

令和 5 年 5 月 24 日現在

機関番号：21602

研究種目：基盤研究(C) (一般)

研究期間：2020～2022

課題番号：20K11956

研究課題名(和文) PSYPHON: Psychoacoustic features for Phonation prediction

研究課題名(英文) PSYPHON: Psychoacoustic features for Phonation prediction

研究代表者

Julian Villegas (Villegas, Julian)

会津大学・コンピュータ理工学部・上級准教授

研究者番号：50706281

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：サボテカ語とミキセ語の単語を収録した新しいコーパスを作成し、3種類の音韻的対比(有声・きしみ声・息もれ声)で特徴づけました。このコーパスは機械学習の音韻予測に役立ちます。主観的実験では、きしみ声に対する感度が母語話者や非専門家の非母語話者を上回りました。この結果は、音響心理学的特徴が発声の知覚を予測する因子として優れていることを支持しました。ファルセットはピッチ、ささやき声は鋭さ、きしみ音は音量と荒さに関連していることも明らかにしました。最後に、主観的研究の再分析から音響心理学的粗さモデルを開発しました。

研究成果の学術的意義や社会的意義

This project is significant because it helps to close the gap between languages that are under-resourced and those that have sufficient resources, contributing directly to the Sustainable Development Goal SDG-10 (reducing inequality within and among countries).

研究成果の概要(英文)：We were able to document and acquire new corpora of words from the Zapotec and Mixe languages. These languages are characterized by a three-way phonemic contrast (modal, creaky, and breathy). These corpora are valuable for training phonation prediction systems based on machine learning. By subjective experimentation, we found that the sensitivity to creakiness observed in classifications made by experts and machine learning systems based on these classifications surpassed that of native and naive listeners. This finding supports our hypothesis that psychoacoustic features, which are universal, are better predictors of perceived phonation compared to existing methods. In addition, we found that falsetto was associated with pitch, whispering with sharpness, and creakiness with loudness and roughness. Lastly, by re-analyzing previous subjective studies, we were able to develop a psychoacoustic roughness model based on machine learning techniques.

研究分野：Computer Science

キーワード：Phonation prediction Machine Learning Psychoacoustics

## 1. 研究開始当初の背景

PSYPHON (Psychoacoustic features for phonation prediction) is a multidisciplinary project that aims to improve the automatic detection of phonemic contrast through the use of psychoacoustic models for better speech recognition systems. Psychoacoustic models establish the nonlinear relationships between physical quantities (such as frequency and pressure level) and sensory counterparts (such as pitch and loudness). These models also provide units (such as mel and sone), scales, and Just-Noticeable Differences (JNDs) that allow for consistent comparison of diverse phenomena.

Typically, researchers utilize prediction models that relate phonation to acoustic features like spectral tilt (the slope of speech's power spectral density) or articulatory features (e.g., open quotient: the fraction of a glottal cycle during which the glottis is open). However, there are numerous ways in which speakers can produce a given phonation. Articulation variations correlate differently with acoustic/physiological features, and the performance of these predictors varies depending on the predominant articulation within a group of speakers. This variation makes it challenging to compare across languages, as researchers often choose models that best suit their needs. While speakers using phonemic contrast may differ physiologically in how they produce it, their interlocutors are ultimately able to distinguish different phonations when necessary. Thus, perceptual features of speech at the listener's end, as opposed to unprocessed acoustic or physiological features at the speaker's end, could be suitable for predicting phonation.

## 2. 研究の目的

The objectives of PSYPHON are twofold: 1) to create a psychoacoustic-based classifier of phonation for digitized speech, and 2) to elucidate correspondences between phonation and psychoacoustic features. We have identified three main challenges in achieving these goals:

A. Lack of high-quality speech corpora: The performance of prediction models greatly depends on the training data. Some available corpora have issues, such as using dictionary entries to determine the phonation of a word without verifying that the actual production corresponds to the intended one. Others have recording issues, where parts of speech are not audible or are masked by noise artifacts.

B. Inadequate psychoacoustic models: While many psychoacoustic models are standardized, some implementations use outdated auditory models, resulting in predictions that are not always accurate. These models are built upon carefully collected subjective data, which can be used to create new prediction models through Machine Learning (ML) techniques.

C. Identification of the main predicting features: While good quality corpora are needed for predicting phonation in principle, the use of psychoacoustic features helps explain the underlying causes of a given classification. However, including all these features is not sufficient to ensure explainability. Therefore, ablation studies (where a set of features is excluded, and the training of a predictor is repeated) are needed to elucidate any correspondence between psychoacoustics and phonation.

### 3 . 研究の方法

First, we conducted an extensive literature review on Zapotec and related languages that are characterized by a three-way phonemic contrast (e.g., normal, creaky, and breathy). We then performed field recordings of Zapotec and Mixe languages in Oaxaca, Mexico. In parallel, we evaluated psychoacoustic models by 1) identifying those from which we could extract features and 2) compiling subjective results from published studies. Additionally, we conducted subjective experiments to compare how naive, native, and expert classifications are related. Several experiments were also conducted to fine-tune the hyperparameters of the machine learning algorithms.

### 4 . 研究成果

- S.Yoshida, J. Villegas. "Development of a psychoacoustic roughness model using machine learning," Master's thesis, University of Aizu, Mar. 2024.
- J. Villegas, K. Akita, and S. Kawahara, "Psychoacoustic features explain subjective size and shape ratings of pseudo-words," in Proc. of Forum Acusticum, the 10 Conv. of the European Acoust. Assoc., (Turin, Italy), Sep. 2023.
- J. Villegas, S. J. Lee, J. Perkins, and K. Markov, "Psychoacoustic features explain creakiness classifications made by naive and non-naive listeners," Speech Comm., vol. 147, pp. 74-81, Jan. 2023. DOI: [10.1016/j.specom.2023.01.006](https://doi.org/10.1016/j.specom.2023.01.006)
- J. Villegas and S. J. Lee, "Creakiness judgments by Burmese and Vietnamese speakers," in Proc. 25 Conf. of the Oriental chapter of the Int. Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Nov. 2022. DOI: [10.1109/O-COCOSDA202257103.2022.9997908](https://doi.org/10.1109/O-COCOSDA202257103.2022.9997908)
- S. J. Lee, J. Villegas, and M. Oh, "The Non-Coalescence of /h/ and Incomplete Neutralization in South Jeolla Korean," Language and Speech, Aug 2022. DOI: [10.1177/00238309221116130](https://doi.org/10.1177/00238309221116130).
- Y. Baldoria, B. P. Fleming, M. W. R. Liu, J. Villegas, and S. J. Lee, ICU language database series 8: Comparative Zapotec," in ICU Working Papers in Linguistics (ICUWPL), no. 22, pp. 225-349, Tokyo: International Christian University, May 2022. [\[Link\]](#)

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 3件／うちオープンアクセス 0件）

1. 著者名 Villegas Julian, Lee Seunghun J., Perkins Jeremy, Markov Konstantin	4. 巻 147
2. 論文標題 Psychoacoustic features explain creakiness classifications made by naive and non-naive listeners	5. 発行年 2023年
3. 雑誌名 Speech Communication	6. 最初と最後の頁 74 ~ 81
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.specom.2023.01.006	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 J. Lee Seunghun, Villegas Julian, Oh Mira	4. 巻 NA
2. 論文標題 The Non-Coalescence of /h/ and Incomplete Neutralization in South Jeolla Korean	5. 発行年 2022年
3. 雑誌名 Language and Speech	6. 最初と最後の頁 NA
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/00238309221116130	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Villegas Julian, Lee Seunghun J.	4. 巻 NA
2. 論文標題 Creakiness Judgments by Burmese and Vietnamese Speakers	5. 発行年 2022年
3. 雑誌名 in Proc. 25 Conf. 0-COCOSDA	6. 最初と最後の頁 NA
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/0-COCOSDA202257103.2022.9997908	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計1件（うち招待講演 0件／うち国際学会 1件）

1. 発表者名 Villegas Julian
2. 発表標題 Creakiness Judgments by Burmese and Vietnamese Speakers
3. 学会等名 25 Conf. of the Oriental chapter of the Int. Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques (0-COCOSDA) (国際学会)
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 Y. Baldoria, B. P. Fleming, M. W. R. Liu, J. Villegas, and S. J. Lee	4. 発行年 2022年
2. 出版社 International Christian University	5. 総ページ数 124
3. 書名 ICU language database series 8: Comparative Zapotec	

〔産業財産権〕

〔その他〕

PSYPHON website <a href="https://onkyo.u-aizu.ac.jp/psyphon/">https://onkyo.u-aizu.ac.jp/psyphon/</a>
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	李 勝勳 (Lee Seunghun) (20770134)	国際基督教大学・教養学部・上級准教授  (32615)	
研究分担者	MARKOV K (Markov Konstantin) (80394998)	会津大学・コンピュータ理工学部・教授  (21602)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------