第50回 東海ファジィ研究会
（蒲研2021）
予稿集

主催：日本知能情報ファジィ学会・東海支部
共催：日本経営システム学会・中部支部
日時： 2021年8月5日（木）
会場：オンライン(ZOOM)

第 50 回東海ファジィ研究会(2021 年 8 月 5 日)
プログラム

8 月 5 日(木)　会場：オンライン(ZOOM)

開会の挨拶 13:30～13:40	東海支部 支部長　川中 普晴（三重大学)
一般講演 13:40～15:00	座長：　ジメネス フェリックス（愛知県立大学)
	1. 多層パーセプトロンへの電力サイドチャネル対策の検討 野崎佑典　吉川雅弥
	2. XYXY 型オノマトペの音響特徴量を用いた 日本語母音の音象徴に関する一考察 兼岩尭希　中村剛士　加納政芳　山田晃嗣　犬塚信博
	3. FCR 法応答の統合値に関する新提案 ― オーケストラ音響刺激の認知を実験課題としたデータの処理 ― 小田哲久　河辺義信　山田洋巳
	4. 少数言語保存活動の支援 ― 録音信号に対するノイズ除去に関する一検討 木村圭吾　松下真也　高瀬治彦　北英彦
休憩（20 分)	
企画(研究紹介) 15:20～16:10	座長：川中 普晴（三重大学)
	➢ 東海ファジィ研究会メンバーによる研究紹介(紹介：5 分，質疑応答：5 分)
	1. 川中普晴（三重大学)
	2. ジメネスフェリックス（愛知県立大学)
	3. 高野敏明（静岡理工科大学)
	4. 矢野良和（愛知工業大学)
	5. 岩堀祐之（中部大学)
閉会の挨拶 16:10 ～16:20	日本知能情報ファジィ学会 東海支部 支部長 川中 普晴（三重大学)

多層パーセプトロンへの電力サイドチャネル対策の検討

○野崎　佑典[1]，吉川　雅弥[2]
○Yusuke Nozaki[1], Masaya Yoshikawa[2]
[1, 2]名城大学
[1, 2]Meijo University

Abstract:　近年，AI に対する様々な攻撃が提案されており，対策技術の検討は非常に重要である．本研究では，AI に対する電力サイドチャネルを使用したモデル抽出攻撃への対策手法を提案する．提案手法では，計算順序をランダム化するシャッフリングによって攻撃への耐性を向上させる．実機を用いた評価実験では，対策手法によって電力サイドチャネルを利用したモデル抽出攻撃への耐性が向上することを明らかにした．

１．はじめに

現在，AI 技術は様々なところで利用されていることから，AI 自身のセキュリティへの関心が高まっている [1]．AI のセキュリティ上の脅威として，AI 推論器で利用する Neural Network（NN）の重みや構造などのモデル情報を推定するモデル抽出攻撃が報告されている．具体的には，これまでにソフトウェア実装の多層パーセプトロン（Multi-Layer Perceptron: MLP）を対象に，回路動作時の処理時間や消費電力，電磁波などのサイドチャネル情報を利用したモデル抽出攻撃が提案されている [2]–[4]．これらの攻撃が対象とするモデル情報は，大量の訓練データを豊富な計算資源を用いて学習することによって得られるため，企業にとっては貴重な知的財産である．また，NN のモデル情報を用いることで，AI の判断を誤らせる Adversarial Examples（AEs [5]）が効率よく生成可能なことも知られている．そのため，NN のモデル情報を保護する対策手法の検討は非常に重要である．

そこで本研究では，MLP に対するサイドチャネルを用いたモデル抽出攻撃への対策手法について検討する．提案手法では，演算の順番をランダム化するシャッフリング手法を導入する．実験では，実デバイスを用いた評価を行い，提案手法の有効性について検証する．

２．MLP に対する電力サイドチャネルを使用したモデル抽出攻撃

ここでは，MLP に対する電力サイドチャネルを使用したモデル抽出攻撃 [4]について説明する．概要を図 1 に示す．電力解析では，MLP の乗算部分（図 1 の例では入力 x_0 と重み $w_{0,1}$ との乗算）に着目する．乗算の計算時には，レジスタの値が遷移し，この遷移による電力を消費する．そこで，電力解析ではこのレジスタ遷移と消費電力との相関を計算することで，重み値の復元を図る．具体的には，マイコンでは各計算前にレジスタ値が 0 に初期化されるため，レジスタ遷移は乗算計算時の値のハミング重み（Hamming Weight: HW）によって計算できる．この計算では，ニューロンへの入力を既知情報として扱い，重み値を未知のパラメータとする．未知の重みについては候補となる値を与えて，既知の入力との乗算を行い，この計算結果の HW を算出する．そして，計算した HW h と消費電力 w とのピアソンの相関係数 ρ を算出する．この計算式を以下に示す．

$$\rho = \frac{\sum_{n=1}^{D}(w_{n,t} - \overline{w_t})(h_n - \bar{h})}{\sqrt{\sum_{n=1}^{D}(w_{n,t} - \overline{w_t})^2}\sqrt{\sum_{n=1}^{D}(h_n - \bar{h})^2}} \quad (1)$$

ここで，D は使用する消費電力波形の数を，t は消費電力波形における時間軸を，$\overline{w_t}$は消費電力w_tの平均を，hは HW h の平均を示している．

この計算において，重みの候補値が正解に近い値の場合，消費電力と HW 間の線形相関から，相関係数にはピークが生じる．一方で，重みの候補値が間違っている場合は相関係数は小さくなる．そのため，相関係数を最大とする候補値を正しい重み値として推定する．

３．提案手法

本研究では，シャッフリング [6]による対策手法を提案する．提案手法の概要を図 2 に示す．電力サイドチャネルを使用したモデル抽出攻撃では，重み値と入力との乗算と，そのときの消費電力との相関を解析に利用している．そこで，図 2 に示すように計算毎に乗算の順序を入れ替えるシャッフリングを導入することで，時間軸上方向で消費電力をランダム化する．

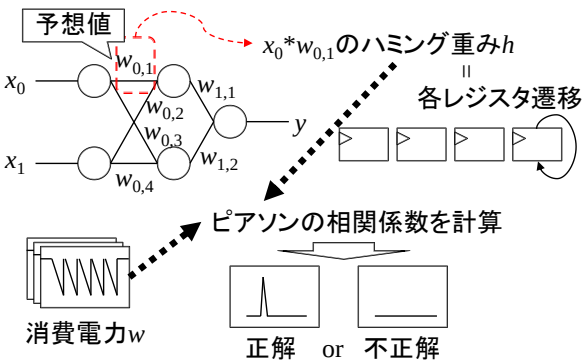

図 1　電力サイドチャネルを使用したモデル抽出攻撃の概要

４．評価実験

実験では，3 層の MLP を Arduino-UNO 上の Atmega328P に実装した．消費電力測定では，Atmega328P の GND 側に 51Ω のシャント抵抗を挿入し，シャント抵抗での電圧降下をオシロスコープによって測定した．使用したオシロスコープは，Agilent DSO 1024A であり，サンプリングレートは 2GSa/sec である．1,000 個の電圧波形を消費電力波形として取得した．また，実験では MLP の 1 つ目のニューロンの重み値を解析対象とした．

実験結果を図 3 と図 4 に示す．図の横軸は重みの候補値を，縦軸は相関係数を示している．また，正解の重み値を赤色で，不正解のものを青色でそれぞれ示している．まず，無対策の場合では，図 3 から正解の重みの候補値で相関係数が最大となっており，重み値の復元に成功していることが確認でき，MLP は電力解析に対して脆弱であることが分かる．

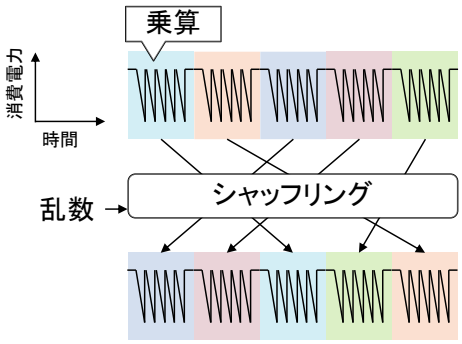

図 2　提案手法

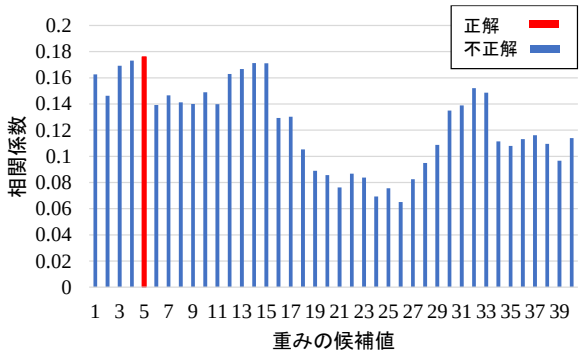

図 3　実験結果（無対策）

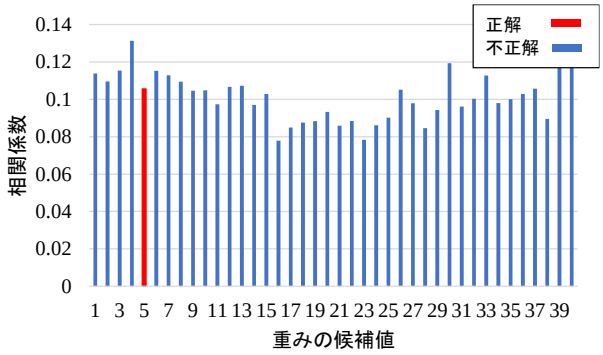

図 4　実験結果（提案手法）

次に，対策手法を施した場合では，図 4 に示すように正解の重みの候補値で相関係数が最大となっておらず，電力解析への耐性が向上していることが確認できる．これはシャッフリングによって，時間軸上での電力サイドチャネルがランダム化されたことによって，乗算時の HW と消費電力との相関が弱まったためだと考えられる．したがって，提案手法による対策が有効であることが分かる．

５．まとめ

本研究では，MLP への電力サイドチャネルを使用したモデル抽出攻撃への対策手法を提案した．提案手法では，ニューロンにおける乗算処理の順番のランダム化によって，時間軸上の電力サイドチャネル情報をランダム化した．評価実験では，Arduino 上のマイコンに MLP を実装し，攻撃耐性について評価した．その結果，シャッフリングを導入することで，電力サイドチャネル攻撃に対する耐性が向上することを明らかにした．

今後は，より多くの消費電力波形を使用した場合の耐タンパ性評価や，より複雑な NN に対する安全性評価を行っていく予定である．

謝辞

本研究の一部は，JSPS 科研費 19K24357 の助成を受けたものです．

参考文献

[1] 独立行政法人情報処理推進機構，“AI 白書 2020，”株式会社角川アスキー総合研究所，2020.

[2] V. Duddu, D. Samanta, D V. Rao, and V. E. Balas, “Stealing Neural Networks via Timing Side Channels,” arXiv:1812.11720, pp. 1–10, Jul. 2019.

[3] L. Batina, S. Bhasin, D. Jap, and S. Picek, “CSI Neural Network: Using Side-channels to Recover Your Artificial Neural Network Information,” arXiv:1810.09076, pp. 1–15, Oct. 2018.

[4] 野崎佑典，吉川雅弥，“ソフトウェア実装した MLP に対する電力解析の検討，”第 48 回東海ファジィ研究会予稿集，no. P1-05, pp. 1–3, Feb. 2020.

[5] E. Quiring, D. Klein, D. Arp, M. Johns, and K. Rieck, “Adversarial Preprocessing: Understanding and Preventing Image-Scaling Attacks in Machine Learning,” Proc. of 29th USENIX Security Symposium, pp. 1–18, Aug. 2020.

[6] T. S. Messerges, “Securing the AES Finalists Against Power Analysis Attacks,” Proc. of Int. Workshop on Fast Software Encryption (FSE 2000), LNCS vol. 1978, pp. 150–164, Springer, 2000.

連絡先

野崎佑典
E-mail: 143430019@ccalumni.meijo-u.ac.jp

XYXY 型オノマトペの音響特徴量を用いた日本語母音の音象徴に関する一考察

A Study on Sound Symbolism of Japanese Vowels using Acoustic Features of XYXY-Typed Onomatopoeia

◯ [1] 兼岩 尭希, [2] 中村 剛士, [3] 加納 政芳, [4] 山田 晃嗣, [1] 犬塚 信博

◯ [1]Takaki Kaneiwa, [2]Tsuyoshi Nakamura, [3]Masayoshi Kanoh, [4] Koji Yamada, [1]Nobuhiro Inuzuka

[1] 名古屋工業大学大学院

[1]Nagoya Institute of Technology

[2] 中部大学

[2]Chubu University

[3] 中京大学

[3]Chukyo University

[4] 情報科学芸術大学院大学

[4]Institute of Advanced Media Arts and Sciences

Abstract: Onomatopoeia is considerd to follow sound symbolism. Sound symbolism is a property that particular sound can give people particular impression. Urata et al. focused on sound symbolism and proposed an onomatopoeic thesaurus map. The map visualizes semantic similarity among onomatopoeias. Urata et al. hired small number of limited onomatopoeias to construct the map. In this paper, we reconstructed an onomatopoeic thesaurus map with many XYXY-typed onomatopoeias and verified whether the map can represent sound symbolism about Japanese voweles or not. Experimental result shows more than half of XYXY-typed onomatopoeias represent the sound symbolism about Japanese voweles.

1 はじめに

Human Robot Interaction では，人と機械の円滑な意思伝達が期待される．意思伝達において，音や物の動きといった感覚的な情報を容易に表現できるオノマトペは有益である．機械がオノマトペを理解することは容易ではない．オノマトペのデータベースは，日本語オノマトペ辞典 [1] 等からデータ収集することよって作成できるが，辞典に掲載されていない未知のオノマトペは数多く存在する．他方，人は未知のオノマトペの意味を音象徴 [3] によってある程度類推できる．音象徴とは，特定の音が特定のイメージを喚起させる性質であり，オノマトペは音象徴性を持つとされる．機械によるオノマトペの意味類推は，人と機械の間の自然な意思伝達の実現に大きく貢献するのではないかと考えられる．

機械によるオノマトペの意味類推に関する研究報告がいくつかある．浦田ら [2] は音象徴に着目し，オノマトペの意味的類似関係を可視化するオノマトペ・シソーラス・マップを提案した．兼岩ら [4] は，浦田らの構築したマップの音象徴の表現能力を篠原 [5] の報告により検証したが，結果の一部は仮説を支持しなかった．浦田らや兼岩らは，マップの構築に特定の用法の少数のオノマトペのみを用いていた．そのため，マップ上に出現する音響情報が偏り，マップが表現する音象徴も偏った可能性がある．

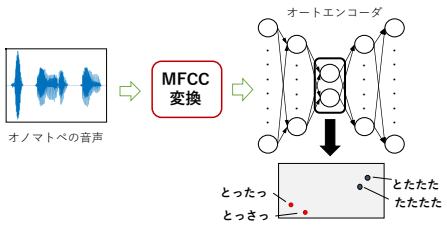

図 1: オノマトペ・シソーラス・マップの概要

本研究では，オノマトペの型として頻出する XYXY 型のオノマトペを網羅的に作成し，音響情報の偏りが少ないマップを構築した．本稿は，この再構築したマップについて，兼岩ら [4] が行った音象徴の検証を改めて行った．

2 オノマトペ・シソーラス・マップ

浦田ら [2] の提案したオノマトペ・シソーラス・マップの概要を図 1 に示す．マップはオートエンコーダ中間層の出力ベクトルによって構成する．

本研究では，オートエンコーダの訓練のため，XYXY 型オノマトペを網羅的に活用した．X には「ン (/N/)」を除く日本語のモーラ 126 種，Y には日本語のモーラ 127 種を用い，合計 16,002 個のオノマトペを作成した．MFCC 特徴量は，Google Cloud の Text-to-Speech[6] によって合成したオノマトペの音声波形から計算した．ここで，i 番目のオノマトペ O_i を式 (1) で表現する．$c^i_{\mathrm{X}}, v^i_{\mathrm{X}}$ は O_i の X の子音と母音，$c^i_{\mathrm{Y}}, v^i_{\mathrm{Y}}$ は Y の子音と母音を表す．

$$O_i = o(c^i_{\mathrm{X}}, v^i_{\mathrm{X}}, c^i_{\mathrm{Y}}, v^i_{\mathrm{Y}}) \tag{1}$$

$$c^i_{\mathrm{X}} \in \boldsymbol{C} = \{-, /\mathrm{k}/, /\mathrm{s}/, \cdots, /\mathrm{by}/, /\mathrm{py}/\}$$
$$c^i_{\mathrm{Y}} \in \boldsymbol{C} \cup \{/\mathrm{N}/\}$$
$$v^i_{\mathrm{X}}, v^i_{\mathrm{Y}} \in \boldsymbol{V} = \{/\mathrm{a}/, /\mathrm{i}/, /\mathrm{u}/, /\mathrm{e}/, /\mathrm{o}/\}$$

3 実験

構築したマップについて，篠原 [5] の報告にある日本語母音に関する音象徴を検証した．検証では，マップ上のオノマトペ間の距離を用いて定量的に評価した．篠原の報告では，日本語の母音のうち，/a/，/u/，/o/が/i/，/e/よりも

大きなイメージを与えるとされる．これを，/a/，/u/，/o/は大きいイメージ群に属し，/i/，/e/は小さいイメージ群に属すると解釈する．ここで，O_i と O_j があり，O_i と O_j は v^i_{X} と v^j_{X} のみ異なる，もしくは v^i_{Y} と v^j_{Y} のみ異なるものとする．マップが音象徴を表現するならば，ある O_i の v^i_{X} に対して v^i_{X} と同じイメージ群に属する v^j_{X} を持つ O_j を O_j^{same}，異なるイメージ群に属する場合の O_j を O_j^{diff} と表すと，以下の式 (2) を満足すると考えられる．

$$d(O_i, O_j^{diff}) > d(O_i, O_j^{same}) \tag{2}$$

ここで，$d(A, B)$ はマップ上でのオノマトペ A と B の間のユークリッド距離を表す．また，式 (2) は，v^i_{Y} と v^j_{Y} のみ異なる O_i と O_j でも同様に満足すると考えられる．

そこで，今回作成した XYXY 型オノマトペが式 (2) の関係を満足することを実験仮説とした．すなわち，v^i_{X} と v^j_{X} に注目し，ある O_i に対して四つの O_j すべてが式 (2) の関係を満たすか否かについて検証した．v^i_{Y} と v^j_{Y} に注目した場合についても同様に検証した．これを，c^i_{X} または c^i_{Y} が/y/，/w/，拗音の場合を除く全ての O_i で実施した．これは，v^i_{X}，v^i_{Y} が取りうる母音が五つでなく，他との比較が困難なためである．

4　実験結果・考察

実験の結果，9,340 個の O_i が仮説に従い，7,105 個の O_i が仮説に従わなかった．過半数の O_i は仮説に従うが，仮説に従うとは言えない傾向を示す O_i も数多く存在した．

そこで，仮説を支持するか否かを分ける要因を探るため，オノマトペ間の距離について詳細を調べた．O_i と四つのオノマトペ O_{j_k} は，v^i_{X}，$v^{j_1}_{\mathrm{X}}$，$v^{j_2}_{\mathrm{X}}$，$v^{j_3}_{\mathrm{X}}$，$v^{j_4}_{\mathrm{X}}$ のみ互いに異なるものとする．ここで，O_i に対する四つの $d(O_i, O_{j_k})$ を算出し，昇順に 1 位から 4 位の順位を付けた．v^i_{Y} と $v^{j_k}_{\mathrm{Y}}$ のみ異なる O_i と O_{j_k} についても同様に行い，その平均順位を v^i_{X} または v^i_{Y} ごとに算出した．

算出した平均順位を表 1 に示す．v^i_{X} または v^i_{Y} が/a/，/i/，/o/の場合は実験仮説に従う傾向の順位となった．一方，v^i_{X} または v^i_{Y} が/u/，/e/の場合はその傾向が見られなかった．最も平均順位が高い母音は/u/の場合/e/であり，/e/の場合/u/となるが，/u/と/e/は異なるイメージ群に属する．

/u/，/e/の群については，大きいイメージを与える/a/，/o/の群と小さいイメージを与える/i/の中間のイメージを与えることを，今回のマップは表現していると考えられる．これは兼岩ら [4] の報告と一致する．しかし，これが人がオノマトペから受ける大きさの印象と一致するかは未検証であり，主観評価実験等で検証する必要がある．

表 1: O_i に対する O_j のマップ上のユークリッド距離の平均順位

		$v^{j_k}_{\mathrm{X}}$ または $v^{j_k}_{\mathrm{Y}}$				
		/a/	/i/	/u/	/e/	/o/
v^i_{X} または v^i_{Y}	/a/		3.96	1.59	2.87	1.58
	/i/	2.97		1.98	1.07	3.98
	/u/	1.82	3.88		1.29	3.02
	/e/	2.08	3.35	1.11		3.46
	/o/	1.03	3.98	2.02	2.97	

5　まとめ

本研究では，浦田ら [2] が提案したオノマトペ・シソーラス・マップを XYXY 型オノマトペを網羅的に用いて構築した．この音響情報の偏りの少ないマップを用いて，兼岩ら [4] が行った篠原 [5] の示した母音に関する音象徴による検証を改めて行った．実験結果の一部は篠原の示した音象徴に従うが，一部は異なる傾向を示した．分析の結果，母音の与える大きさのイメージは明確に 2 クラスに分かれるのではなく，中間クラスが存在する可能性がある．今後，この関係について更なる検証を行いたい．

また，本実験では母音による大きさの印象のみを用いて検証したが，篠原の報告は子音のものも含め他にも存在する．今後，これらによる検証を行い，マップの表現能力と音象徴そのものを定量的に明らかにしたい．

参考文献

[1] 小野: 擬音語・擬態語 4500 日本オノマトペ辞典, 小学館, 2007

[2] Urata et al.: Visualized Onomatopoeia Thesaurus Maps Based on Deep Autoencoder, SCIS & ISIS, pp. 215-219, 2018

[3] Hinton et al.: Sound symbolism, Cambridge University Press, 2006

[4] 兼岩他: オノマトペ・シソーラス・マップを用いた音象徴性の可視化に関する一検証, 知能と情報, Vol. 32, No. 5, pp. 907-911, 2020

[5] 篠原: オノマトペと認知科学：実験をもちいた音象徴の基礎研究, 日本語学, Vol. 34, No. 11, pp. 44-54, 2015

[6] Text-to-Speech: 自然な音声合成 — Google Cloud: https://cloud.google.com/text-to-speech/?hl=ja

連絡先

兼岩 尭希

E-mail: t.kaneiwa.747@stn.nitech.ac.jp

FCR 法応答の統合値に関する新提案
—オーケストラ音響刺激の認知を実験課題としたデータの処理—

A New Model for Integrating a Response Data Set of the FCR-methods: Processing of Experimental Data to Recognize the Sound Flow of a Symphony

○小田　哲久[1]，河辺義信[1]，山田　洋巳[2]
○Tetsuhisa Oda[1], Yoshinobu Kawabe[1], Hiromi Yamada[2]
1 愛知工業大学，2 京都西山短期大学
1 Aichi Institute of Technology, 2 Kyoto Seizan College

１．はじめに

小田の提案した FCR 法（ファジィ多項目並列評定法）は，２項目並列尺度の利用が普通であり，その場合は，各種の統合アルゴリズムが提案され，対象によって公式を変えることも可能である．(統合値が一位に定まらないのは，不便とも言える)

一方，応用法のバリエーションとして，項目数を３項目以上にする方法も提案されている．しかし，その場合，利用できる統合式としては，ファジィ推論の重心法以外は提案されていない．

本研究では，ある対象の主観的評定に対して，５つの設問の全てに答えさせるような形式を採用した課題について，それを FCR 法の課題と捉えて，実際の実験結果に適用した場合の分析結果を示す．

今回の課題では，５つの設問のうちの２問は，ある基準では，同じ評価傾向を示すので，OR 演算（Max 演算）をすることで一つに集約することで，結果的に４項目法に帰着させている．

ここでの提案は，４項目 FCR 法応答値(a,b,c,x)から，新統合値 y を求める過程で，重心法ではなく，Mode-Median 法を採用することである．但し，データの処理は．最終的に元の項目{A,B,C,X}にデファジィ（非ファジィ化）した結果を求める．

実験は，音響刺激に対する被験者応答による．具体的には，シューベルト作曲，交響曲第７番（これまでは８番とされていた)「未完成交響曲」の，特定レコード，カールシューリヒト指揮，ウィーンフィルハーモニー交響楽団による録音を使用．その演奏の中で，「第一楽章展開部で，人によっては鐘の音のように聴こえるパート」を聴かせて，それがどう聴こえたか，聴こえ方に関する５つの文章を言語刺激とし，それぞれに評定尺度を付して応答させた．但し，５項目全てに，該当する度合いを応答させた．その際，矛盾した応答も排除しない事をあらかじめ教示した．また，足して１になるような，いわゆる定和尺度ではないことも，良く教示しておいた．

測定結果は，まず，重心法による統合値を計算し，それをさらに{A,B,C,X}にデファジィした．その操作を，ここでは，G+Defuzzy と呼ぶことにする．

次に，新しい試みとして，一旦，Mode-Median 処理で複数のパターンに集約してから，さらにデファジィして，{A,B,C,X} に戻す．

この新旧２つの方法による分析結果を比較して，新技法の特徴を探る．

２．実験

被験者：愛工大の情報科学部学生，１３０名
教示を理解できていなかった２名分をスクリーニングで除去．
環境：一般教室の PA 装置を使用．
刺激：大学図書館の契約している，音楽配信サービス（NAXOS ミュージックライブラリ）を利用．
応答方法：Moodle による．
教示と手続き：
（１)「これまで，何か楽器を学んだことがありますか？」などの音楽経験を聞いた．(問１から問８）
（２）楽曲を流したのち，五つの文章に対する印象（その文の内容が，音楽を聴いて自分が感じた印象と，どの程度合致しているかを答えてください．「1 （まったくそうは思わない)」から「9 （完全にそのとおりだと思う)」まで．)
(A) 教会の鐘の音が聞こえた．
(B) オーケストラの音だが「鐘の音に似た音」も聴こえた．
(C)オーケストラの音で，そこに「あえてそう言われれば，教会の鐘の音に聴こえなくもないような音」が聴こえた．
(D) オーケストラの音として聴こえ，鐘の音らしき音は聴こえない．音は「個々の楽器の音が分離して」聴こえる．
(E) オーケストラの音として聴こえ，鐘の音らし

き音は聴こえない．音は「ぼやっとした塊の音色(いわゆる，ハーモニー)」として聴こえる．

３．分析１（重心法による）

各被験者の応答結果は，５次元情報（a，b，c，d，e）として得られる．（ a,b,c,d,e∈[0,1] ）しかし，D と E は，理由は異なるけれども，「鐘の音のようには聴こえない」，という点では同じなので，X ＝D∨E とし，値は，x＝ max(d,e)とする．こうして，４次元情報(a，b，c，x）に集約される．
これは，FCR 法としては異例の前処理である．
そして，A，B，C，X には，それぞれスコア約１点，約 2/3 点，約 1/3 点，約０点，を付与する．（鐘の音に聴こえる程度を等間隔の三角ファジィ数に数値化した．三角ファジィ数は左右対称で底辺長は 1/3 の,いわゆるファジィ分割である.）
重心法による統合式は次の通りで，要するに，重みつき平均になっている．

$$G = \frac{a+\frac{2}{3}\times b+\frac{1}{3}\times c}{a+b+c+x}$$

統合値 G は，[0,1]の実数であり，まだ曖昧さを含んでいる．ここで，Defuzzy 操作を適用する事により，カテゴリー{A,B,C,D,X} のいずれに該当するかを強制的に決めて，各カテゴリーの頻度を求めることができる．
Defuzzy 操作の閾値は，スコア設定の時の逆で，下記のように定める．
(G≦1/6)→X ，
(1/6<G≦1/2)→C ，
(1/2<G≦5/6)→B ，
(5/6<G≦1)→A
結果は次の通り．

A の人数	0	0.00%
B の人数	32	25.00%
C の人数	57	44.53%
X の人数	39	30.47%
合計	128	100%

４．分析法２（Mode-Median 法による）

各被験者応答において最大値をとる刺激値（Mode）で分類する方法は，単純だが，案外有効かもしれない．その方式の問題点は，２つ以上の最大値がある場合に,Mode 値をどこに決めるべきか困る点である．そこで，刺激に順序があることを利用し，中央値(Median)で決めるようにする．刺激項目の中間値が出現するが，それは，均等配分で処理することにする．（均等にできない場合が出現しうるので，その場合は，あらかじめ決めておいた方式で配分する）また，実際の測定結果には，凸でない状態（複数のピークの間に，ピークよりも低い値が出る事）も出現しうるので，あらかじめ，その場合の対策を指定（定義）しておく必要がある．なお，ピークの外側において非凸な観測値があっても，無視することにする．

鐘の音に聴こえる順序は，X＜C＜B＜A である．
聴こえない順位は，A ＜ B ＜ C ＜ X である．
実際の分析では，これらのどちらかを使う
重心法の場合と同様に，等間隔ファジィ数を仮定する．
2 項目が同値の場合，仮に A と B に同じ最大値が出現した場合，(A,B) と表記する．
今回の課題で出現するのは，
(X,C)，　(X,B) → C，　(X,A) → (C,B)，
(C,B)，　(C,A) → B
(B,A)
ここで，矢印→の右側は，Median が直ちに求まるパターン．または，より近接した項目パターンに集約できるパターン．
３項目が同値では，
(X,C,B) → C，(X,B,A) → B，(C,B,A) → B
4 項目が同じ値のパターンは，
(X,C,B,A) → (C,B)
この変換表の中で，(X,B,A) → B だけは，やや不自然である．しかし，右辺を(C+B)/2 としたら，C が Mode 値でないので,さらに不自然であろう，

以下，実際の結果を「聴こえない順位」で分析する．
Mode-Median 法による分類結果の頻度は，
counts:

(A,B)	(B,C)	(C,X)	A	B	C	X
2	5	7	6	17	32	59

percentages:

(A,B)	(B,C)	(C,X)	A	B	C	X
1.56	3.91	5.47	4.69	13.28	25.00	46.09

なお，(A,B,C)など，Median で集約できる結果（先に示した変換表の矢印表記）は，そのように処理してある．

この方法では，聴こえない人(X)が約４６％で，残りの約５４％が何らかの形で鐘の音のような音響を聴いていることになる．
G+Defuzzy の結果よりは，X の比率が上がるが，一方で，G+Defuzzy 結果では見られなかった A の割合も 4.69%が確認される． A と B の中間が 1.56%，後者を A と B に半々に振り分けたとして，4.69%+0.78%=5.47%が A となる．

ここまでの段階では，複数の Mode 値を持つ被験者がそのままで計上されていて，曖昧さが残っている．そこで，例えば，(A,B)なら，A へ 1/2，B へ

1/2 という形で被験者数を分配する．それによって，{A, B, C, X} の頻度がわかる．

以下，(B,C),(C,X)，を分配すると，
(B,C)の 3.91%は，1.955%の B と，1.955%の C に分配．
(C,X)の 5.47%は，2.735%の C と 2.735%の X に分配，
今回の実験では，分配が必要となるような応答パターンは，以上の２種で全てであった．

以上をまとめてみる．
A=5.47%
B=13.28% + 0.78% +1.955% =16.015%
C=25.0% + 1.955% + 2.755%=29.71%
X=46.0% + 2.735%=48.735%
丸め誤差で，合計が１００%にならないので，正確に計算すると，次の表１のようになる．

表１　Mode-Median 法の中間結果

	Count	%	左の 1/2
AB	2	1.5625%	0.7813%
BC	5	3.9063%	1.9531%
CX	7	5.4688%	2.7344%
A	6	4.6875%	
B	17	13.2813%	
C	32	25.0000%	
X	59	46.0938%	
Total	128	100.0000%	

２項目にピークのある応答を，各項目に分配すると，表２のようになる．

表２　Mode-Median 法の結果を分配した結果

	本体	1/2（1）	1/2（2）	合計
A	4.6875%	0.7813%		5.4688%
B	13.2813%	0.7813%	1.9531%	16.0156%
C	25.0000%	1.9531%	2.7344%	29.6875%
X	46.0938%	2.7344%		48.8281%
Total				1

表２の合計欄の小数点以下の桁数を少なく表示．

表３　Mode-Median 法の分配結果

	合計
A	5.5%
B	16.0%
C	29.7%
X	48.8%

この分析方法は，項目軸方向の Median をとっているので，伝統的な分析法とはいえないだろう．計算順序に従って，Mode-Median 法としておく，
(Min-Max-重心法の命名法から類推)
（ファジィ積分にも似ている）

５．新旧両技法による処理結果の比較

重心法の結果 G を Defuzzy した場合と，この Mode-Median 法の結果を比較してみる．（表４）

表４　G+Defuzzy 法と Mode-Median 法の比較

	Gravity-Center+ Defuzzy		Mode-Median	
	Count	%	Count	%
A	0	0.0000%	7	5.4688%
B	32	25.0000%	20.5	16.0156%
C	57	44.5313%	38	29.6875%
X	39	30.4688%	62.5	48.8281%
Total	128	100%	128	100%

G+Defuzzy 法と比べて，Mode-Median 法の方が，X に分類される応答（被験者）が増えて，全体の約半数となった．Mode-Median 法では，Mode や Median といった演算で，細かい情報をバッサリと切り落としているために，微妙な感覚の差異が結果に反映されず，こういう結果が得られたのではないか，と思う．
それでも，約半数の被験者は，「鐘のような音響が聴こえる」と反応したことがわかる．

６．結び

従来法の G+Defuzzy 操作と，新提案の Mode-Median 操作の結果を比較すると，両者は同じにはならず，それぞれ特徴が異なり，課題を異なる視点から観察した結果となる．

Mode-Median 法は，一般化してファジィ推論の出力ファジィ集合の処理にも使う事が可能か，検討に値すると考えられる．

謝辞：本研究は，科研費・基盤 C(20K11822）の助成を受けて行われた．また，多くの被験者各位，音楽関係者，音響学の研究者各位，並びに，知人諸氏に協力いただいた．諸氏に深謝します．

参考文献

[1] 小田 哲久，「ファジィ多項目並列評定法の特徴について」，日本経営システム学会誌 12(1), pp.23-32, 1995
[2]小田 哲久 , 林 文俊,「FCR 法の心理測定への応用について：新しい定義と実践手続きの提案」, 日本経営工学会論文誌 52(6), pp.396-406, 2002
[3] 小田哲久，林文俊，「心理測定におけるファジ

ィ性」, ファジィとソフトコンピューティングハンドブック, 日本ファジィ学会編, 共立出版, pp.889-886, 2000
[4]小田哲久,「評定尺度法へのファジィ論的アプローチ: FCR 法,IR 法について」,心理学評論, Vol.56, No.1, pp.67-83, 2013
[5] 小田 哲久,「特集③人間工学のための計測手法　第 3 部：心理計測と解析（3）:ー感性評価による計測と解析ー」, 人間工学 51(5), pp.293-303, 2015
[6]高萩 栄一郎, 李 彦雯, 小田 哲久, 「集合関数による並列評定値（FCR 法）の分析:分析手法の提案」, 知能と情報 28(4), pp.723-733, 2016

付録　数値例

処理過程を理解しやすくする目的で，今回の測定結果から, ID 番号の若い順に 5 人分の応答を抜き出して，処理過程をモデル的に示す.
（ID 番号は被験者にランダムに付与されている）

（1） 元データ（9 段階）

ID	A	B	C	D	E
1	2	4	5	5	6
2	7	9	7	4	8
3	3	5	7	2	7
4	1	1	1	1	3
5	5	6	8	2	1

（2） D,E から X への変換（X=D∨E）

ID	A	B	C	X
1	2	4	5	5
2	7	9	7	4
3	3	5	7	2
4	1	1	1	1
5	5	6	8	2

（3） 重心法とその Defuzzy 結果
9 段階のデータを[0,1]データに直して, 重心 G を計算する. さらに Defuzzy して{A,B,C,X}に戻す.

ID	a	b	c	x	G	Defuzzy
1	0.13	0.375	0.5	0.63	0.333	C
2	0.75	1	0.75	0.88	0.494	C
3	0.25	0.5	0.75	0.75	0.370	C
4	0	0	0	0.25	0.000	X
5	0.5	0.625	0.88	0.13	0.569	B

（4） 4 項目に縮約，Mode-Median 法

ID	A	B	C	X	Mode	Mode-Median
1	2	4	5	**6**	X	X
2	7	**9**	7	8	B	B
3	3	5	**7**	**7**	(C,X)	C,X へ各 1/2
4	1	1	1	**3**	X	X
5	5	6	**8**	1	C	C

（5） Mode-Median 法の分配結果

ID	A	B	C	X
1				1
2		1		
3			0.5	0.5
4				1
5			1	

少数言語保存活動の支援
録音信号に対するノイズ除去に関する一検討

木村圭吾，松下真也，○高瀬治彦，北英彦
三重大学

１．はじめに

現在，少数言語の保護活動が言語学者らによって行われている．少数言語は　，使用者が少なく絶滅を危惧されている言語であり，独自の文字体系を持たず音声だけの言語も少なからず存在する．そのため保護活動は，音声を収集するフィールドワーク・収集した音声の体系化・語られている内容の理解などのステップをふむ．フィールドワークでは，対象言語の使用者と現地で数か月間の共同生活を行い，資料や史料，音声データの録音等を行う．収集した音声の体系化では，国際音声記号（IPA: International Phonetic Alphabet）等を用いた書き起こし，単語・文法の認識を行う．最後に，語られている内容の理解・記録を行う．

本研究では，保護活動における音声の体系化に着目する．音声の体系化では，(1) IPA 等を使った書き起こし，(2) 単語の切り出し，(3) 文法の理解の手順をふむ．(1) では，音声認識の技術を利用できる．IPA は，音声学から「あらゆる言語の発音を説明することが可能な記号」[1]と定義されており，既知・未知に関わらず言語として発音された音声を記録するのに適している．(2) では形態素解析などの技術を，(3) では構文解析などの技術が利用できる．

本稿では，録音した音声を，IPA を用いて書き起こすための前処理としてのノイズ除去に着目し，その処理が書き起こしに与える影響について簡単な調査をしたので報告する．

２．書き起こし作業とノイズ除去

先に述べたように，少数言語の保護の第 1 段階として，フィールドワークを通じて音声情報を収集する．このフィールドワークは，音響的に管理された録音スタジオで行われることはまれで，日常の生活環境下で行われる．そのため，定常的に発生するノイズだけでなく，環境音の混入もある．そのため，ノイズ除去でなく，音声分離技術の適用が望ましい．

また，一般的な音声認識技術では，音声情報だけでなく，さまざまな事前知識を利用することで，認識率を高めている．しかし，少数言語の保護活動においては，十分な事前知識が無いため，音声情報だけから，書き起こしをする必要がある．

これらの状況から，音声分離を行った後の音声が，音声認識に必要な特徴量を維持しているかどうかを検討する必要がある．

３．音声分離技術

音声分離技術は，大きく複数チャネルを利用した方法と単一チャネルによる方法に分かれる．複数チャネルを利用する方法では，マイクを複数使用し，それらから音の空間的特徴を得ることで，音声分離を行う．しかし，少数言語の保護におけるフィールドワークでは，録音環境を整えるのが難しいため，ここでは考慮に入れない．単一チャネルによる方法では，単一のマイクから録音した情報から，周波数的な特徴に基づいて音声分離を行う．ここでは，これに基づく手法を 2 つ取り上げ，分離処理が音声認識に与える影響を議論する．

本稿で取り上げる音声分離手法の一つ目は，Open-unmix である．これは，SONY と INRIA（Institut National de Recherche en Informatique et en Automatique）により開発された，深層学習を用いた音源分離手法である[2]．これにより，音楽のボーカル・ベース・ドラム・その他伴奏を分離することができる．

本稿で取り上げる音声分離手法の二つ目は，Spleeter である．これは，フランスの楽曲配信サイトを運営する Deezer が開発した深層学習を用いた音源分離手法である[3]．これにより，音楽のボーカル・ベース・ドラム・ピアノ・その他伴奏を分離することができる．

いずれも，音楽からのボーカル分離という意味では音声分離が可能である．実際，分離した結果は，人の耳では音声を分離できているように聞こえるが，音声認識にかけた場合にどうなるかは不明である．

４．音声分離した音声の認識実験

この章では，クリアな音声にノイズを加え，それから 2 種類の音声分離方法により分離した音声を認識し，その結果を比較する．

認識対象の音声には，音声資源コンソーシアムにより提供されている音声コーパスの日本語音声を用いた．音声は，152 のファイルに分かれており，いずれも 10～40 の音素を含んでいた．加えるノイズの SN 比は 10dB として，実験を行った．

音声分離の影響を評価するために，ノイズ印加前・後の音声それぞれを，単語辞書・言語モデルを空にした Julius[4]を用いて認識した．日本語音声を用いているが，単語辞書・言語モデルを空にしているので，未知の言語を対象としている状況と見なすことができる．結果は，式(1)の一致率により，評価する．ここで，$S1$ はノイズ印加前の音声の認識結果（文字数は$|S1|$），$S2$ は ノイズ印加後の音声の認識結果（文字数は$|S2|$），これらが一致した文字数を Kとする．

$$一致率 = \frac{2K}{|S1| + |S2|} \tag{1}$$

図 2 にノイズを印加した場合の結果を示す．この図は， 152 個のファイルそれぞれに対して式(1)により求めた一致率のヒストグラムである．この図では，Spleeter，open-unmix それぞれにより音声分離した結果を示している．この結果から，二つの音声分離技術によるヒストグラムの大きな差は見いだ

せなかった．いずれも，ほとんどのファイルで一致率が 50%を下回っており，ノイズ印加・音声分離の工程を経ることで，音声認識に重篤な影響を与えているといえる．

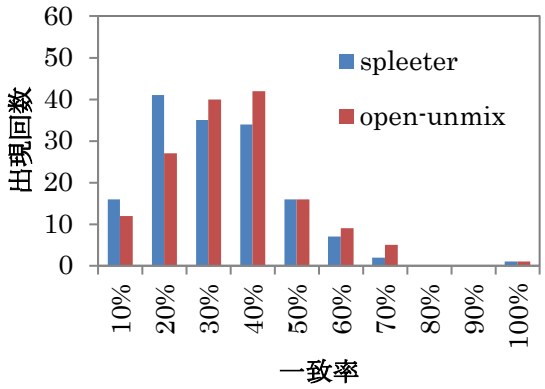

図 2　さまざまな音の周波数帯

5．スペクトログラムによる比較

この章では，ノイズ印加・音声分離による音素の特徴量の変化を議論する．ここでは，スペクトログラムと呼ばれる周波数スペクトルの時間変化を表した図を比較する．

図 3 にノイズ印加前の信号，図 4 にノイズ印加後 Spleeter により音声分離した信号に対するスペクトログラムを示す．いずれの図も横軸が時刻，縦軸が周波数を表し，各点の色でその時刻のその周波数成分の強度を表している．これらを比較すると，音声分離後は，ノイズ印加前と比べ高周波成分の強度が減少していることが分かる．音声認識には，フォルマント周波数と呼ばれる周波数の信号強度が重要である．特に母音の明瞭さには周波数が低い方から 3 個分のフォルマント周波数（F1, F2, F3）に対応する信号の強度が強く関わっている[5]．この点に留意して，図 3, 4 を比較する．時刻 0.5 近辺におけるフォルマント周波数は，F1=300Hz，F2=2.4kHz, F3=3.2kHz 程度である．音声分離に伴い，フォルマント周波数近辺の信号強度は弱くなっている．特に F2, F3 近辺では，フォルマント周波数から外れている部分の信号が強く減衰していることが確認できる．人の耳には，F1, F2, F3 に対応する信号が残っているため，音声として聞き取れるが，音声認識においては，これが悪影響を与えているのではないかと推察される．

4 章の実験結果と併せると，ノイズ印加・音声分離により，音声信号の高周波数成分が除去され，音声認識に利用される特徴量が大きく変化したことにより，音声認識処理に重篤な影響を与えたと考えられる．

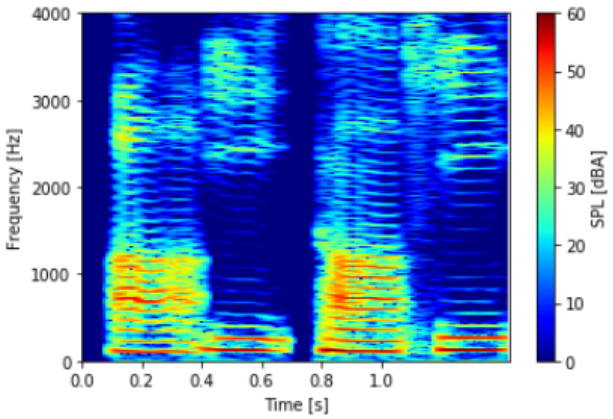

図 3　ノイズ印加前の信号のスペクトログラム

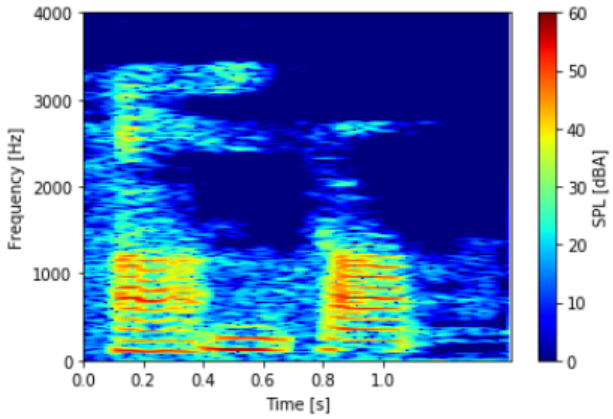

図 4　音声分離後の信号のスペクトログラム

6．まとめ

本稿では，少数言語の保護における，フィールドワークにおいて収集した音声情報の書き起こし作業に関して，ノイズ除去処理について検討した．具体的には，Spleeter および open-unmix といった音声分離技術をとりあげ，これらによる処理が書き起こし作業に与える影響について評価した．実験の結果，音声分離は，音声認識処理に対して重篤な影響が与えられることが示された．

参考文献

[1] 河原繁人，ビジュアル音声学，三省堂, 2018
[2] Fabian-Robert Stöter, et. al., open.unmix.app - towards audio separation on the edge, Proceedings of 2021 Web Audio Conference, 2019
[3] Romain Hennequin, et. al., SPLEETER: A fast and state-of-the art music source separation tool with pre-trained models, The International Society for Music Information Retrieval Conference (ISMIR), 2019
[4] A. Lee, et.al., Julius v4.5, https://doi.org/10.5281/zenodo.2530395, 2019
[5] 柘植佑好 他，母音のフォルマント周波数と発語明瞭度との関係について，日本口腔科学会雑誌, 29 巻, 2 号, pp. 268–271, 1980