、車内のさまざまな騒音環境で運転者の発話区間のみを高い精度で検出するのに有効であることも検証した [2]。現在、遠隔操作型ロボットシステムにおいても、ロボッ

ト側の音環境を操作者側で再現して臨場感を高めるなど、AR (augmented reality) 技術としても応用・発展を進めている。

[成果の発表、論文など]

1. 劉超然, 石井カルロス (2019). “マイクロフォンアレイおよびデプスセンサーのオンラインキャリブレーションに関する考察” 人工知能学会研究会 AI チャレンジ研究会, pp. 18-23, 2019.
2. C. T. Ishi, A. Utsumi, S. Nagasawa (2020). “Analysis of sound activities and voice activity detection using in-car microphone arrays,” Proc. of the 2020 IEEE/SICE International Symposium on System Integration (SII 2020), pp. 640-645, 2020.
3. C. T. Ishi, C. Liu, T. Minato (2023). An attention-based sound selective hearing support system: evaluation by subjects with age-related hearing loss, Proc. of the 2023 IEEE/SICE International Symposium on System Integration (SII 2023), accepted.
4. C. Liu, C. T. Ishi (2023). A Smartphone Pose Auto-calibration Method using Hash-based DOA Estimation, Proc. of the 2023 IEEE/SICE International Symposium on System Integration (SII 2023), accepted.