Vowel length contrasts in deep learning: Generative Adversarial Phonology and duration

Wing-sze Kat
The Chinese University of Hong Kong
1155016244@link.cuhk.edu.hk / cuhkpgdewing@gmail.com

This paper models unsupervised learning of identity-based pattern (or vowel length contrast) in speech called the ratio of the acoustic duration from raw consecutive data of multiple natural languages of southeast Asian languages with deep convolutional neural networks. Vowel length contrast, e.g. /äː/ - /ɐ/ in Cantonese, is a common phenomenon and they often correlate with systematic changes in vowel quality and / or duration found in East and Southeast Asian languages, like Cantonese, Vietnamese and Thai. The fiwGAN architecture integrates the Deep Convolutional GAN framework for audio data (InfoWaveGAN) with categorical variables from InfoWaveGAN. Unlike InfoWaveGAN, the latent code in fiwGAN is distributed binomially, and training is conducted using sigmoid cross-entropy. The network is designed to learn and represent identity-based patterns, such as vowel length contrast, in its latent space. By manipulating the first two categorical variables (i.e., vowel quality and duration), the model can transform a short vowel into a long vowel without significantly altering other aspects of the output. The study explores how meaningful representations of identity-based patterns emerge in CNNs and examines the correlation between latent space variables and vowel length contrast. This has broader implications for understanding and interpreting neural networks. We propose a technique for CNNs trained on speech and, based on generative tests, argue that the network learns to represent an identity-based pattern in its latent space (Z). By manipulating several categorical variables (i.e., vowel quality or vowel closeness, duration, tone) in the latent space, we can actively turn a short vowel into a long vowel with no other substantial changes to the output in the majority of cases. Exploration of how meaningful representations of identity-based patterns emerge in CNNs and how the latent space variables outside of the training range correlate with vowel length contrast in the output has general implications for neural network interpretability. The paper discusses four generative tests that demonstrate the network's ability to encode phonetic properties and abstract processes like vowel substitution, which mimics language acquisition in Cantonese and Southeast Asian languages. The manipulation of categorical variables in a fiwGAN is not only to create a long vowel from a short vowel, but also to imitate the intricate tone-vowel interaction among these Southeast Asian languages. Furthermore, the results suggest that the deep convolutional network can imitate language acquisition processes found in language, which means acquisition of the data mirrored that of humans learning the language from which the data originate.

While this project is interesting conceptually and fills a necessary gap in the research, this abstract seems somewhat disorganized and does not include a lot of information that is relevant to the interpretation of these findings. Although the manipulation of categorical variables in a fiwGAN to create a long vowel from a short vowel is a novel and interesting finding, many key details are left somewhat ambiguous. It is unclear whether two categorical variables or several are being manipulated, and how exactly they and the other variables of the network are being manipulated to produce the results.

It also seems to be a glaring omission that the author(s) do not mention what the training data are: although they call it "raw consecutive data," whether they came from one natural language, multiple, or were synthetic seems like a necessary detail. Vowel length and quality alternations in Southeast Asian languages are mentioned towards the end, but it would be a much more interesting finding if the model were trained on data from a Southeast Asian language and its 'acquisition' of the data mirrored that of humans learning the language from which the data originate, but it is actually unclear whether this is the case or not. And while a detailed review of the literature is not necessary in an abstract, it does not inspire confidence that the author(s) do not cite or engage with any of the prior work regarding phonological alternations in InfoWaveGAN models, aside from a GitHub page and a YouTube link.

1. Introduction

Recent advances in artificial intelligence have enabled phonological studies to more precisely model the human encoding process, shedding light on how humans organize sounds and link them to cognition. One important aspect of phonological encoding is the contrast in vowel length. Vowel length contrast refers to the phenomenon where the duration of a vowel sound changes the meaning of a word. Languages with vowel length contrast may use either vowel quantity (duration) or vowel quality (formant frequencies) to mark distinctions. For example, in Japanese od East Asian language, "biru" (ビル, building) and "biiru" (ビール, beer) differ only in the length of the vowel "i". This phenomenon is common among Asian languages, including Japanese, as well as languages in the Kra–Dai family (e.g., Thai, Vietnamese) and the Sino-Tibetan family (e.g., Cantonese, Tibetan). However, these languages differ in whether they use vowel quantity or quality to mark contrasts, due to different historical developments. For instance, Tibetan exhibits vowel harmony in its disyllabic words.

Across these diverse regions, three common features emerge among languages with vowel length contrast: the presence of monosyllable, a rigid tonal system and a preference for CV or (C)VC syllable structures. For example, Cantonese is notable for its syllable-stress system among Chinese languages. Nevertheless, vowel length contrast is not unique to East and Southeast Asian languages; it also occurs in non-tonal languages such as those in the Austronesian, Slavic, and Germanic families. This paper focuses on East and Southeast Asian languages—specifically Cantonese, Vietnamese and Thai—to investigate tone-vowel interactions within syllabic constituents and other categorical contrasts in classification.

There is a consensus that the tones of Sino languages (e.g., Cantonese of Yue Chinese language) and Vietnamese originate from laryngeal features of onset and coda consonants, while Thai retains both voiced obstruents, which can complement the process of tonogenesis. The training data for this study include Cantonese, Vietnamese and Thai, all of which are rich in tones and also demonstrate vowel length contrast. Given that a large number of tones are realized on full vowels with bright timbre, the interaction between tone and vowel length should be examined together.

Previous studies (Hu, 1980, 2003; Sun, 1997; Zhou, 1983) have suggested that pitch contours are linked to vowel duration, but the relationship between tone and vowel length remains vaguely described. These researchers have proposed various analyses for capturing the pitch patterns of Lhasa Tibetan, incorporating distinctions such as falling versus non-falling pitch contours or short versus long tone lengths into their tonal descriptions. Therefore, this paper aims to investigate the significance of vowel length contrast in CV and CVC syllable structures and their impact on vowel systems, by utilizing Generative Adversarial Networks (GANs) to model both the formation and constraints of phonologies in East and Southeast Asian languages.

This paper aims to investigate the mechanisms underlying the phonological feature of vowel length contrast, and to explore the feasibility of applying deep learning methods to this area. The following research questions are proposed:

1. In the phonological alternation, the two categorical variables, vowel length and quality alternations, are being manipulated. Can Generative Adversarial Networks (GANs) effectively model and elucidate vowel length contrast across Southeast Asian languages with different historical origins of this phenomenon?

2. Are the phonological systems of these southeast Asian languages and the interaction between vowel length and vowel quality as contrastive features? There is no universal phonological principle that makes vowel length and vowel quality mutually exclusive as contrastive features. Cantonese relies on vowel quality for the vowel contrast and length is a major dimension in Thai. For Vietnamese, vowel quality is the main contrastive feature, but intrinsic duration differences exist (i.e., some vowels are naturally longer or shorter, but this is not phonemic length). Are there perceptual and phonetic reasons why languages may prefer to use one feature over the other?

3. In language acquisition, what are the perceptual processes involved in distinguishing categorical contrasts in vowel quality and duration within #CV(C) and #(C)VC syllable structures? (e.g., #(C)VT, #(C)VD)

4. How is vowel length contrast reflected in the vowel charts (monophthongs) of these languages?

5. Duration and vowel quality are the two categorical variables in GANS, is vowel length contrasts enough for meaning distinguishing, what are the importance of other elements, like suprasegment? Specifically, can GANs capture the complex phonological patterns arising from tone-vowel interactions in the syllabic constituents of Cantonese, Vietnamese, and Thai?

This paper is to answer these above questions.

GANs have shown promise in modeling complex distributions in speech data, including phonetic and phonological contrasts. In principle, GANs can be trained to generate and discriminate between vowel tokens differing in length and quality, even when the historical origins of these contrasts differ across languages. However, the effectiveness of GANs depends on the quality and quantity of training data, the architecture used, and the specific features extracted (e.g., formant trajectories, duration, spectral shape).

For vowel length contrasts, GANs can learn to distinguish and generate vowels of different durations if the training data is annotated accordingly. They can also help elucidate how length contrasts are realized acoustically in different languages, potentially revealing language-specific patterns or constraints. However, GANs are less interpretable than traditional statistical models, so elucidating the phonological nature of the contrast (e.g., whether it is phonemic or allophonic) may require additional analysis beyond the GAN outputs.

1. Can Generative Adversarial Networks (GANs) effectively model and elucidate vowel length contrast across Southeast Asian languages with different historical origins of this phenomenon?

GANs have shown promise in modeling complex distributions in speech data, including phonetic and phonological contrasts. In principle, GANs can be trained to generate and discriminate between vowel tokens differing in length and quality, even when the historical origins of these contrasts differ across languages. However, the effectiveness of GANs depends on the quality and quantity of training data, the architecture used, and the specific features extracted (e.g., formant trajectories, duration, spectral shape).
For vowel length contrasts, GANs can learn to distinguish and generate vowels of different durations if the training data is annotated accordingly. They can also help elucidate how length contrasts are realized acoustically in different languages, potentially revealing language-specific patterns or constraints. However, GANs are less interpretable than traditional statistical models, so elucidating the *phonological* nature of the contrast (e.g., whether it is phonemic or allophonic) may require additional analysis beyond the GAN outputs.

---

2. Are vowel length and vowel quality mutually exclusive as contrastive features? Are there perceptual and phonetic reasons why languages may prefer to use one feature over the other?

There is no universal phonological principle that makes vowel length and vowel quality mutually exclusive as contrastive features. Many languages use both, either independently or in interaction. For example:

- **Cantonese:** Primarily uses vowel quality for contrast; length is not phonemic.
- **Thai:** Vowel length is a major contrastive feature; both short and long vowels exist for most qualities.
- **Vietnamese:** Vowel quality is the main contrastive feature; some vowels are intrinsically longer or shorter, but length is not phonemic.

Perceptual and phonetic reasons for preference:

- **Perceptual distinctiveness:** Languages may prefer the feature that provides the greatest perceptual distance between categories, given their phonetic inventory and phonotactic constraints.
- **Phonetic naturalness:** Some vowel qualities are more easily distinguished by duration (e.g., high vowels tend to be shorter than low vowels cross-linguistically).
- **Historical development:** The diachronic evolution of a language may favor one feature over another due to mergers, splits, or contact with other languages.
- **Functional load:** If one feature (e.g., tone) already carries a heavy functional load, the language may avoid overloading the same syllable with multiple contrasts.

---

3. In language acquisition, what are the perceptual processes involved in distinguishing categorical contrasts in vowel quality and duration within #CV(C) and #(C)VC syllable structures?

Infants and young children are sensitive to both spectral (quality) and temporal (duration) cues from an early age. The processes involved include:

- **Categorical perception:** Learners develop category boundaries for both vowel quality and duration, often influenced by the distribution of tokens in the input.
- **Cue weighting:** Depending on the language, children learn to attend more to the relevant cue (quality or duration) for distinguishing meaning.
- **Syllable structure effects:** The position of the vowel within the syllable (onset, nucleus, coda) can affect the salience of duration and quality cues. For example, vowels in open syllables may have more variable duration, while those in closed syllables may be more stable.
- **Interaction with other features:** In tone languages, learners must also parse pitch cues, which can interact with vowel duration and quality.

---

4. How is vowel length contrast reflected in the vowel charts (monophthongs) of these languages?

- **Thai:** Vowel charts typically show pairs of vowels at each quality, one short and one long (e.g., /a/ vs. /aː/), often occupying the same position in the F1-F2 space but differing in duration.
- **Cantonese:** Vowel length is not phonemic, so the chart reflects only quality contrasts.
- **Vietnamese:** Vowel charts show quality contrasts, with some vowels being intrinsically longer or shorter, but not as a phonemic length contrast.

In acoustic vowel charts (F1-F2 plots), length contrasts may not be visible unless duration is explicitly encoded (e.g., by color or symbol size).

---

5. Is vowel length contrast alone enough for meaning distinction? What is the importance of other elements, like suprasegmentals? Can GANs capture complex tone-vowel interactions in Cantonese, Vietnamese, and Thai?

- **Vowel length alone:** In some languages (e.g., Thai), vowel length alone can distinguish meaning (minimal pairs exist). In others (e.g., Cantonese, Vietnamese), vowel quality or tone is more important.
- **Suprasegmentals:** Tone (pitch), stress, and other suprasegmental features are crucial in these languages, especially for distinguishing meaning in monosyllabic words.
- **GANs and tone-vowel interactions:** GANs can, in principle, model the joint distribution of segmental (vowel quality, length) and suprasegmental (tone) features, provided the input features capture both spectral and pitch information. However, modeling the *interaction* (e.g., how tone affects vowel duration or quality) requires careful feature engineering and possibly multi-modal GAN architectures.

Summary:

- Vowel length and quality are not mutually exclusive; languages choose based on perceptual, phonetic, and historical factors.
- GANs can model these contrasts if provided with appropriate data and features, but interpreting the results requires linguistic expertise.
- Suprasegmental features like tone are essential in Southeast Asian languages, and advanced models are needed to capture their interaction with segmental features.

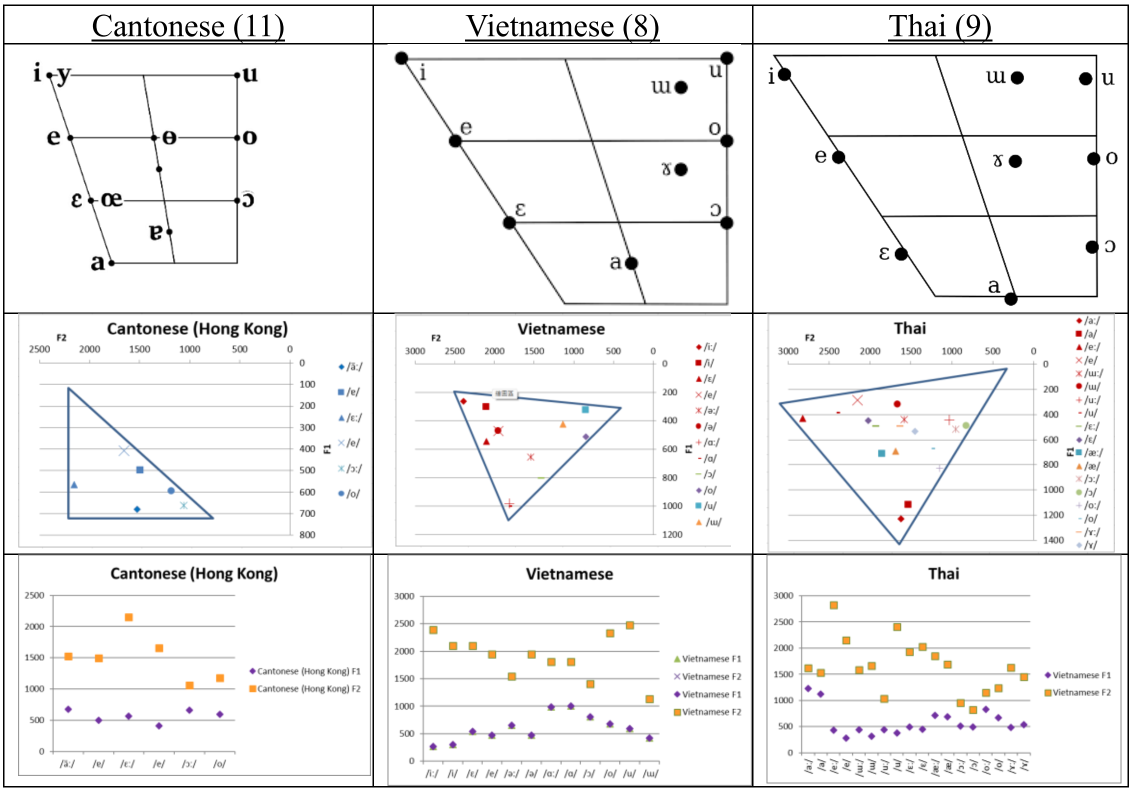


Figure 1. Vowel charts of monophthongs and Formants of contrasting vowels

The term "vowel highness" can be somewhat ambiguous in phonological discussions, as vowel charts may differ significantly across languages. To address this variability, it is useful to represent vowel systems using formant extraction, which provides a more objective and quantifiable depiction of vowel quality. In our experiment, we observed that the distribution of vowel height and backness varies across the three languages under investigation. Specifically, the contrasting vowels in Cantonese (here, we focus on a subset of six vowels) tend to be realized with relatively low formant values, indicating lower tongue positions. In Vietnamese, back vowels are also produced with low tongue positions (i.e., they are phonetically low), whereas in Thai, back vowels are articulated with higher tongue positions (i.e., they are phonetically high). These findings highlight the importance of using acoustic measurements, such as formant frequencies, to accurately compare vowel systems across languages, rather than relying solely on traditional phonological categories.

For explaining the genesis of contrasts and simulating evolution of contrasts, GANs do not inherently model the historical or linguistic processes (the "genesis") that led to vowel length contrast in different languages. For that, the future work would need linguistic analysis, possibly combined with other machine learning models designed for historical linguistics.

2. The Model: Explain WaveGAN and InfoGAN briefly

Humans demonstrate a very early sensitivity to the acoustic features of human languages, particularly to those that form the foundational building blocks of sound systems. One key aspect of this sensitivity is the perception of certain phonetic contrasts as categorical distinctions, even when these contrasts are based on continuous, incremental changes in the acoustic signal. This phenomenon is especially relevant in the lexical domain, where such distinctions can signal differences in word meaning.

This study aims to investigate the long-short vowel contrast as a categorical distinction, focusing on both the quantitative aspect (duration, as a continuous incremental change) and the qualitative aspect (vowel quality) within the boundaries of the lexical domain. Specifically, the research seeks to define how duration and quality interact to form perceptually distinct vowel categories.

Given this background, there is a need for research utilizing deep learning approaches. Deep learning models can be employed to analyze and model the perception and production of vowel contrasts, offering insights into how these categorical distinctions are learned and represented in both human cognition and artificial systems.

Previous research in deep learning applied to phonology (Beguš, 2021) has demonstrated that methods such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) are effective for interpreting identity-based patterns and modeling generative phonological processes, such as reduplication, in language acquisition. Among various deep learning approaches, this paper selects GANs as the primary method because they can be trained to imitate vowel length contrasts, provided that sufficient training data containing these contrasts is available. For instance, models like WaveGAN (for audio) or spectrogram-based GANs can learn to generate speech samples in which vowel length serves as a distinguishing feature, as long as the training data includes examples of both short and long vowels.

WaveGAN is a type of Generative Adversarial Network (GAN) designed specifically for generating raw audio waveforms. Introduced in 2018, WaveGAN adapts the architecture of DCGAN (Deep Convolutional GAN) to work with one-dimensional audio signals instead of two-dimensional images. The generator creates audio samples from random noise, and the discriminator tries to distinguish between real and generated audio. WaveGAN has been used for tasks like generating short audio clips, such as spoken digits or drum sounds. InfoGAN is another GAN variant that aims to learn interpretable and disentangled representations in an unsupervised way. InfoGAN extends the standard GAN by maximizing the mutual information between a subset of the generator's input (called the "latent code") and the generated data. This encourages the model to learn meaningful and controllable features, such as digit rotation or thickness in generated images of handwritten digits, without explicit labels. In summary, WaveGAN is GAN for generating raw audio waveforms, while InfoGAN in advance is GAN that learns interpretable, disentangled features by maximizing mutual information.

2.1 why not use Variational Autoencoders

There are other several types of artificial intelligence tools, one of the main one is Variational Autoencoders (VAEs) which can also learn complex distributions of sound data and reveals how humans perceive variations in sounds and musical tones. While Variational Autoencoders (VAEs) are powerful generative models capable of learning complex data distributions, there are several reasons why they may not be the optimal choice for certain tasks:

Firstly, it is because that VAEs typically assume that the latent variables follow a simple prior distribution, such as a standard normal (Gaussian). If the true underlying data distribution is multimodal or significantly deviates from this assumption, VAEs may struggle to accurately capture the data structure. The reason not to use it in this research is that it uses maximum likelihood estimation to model the parameters of a unimodal Gaussian distribution. Naturally, this leads to problems, because the true distribution is bimodal rather than unimodal.

Secondly, VAEs are known to produce blurry or less sharp reconstructions, especially in image and speech data, due to the regularization imposed by the KL-divergence term. This can be problematic when high-fidelity reconstruction is required.

Thirdly, the latent space learned by VAEs may not always be easily interpretable, which can be a disadvantage if the goal is to analyze or visualize specific features (such as phonetic or linguistic properties), given that its input and output are so similar, it cannot analyze the perception process well.

Last but not least, GANs, another model that deploys Gaussian Mixture Models (GMMs), create sounds with generator and discriminator, the later one distinguish the accuracy of sound imitation. As a result, this paper continues to utilize the Deep Convolutional GAN framework. GANs can imitate and generate vowel length contrasts if trained on relevant data, but they do not inherently model or explain the historical genesis of these contrasts across languages.

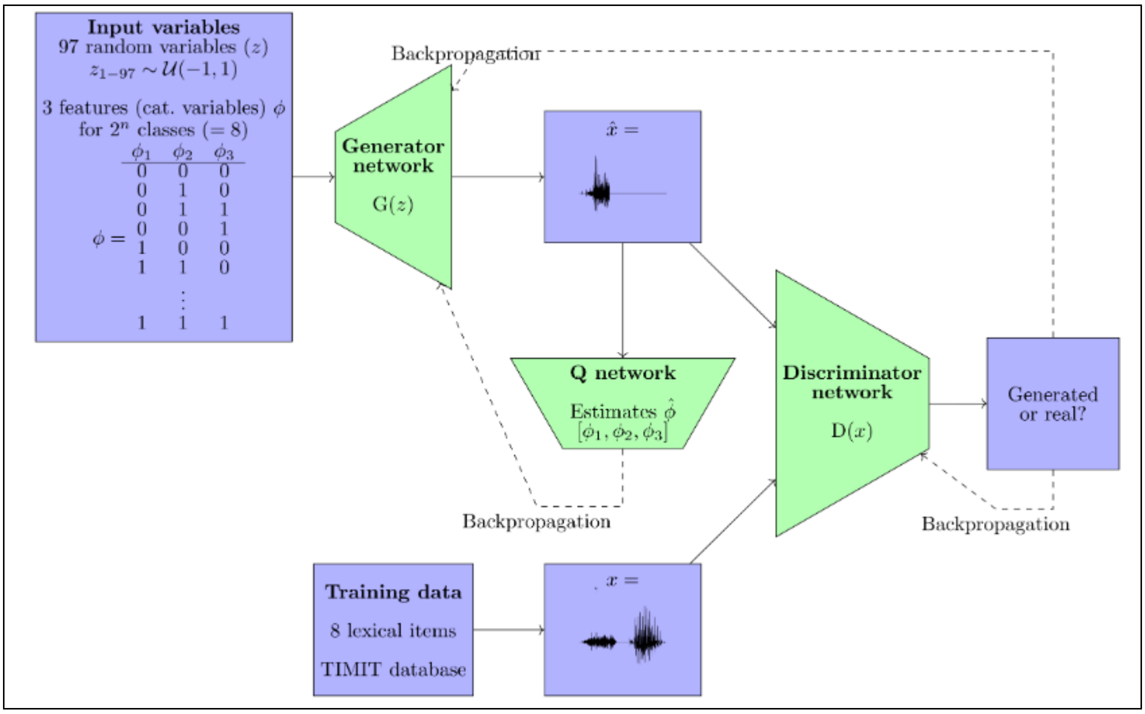


Figure 2. The mechanism of ciwGAN (Categorical InfoWaveGAN)

2.2 the core research question in phonology and 'acquisition' way in WaveGAN

the manipulation of categorical variables in a fiwGAN is to create a long vowel from a short vowel, which imitates the situation that infants of first language acquisition or second language learner acquire long vowels from short vowels.

3. Vowel length contrast in Training Data

The code testing environment is Python3.6.0, Keras 2.1.6 and Tensorflow 1.8.0.

Recordings of the training data (N=553) were made using Audacity. The audio was originally sampled at 44.1 kHz and then downsampled to 16 kHz. The training data includes Cantonese (Hong Kong, N=40), Vietnamese (both Northern and Southern dialects, N=353), and Thai (Suphan Buri, N=164), all of which are tonal languages. The nativity of the speakers ensures the accuracy of WaveGAN's input data, allowing the output to closely mirror actual human language acquisition phenomena.

The syllabic constituent's condition for Cantonese vowel length contrast is #CVT, while it is #CV, #(C)VT in Vietnamese, and #TV, #DV, #(C)VT and #(C)VD in Thai. # means boundary, T represents unaspirated obstructs /p t k/ and D represents voiced obstructs /b d g/.

Thai has the most flexible system, allowing vowel length contrast in syllables with both unaspirated and voiced stop onsets and codas, and in both open and closed syllables, e.g., /ka/ (short) vs. /kaː/ (long); /kat/ (short) vs. /kaːt/ (long); /gad/ (short) vs. /gaːd/ (long)
So, there are 3 kinds of combinations, vowel length is phonemic and independent of tone., shed light on the construction of syllabic constituent with the factor, vowel length contrast, in this three phonology. Tone development in these languages interacts with syllabic structure and vowel length in different ways, reflecting their unique phonological histories and systems.

4. Experiment designs in CiwGAN (Beguš, 2021a)

The GAN is unsupervised that category labels are automatically added. But to enhance the accuracy rate for more 27% to 89%, we use the CGAN.

4.1 the #TV, #T^{h}V, #DV, #(C)VT and #(C)VD syllabic structures

Voice onset time (VOT) in phonetics describing the time difference between the release of a stop consonant and the start of vocal cord vibration (voicing). It's a key feature for distinguishing between voiced and voiceless consonants in speech. In the training, GAN imitates the sound structures but with an insensitive to ± voiced consonants. In this paper, /s/ is excluded as it is reported by Beguš () of its bad effects, so we just use /p-, t-, k-, Ø-/, /p^h-, t^h-, k^h-/ and /b-, d-, g-/ for CV structure and /-p, -t, -k/, /-b, -d, -g/, /-m, -n, -ŋ/ and /-ɪ, -ʊ/ for VC structure.
We conduct VOT measurement to compare with GAN's performance.

5. Vowel length contrast: Bare WaveGAN

The bare WaveGAN is to imitate CV, VC structures and ton-vowel interactions of southeast Asian languages, and its 'acquisition' of the data mirrored that of humans learning the language from which the data originate

5.1 CV structures

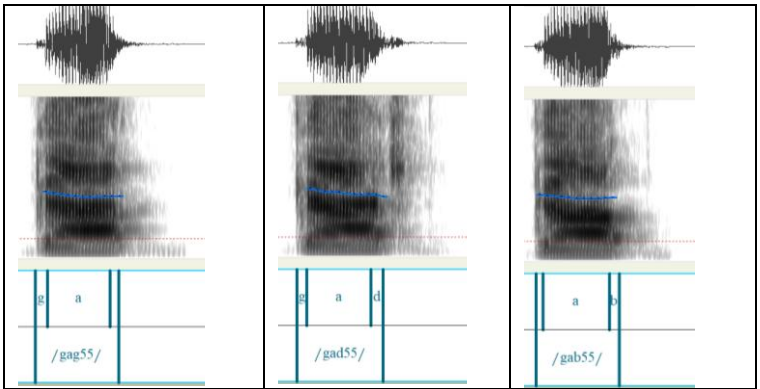
In this research, the issue of vowel length contrast triggers acquisition on syllabic structures. In first language acquisition, after approximately six months, children use what look like units with "syllable structure and Consonant-vowel combinations (CV) are preferred". (Meisel, 2011). It indicates that CV is the easiest and basic format to learn, so this paper deploys WaveGAN to imitate the various Voice Onset Time (henceforth VOT) of three ways of contrast in **#CV (#TV, #T^{h}V, #DV)** of Thai, but the conditions for each combination of syllabic constituents are different: Vowel length contrast on aspiration condition. Inaspirated and voiced initial consonants only match with short vowels.

- Formants, intensity of contrasting vowels

5.2 VC structures

The investigation in VC structure is to analyze **the condition of vowel length contrast**, which is the only form in Cantonese and **vowels contrast in both length and quality in #VT (#VVT) in Vietnamese. #VT** is also the fixed way to keep voiced obstuents /-b, -d, -g/ in Thai. But controversially, when a vowel is followed by a obstuents coda that accompanies with glottalization, the vowel is pronounced more briefly than when followed by other codas or when open (no coda). This shortening effect is due to the abrupt closure of the glottal stop, which cuts off the vowel. (It has been well established that vowels are phonetically longer before voiced than voiceless stops. The effect that voiced consonants have on preceding vowel duration has been observed in various languages across language families: among others in English, French, Russian, Korean, and Bengali (Beguš 2016). But why short vowels in Thai are preceding /-b/, /-d/ and /-g/?) In optimal theory of symbolism, the length of vowel, duration, is also considered as mora-bear-unit. But how is to explain the vowel contrast? Then, the comparison on ± voiced obstruents aims to reveals acoustic phonetic cues that keep the robustness of codas, which is **#VT in both long and short vowels** (Cantonese /-p, -t, -k/ and Vietnamese /-p, -t, -k/) vs. **#VD in short vowels** (Thai /-b, -d, -g/).

(VOT) following coda which is voiced obstruents should have long vowels



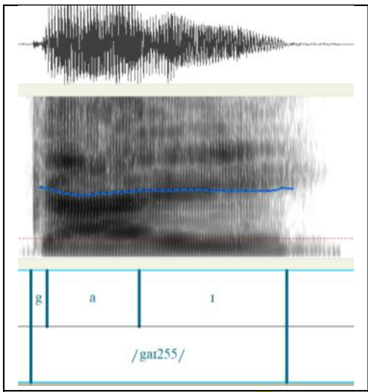
Thai : **#VD** /-ab/	Thai : **#VD** /-ad/	Thai : **#VD** /-ag/

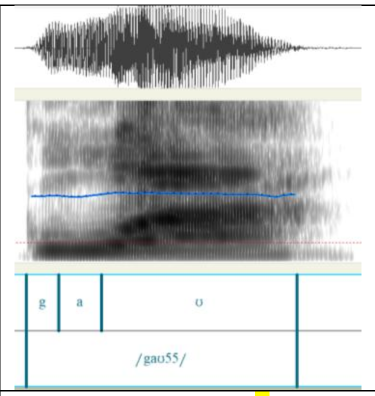
Figure X. Waveform and VOT on spectrogram of **#VD** in Thai with a coda /-b, -d, -g/

Since vowel contrasts differing in both duration and quality (can be attested in Praat spectrograms) in Cantonese, which features are salient in acquisition?

Furthermore, under the condition of vowel length contrast, the comparison on coda /-ɪ, -

ʊ/ aims to reveals how the long and short vowels piece together with the /-i, -u/, which is **#VT in both long and short vowels** in Cantonese /-ɪ, -ʊ/ vs. Vietnamese **#VT in both long vowels** in /-ɪ, -ʊ/ vs. **#VD in short vowels** in Thai /-ɪ, -ʊ/





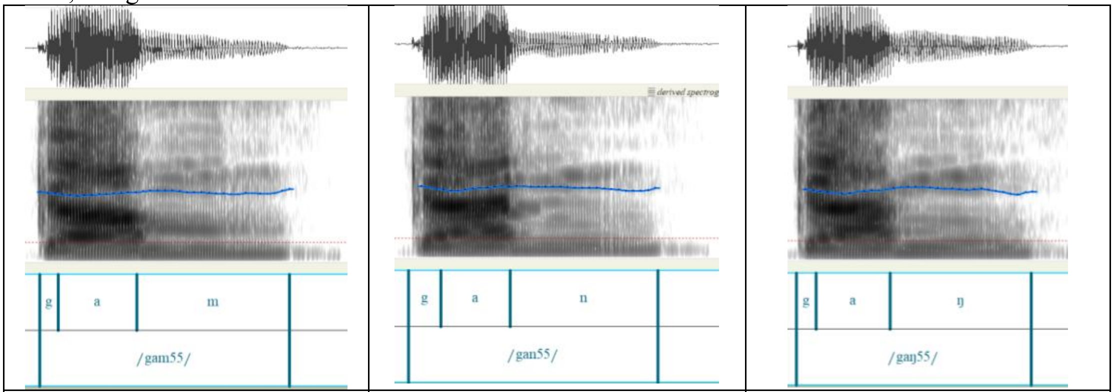
Thai **: #VD** /- aɪ /	Thai **: #VD** /- aʊ /

Figure X. Waveform and VOT on spectrogram of **#VD** in Thai with a coda /-b, -d, -g/

5.3 the tone-vowel interactions of syllabic constituents

In a previous study (Kat, 2018), syllabic constituents are tested for preference of curving tones contours in Northern Wu dialects. Other previous studies, including those on African languages (Connell, 2002) and Min Chinese (Bao & Shi & Xu, 2010), have shown that **intrinsic pitch height**—the natural pitch associated with high vs. low vowels—can influence tone systems. This is often cited as a reason why certain tones are more likely to occur with certain vowel heights. In Cantonese, low vowels /ä:j/ and /ɐj/ but only combine with high level tone /55/. The technique problem is that can GANs imitate sound that with different tones with a small amount of data? (N=X) Do GANs knows the syllabary of these Southern Asian languages, otherwise, it will create nonce.

To cope with complex tone system in Southeast Asian languages, **'duration'** is the phonetic cue to be considered. How do learners gradually develop a new duration category to cope with the new vowel durational distributions in Thai? Thai retains phonemic vowel length distinctions as a key feature of its phonological system. The comparison of **#VD** in Thai with **long vowels** followed by /-m, -n, -ŋ /vs. **short vowels** followed by /-m, -n, -ŋ / aims to reveals how tones are manifest on their attached long and short vowels.



Thai **: #VD + coda** /-m/	Thai **: #VD + coda** /- n/	Thai **: #VD + coda** /- ŋ/

Figure X. Waveform and VOT on spectrogram of **#VD** in Thai with the coda /-m, -n, -ŋ /

In Thai, the phonological features of tone that are proposed by Wang (1967) can be represented by rule-like formula. The tone features are [Contour], [High], [Central], [Mid], [Rising], [Falling] and [Convex]. The original tone is /21/ by default in #VC (short vowels). It also becomes a lowing falling tone /21/ of [-HIGH, +FALLING] features with these initial /p, t, k, p^h-, t^h-, k^h-, b-, d-, g-/ as in #CVD (short vowels). Given that short vowel in #VD (/-m, -n, -ŋ, -e, -ʊ/) is so productive in tones, the "tone feature" is analyzed to set up as one of the categorical variables besides "duration" and "vowel quality" in vowel length contrasts. Below are two redundancy conventions for tone alternations in #VD (short vowels that end with nasals and voiced codas /-m, -n, -ŋ, -e, -ʊ/):

1. #/p^h-, t^h-, k^h-/+VT $\begin{bmatrix}-HIGH \\ +FALLING\end{bmatrix} \rightarrow \begin{bmatrix}+HIGH \\ -FALLING\end{bmatrix}$ / nasals and voiced codas____

   The lower tone /21/ that attached on short vowel turns to [33], [45] and a low-raising tone in #CVD, which elucidates a [-falling] feature that disables a glide, which is consistent in the monothrongs in vowel length contrast of Thai.

2. # /k/+VT $\begin{bmatrix}-HIGH \\ +FALLING\end{bmatrix} \rightarrow \begin{bmatrix}-HIGH \\ +CONTOUR\end{bmatrix}$ / nasals and voiced codas____

   The lower tone /21/ turns to /14/, /21/ and /41/ with an unaspirated initial /k/ as in #CVD (short vowels). Since the tone contours may undergo historical change, both falling and rising tones occurs occasionally under the same syllabic constituent's condition.

The issues raised above regarding Thai are language-specific, but a broader question arises: under the same mechanisms of vowel length contrast (i.e., duration and/or vowel quality), do tonal features serve as supplementary cues to vowel length distinctions in Southeast Asian tonal languages? In Thai, two redundancy conventions for tone alternations in #VD (short vowel) contexts are particularly salient: the features of [±HIGH], [±FALLING], and [+CONTOUR] are clearly distinguished. In contrast, in Cantonese, the /äː/–/ɐ/ vowel length pairs are only found with the high level tone /55/. These patterns suggest that vowel length contrasts are not solely realized through suprasegmentally features such as duration and vowel quality, but also through tone-vowel interactions. For example, in Cantonese, the low vowels /äː/–/ɐ/ in #VT contexts combine with high tones, whereas in Thai, short vowels in #VD contexts combine with non-high tones. This indicates that the interaction between tone and vowel length is an important factor in the phonological systems of Southeast Asian tonal languages. So, this paper intends to imitate the intricate tone-vowel interactions in Southeast Asian languages that have vowel length contrasts.

However, tone systems do not necessarily develop in the same way, even when languages employ similar or different mechanisms, such as duration and/or vowel quality, for vowel length contrasts. For example, Thai primarily uses duration, while Cantonese relies on vowel quality. Although similar mechanisms for vowel length contrast may exist across languages, the development and interaction of tone systems with these mechanisms can vary significantly. Vietnamese, which employs a mixed system—using vowel quality for monophthongs and duration for diphthongs—demonstrates that the relationship between vowel contrast mechanisms and tone is not straightforward. Both historical and phonetic factors contribute to this complexity. The Vietnamese system, in particular, suggests that the mechanism of vowel contrast (quality vs. duration) can influence tone realization and potentially tone development. This mixed system thus provides a unique opportunity to study how different mechanisms of vowel contrast interact with tone.

6. Discussion

The relation of contrasting vowels is not investigated. '…the poor of available literature on infants' sensitivity towards the internal organization of syllables such as consonant and vowel contrasts, by affirming their discrimination ability on the suprasegmental feature of tone in a tonal language.' (Lei 2007) This research fills the research gap of how native speakers of Cantonese, Thai and Vietnamese perceive syllabic constituents under different phonetic conditions.

Previous acoustic studies have confirmed systematic durational differences between the contrastive pairs, indicating that vowel quantity is a significant feature of vowel length in Cantonese (Kao, 1971; Lee, 1983, 1985; Shi & Liu, 2005). Lee (1983, 1985) observed and Shi and Liu (2005) further confirmed that, at a similar speech rate, the durational ranges of the long and short vowels of a vowel pair did not overlap, although there was a large degree of spectral overlap (Zee, 2003). The large spectral overlap implied that vowel quality is not the only distinctive feature in signaling

the contrast. Thus, they suggested that vowel quantity also functions as a crucial distinctive feature

Cumulative Enhancement Model (CEM) (Flynn, Foley, & Vinnitskaya, 2004) and the Typological Primacy Model (Rothman, 2010, 2015)

Phonological Permeability Hypothesis (Cabrelli Amaro & Rothman, 2010)

Speech Learning Model

(SLM) (Flege, 1989, 1995, 1999a; Flege, Bohn, & Jang, 1997), the Perceptual Assimilation Model (PAM) (Best, 1995; Best & Tyler, 2007), and the Second Language Linguistic Perception (L2LP) model (Escudero, 2005, 2009)

Single-Category (SC) Type assimilation

Category-Goodness (CG) Type Assimilation

Uncategorized versus Categorized (UC) Type assimilation

There are perceptual and phonetic reasons why languages may *prefer* to use one feature over the other, or why the two features may not be fully independent:

- **Perceptual distinctiveness:** In some vowel systems, the acoustic cues for length and quality can overlap. For example, high vowels tend to be shorter than low vowels, even when length is not phonemic. This can make it harder for listeners to reliably distinguish length and quality *independently*.
- **Functional load:** If a language already has a rich system of vowel quality contrasts, adding a length contrast may not add much communicative value, and vice versa.
- **Historical development:** The phonological history of a language may favor the development of one feature over the other.

7. Conclusion

The results of our computational linguistics experiment on phonological alternation in vowel length contrast—where duration serves as a phonemic distinction in Thai, and vowel quality fulfills this role in Cantonese and Vietnamese—address a long-standing research gap in the debate between symbolism (as represented by the Optimality Theoretic framework, 1990) and connectionism in cognitive science (Beguš 2022). Our discriminator not only distinguishes long vowels from short vowels, but also generates tonal distinctions from these two categories. Notably, tones are especially evident in the #VT (vowel-tone) context, raising questions about categorical levels and the nature of incremental change. Furthermore, training a deep convolutional GAN on the data reveals an approximation of rule-like generalizations, suggesting that such models can capture systematic phonological patterns.

8. Acknowledgement

It was a ‘poster presentation’ at the Annual Meeting on Phonology 2025 at the University of California, Berkeley, the U.S. This research was not funded by any grant but I would like to thank my family. I would also like to thank Professor Gašper Beguš, Dr. Stephan Hyeonjun Stiller, Professor Lawrence Cheung Yam-leung and Dr. Wong Tak-sum for their useful comments. Any errors are my responsibility.

Reference

Abramson, A. S. (1962). *The Vowels and Tones of Standard Thai*.

Bao, Z., Shi, J., & Xu, D. (2010). The theory and application of generative phonology [生成音系學理論及其應用]. Beijing: Peking University Press.

Beguš, Gašper & Zhou, Alan & Wu, Peter & Anumanchipalli, Gopala K. (2023). Articulation GAN: Unsupervised modeling of articulatory learning. / arXiv:2210.15173v2

Beguš, Gašper. (2020). Artificial sound change: Language change and deep convolutional neural networks in iterative learning. https://doi.org/10.48550/arXiv.2011.05463 / arXiv:2011.05463v2

Beguš, Gašper & Leban, Andrej & Gero, Shane. (2024). Approaching an unknown communication system by latent space exploration and causal inference. https://doi.org/10.48550/arXiv.2303.10931 / arXiv:2303.10931v2

Beguš, Gašper & Lu, Thomas & Zhou, Alan & Wu, Peter & Anumanchipalli, Gopala K. (2023). Ciwagan: Articulatory Information Exchange. https://doi.org/10.48550/arXiv.2309.07861 / arXiv:2309.07861v1

Beguš, Gašper. (2021). CiwGAN and fiwGAN: Encoding information in acoustic data to model lexical learning with Generative Adversarial Networks. Neural Networks 139 (2021) p.305–325.

Beguš, Gašper. (2021). Deep Sound Change: Deep and Iterative Learning, Convolutional Neural Networks, and Language Change. https://doi.org/10.48550/arXiv.2011.05463 / arXiv:2011.05463v2

Beguš, Gašper. (2020). Generative Adversarial Phonology: Modeling Unsupervissed Phonetic and Phonological Learning With Neural Networks. Frontiers in Artifical Intelligence. 3:44. doi: 10.3389/frai.2020.00044

Beguš, Gašper. (2021). Identity-Based Patterns in Deep Convolutional Networks: Generative Adversarial Phonology and Reduplication. https://doi.org/10.48550/arXiv.2009.06110 / arXiv:2009.06110v2

Beguš, Gašper & Zhou, Alan. (2022). Interpreting Intermediate Convolutional Layers of Generative CNNs Trained on Waveforms. IEEE/ACM Transactions on audio, speech, and language processing, vol. 30, 2022, p.3214-3229.

Beguš, Gašper. Modeling unsupervised phonetic and phonological learning in Generative Adversarial Phonology. https://doi.org/10.48550/arXiv.2006.03965 / arXiv:2006.03965v1

Beguš, Gašper. (2022). Local and non-local dependency learning and emergence of rule-like representations in speech data by deepconvolutional generative adversarial networks in Computer Speech & Language 71 (2022).

Boersma, Paul & Weenink, David. (2024). *Praat: doing phonetics by computer* (Version Praat (version 6.4.27) [Computer program]. Retrieved March 15, 2024, from http://www.praat.org/

Brato, Thorsten. (2025). Praat Script: TB-Track Vowels (V2-deprecated). Retrieved July 15, 2025, from https://www.uni-regensburg.de/language-literature-culture/english-linguistics/staff/brato/praat-scripts/index.html

Brato, Thorsten. (2025). Praat Script: TB-Obstruent analyis (deprecated) (V2-deprecated). Retrieved July 15, 2025, from https://www.uni-regensburg.de/language-literature-culture/english-linguistics/staff/brato/praat-scripts/index.html

Connell, Bruce. (2002). *Tone languages and the universality of intrinsic F0: evidence from Africa*, Journal of Phonetics, Volume 30, Issue 1, 2002, p.101-129, ISSN 0095-4470, https://doi.org/10.1006/jpho.2001.0156.

Donahue, Chris. (2025). Wavegan. Retrieved July 17, 2025, from https://github.com/chrisdonahue/wavegan?tab=readme-ov-file

Donahue, Chris. (2025). Specgan. Retrieved July 17, 2025, from https://github.com/chrisdonahue/specgan

Hu, Tan. [胡坦]. (2003). Central Tibetan (Lhasa Tibetan) [衛藏方言(拉薩話)]. In X. Ma (Ed.) [馬學良], A General Introduction to Sino-Tibetan Languages [漢藏語概論] (2nd ed., pp. 149–161). Ethnic Minority Publishing House [民族出版社].

KAT, Wing Sze. (2018). *An Acoustic Study of Syllabic Constituents and Their Interaction in Prosodic and Morphological Contexts in Liyang and Surrounding Wu Dialects.* Thesis (MPhil.), HKUST.

Kirby, J. (2011). *Vietnamese (Hanoi Vietnamese)*. Journal of the International Phonetic Association.

Langr, Jakub & Bok, Vladimir. (2019). GANs in action: deep learning with generative adversarial networks. Shelter Island, New York: Manning. (https://github.com/GANs-in-Action/gans-in-action)

Lei, Margaret Ka-Yan. (2007). 2007. Discrimination of level tones in Cantonese-learning infants. Poster presented at the 16th International Congress of Phonetic Sciences (ICPhS-16), August 6-10, Saarbrücken, Germany.

Meisel, Jurgen M. (2011). First abd second Language Acquisition: Parallels and Differences. Cambridge: Cambridge University Press.

Sun, J. T. S. (1997). The typology of tone in Tibetan. In Chinese Languages and Linguistics IV: Typological Studies of Languages in China (pp. 485–521). Symposium Series of the Institute of History and Philology, Academia Sinica.

Wang, Shi-yuan. (1967). Phonological features of tone. International Journal of American Linguistics, vol. 33, no. 2, pp. 93-105.

ZHANG, Yubin. (2025). Central Tibetan (Lhasa), in Journal of the international Phonetic Association (2025), 54, p.788-810.

Zhou, Jiwen. [周季文]. (1983). A textbook of Tibetan Pinyin [藏語拉薩話的文白異讀]. Ethnic Minority Publishing House [語言研究]. 2, 173–193.

Language acquisition paper

泰國

Tai ji hua yu xiao xi zhe shui hua qing xi du yu sheng xiao fen xi = Acoustic Analysis and Speech Intelligibility in Thai Mandarin Learners.

Noppadolsathan, Kamolwan (author) ; National Taiwan Normal University (Taiwan). Department of Chinese as a Second Language (degree granting institution)

Thesis (Ph.D.)--National Taiwan Normal University (Taiwan), 2021.; 2021

Identity Construction and Second Language Acquisition: A Multiple Case Study of Thai Immigrants in Hong Kong

Zhang, Zhe (Victor)

Philadelphia: Routledge

Journal of language, identity, and education, 2024-07, Vol.23 (4), p.498-511

Computer-Assisted English Lessons for Logistics with Task-Based Language Teaching for Enhancing Receptive and Productive Vocabulary Knowledge of EFL Thai Learners

Durongbhandhu, Nunpaporn ; Suwanasilp, Danuchawat

Language Institute of Thammasat University

Consonant Cluster Acquisition by L2 Thai Speakers

Rungruang, Apichai

Canadian Center of Science and Education

English language teaching (Toronto), 2017-06, Vol.10 (7), p.216

Varapongsittikul, T., & Jitwiriyanont, S. (2025). English Plosive Consonants Produced by Thai Speakers: An Analysis of Voice Onset Time. LEARN Journal: Language Education and Acquisition Research Network, 18(1), 724–747. https://doi.org/10.70730/LJIC7341

Influence of the first and second language on the perception of Thai tones.
Schaefer, Vance (author) ; Indiana University. Second Language Studies. (degree granting institution)

Effects of L1 prosody and phonetic variability on second language perception of Mandarin tones by Thai and Indonesian speakers
Chow, Wen Chun ; Faculty of Humanities
Hong Kong Polytechnic University; 2021
PolyU Electronic Theses

越南

Vietnamese-Speaking Children's Acquisition of Consonants, Semivowels, Vowels, and Tones in Northern Viet Nam
Phạm, Ben ; McLeod, Sharynne
ROCKVILLE: Amer Speech-Language-Hearing Assoc
Journal of speech, language, and hearing research, 2019-08, Vol.62 (8), p.2645-2670

Multilingual Speech Acquisition by Vietnamese-English-Speaking Children and Adult Members
McLeod, Sharynne ; Verdon, Sarah ; Margetson, Kate ; Tran, Van H. ; Wang, Cen ; Pham, Ben ; To, Lily ; Huynh, Kylie
ROCKVILLE: Amer Speech-Language-Hearing Assoc
Journal of speech, language, and hearing research, 2023-07, Vol.66 (7), p.2184-2229

STRENGTHENING OF FRICATIVES IN LANGUAGE ACQUISITION AND LEXICAL BORROWING THE CASE OF SINO-VIETNAMESE / 语言习得和语言借用中摩擦音的闭塞音化一以越南汉字音为实例
Jang, Youngjun ; 张荣俊
Project on Linguistic Analysis
Journal of Chinese linguistics, 2015-01, Vol.43 (1A), p.150-169

The Lexical Tones of Vietnamese Metropoles.

Slówik, Ondřej ; Volín, Jan.
Prague : Karolinum Press; 2020 https://www.jstor.org/stable/jj.7418734?

Vietnamese tone tone is not pitch.
Pham, Hoa Thi ; Rice, Keren. ; University of Toronto
Thesis (Ph.D.)--University of Toronto (Canada), 2001.; 2001

粵語

Assessing the Link Between Perception and Production in Cantonese Tone Acquisition
Mok, Peggy Pik Ki ; Fung, Holly Sze Ho ; Li, Vivian Guo
ROCKVILLE: Amer Speech-Language-Hearing Assoc
Journal of speech, language, and hearing research, 2019-05, Vol.62 (5), p.1243-1257

Mok, Peggy Pik Ki ; Li, Vivian Guo ; Fung, Holly Sze Ho. (2020). Development of Phonetic Contrasts in Cantonese Tone Acquisition in *ROCKVILLE: Amer Speech-Language-Hearing Assoc*, Journal of speech, language, and hearing research, 2020-01, Vol.63 (1), p.95-108.

The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English
MOK, PEGGY P. K.
Cambridge, UK: Cambridge University Press
Bilingualism (Cambridge, England), 2011-10, Vol.14 (4), p.458-472

Cantonese tone production in pre-school Urdu–Cantonese bilingual minority children
Yao, Yao ; Chan, Angel ; Fung, Roxana ; Wu, Wing Li ; Leung, Natalie ; Lee, Sarah ; Luo, Jin
London, England: SAGE Publications
The international journal of bilingualism : cross-disciplinary, cross-linguistic studies of language behavior, 2020-08, Vol.24 (4), p.767-782, Article 1367006919884659

The roles of Cantonese speakers' L1 and L2 phonological features in L3 pronunciation acquisition
Chen, Hsueh Chu ; Tian, Jing Xuan
Abingdon: Routledge
International journal of multilingualism, 2024-01, Vol.21 (1), p.1-17

Appendix: Vocabulary lists

Table 1. Cantonese #(C)VT (11 monothrongs with 3 pairs of vowel length contrast)

<table>
<tr><th>Minimal pairs</th><th>Long vowel</th><th>Short vowel</th></tr>
<tr><td rowspan="12">/ä:/ - /ɐ/

(mostly with a high level tone /55/)</td><td>挨 /Øä:j^{55}/</td><td>閪 /Øɐj^{55}/</td></tr>
<tr><td>齋 /ʦä:j^{55}/</td><td>濟 /ʦɐj^{55}/</td></tr>
<tr><td>炸 /ʦʰä:j^{55}/</td><td>寨 /ʦʰɐj^{55}/</td></tr>
<tr><td>曬 /sä:j^{33}/</td><td>細 /sɐj^{33}/</td></tr>
<tr><td>揩 /hä:j^{55}/</td><td>閪 /hɐj^{55}/</td></tr>
<tr><td>趴 /pʰä:j^{55}/</td><td>批 /pʰɐj^{55}/</td></tr>
<tr><td>拜 /pä:j^{55}/</td><td>幣 /pɐj^{55}/</td></tr>
<tr><td>梯 /tʰä:j^{55}/</td><td>軑 /tʰɐj^{55}/</td></tr>
<tr><td>低 /tʰä:j^{55}/</td><td>呆 /tɐj^{55}/</td></tr>
<tr><td>街 /kä:j^{55}/</td><td>雞 /kɐj^{55}/</td></tr>
<tr><td>凱 /kʰä:j^{55}/(literary pronunciation)</td><td>溪 /kʰä:j^{55}/</td></tr>
<tr><td>掉 /tʰɛ:ʊ^{22}/</td><td>掉 /teʊ^{22}/ (literary pronunciation)</td></tr>
<tr><td rowspan="7">/ɛ:/ - /e/</td><td>舔 /tʰɛ:m^{25}/</td><td>舔 /lem^{25}/ (literary pronunciation)</td></tr>
<tr><td>虔 /kʰɛ:n^{21}/</td><td>虔 /kʰen^{21}/ (literary pronunciation)</td></tr>
<tr><td>驚 /kɛ:ŋ^{55}/</td><td>驚 /keŋ^{55}/ (literary pronunciation)</td></tr>
<tr><td>夾 /kɛ:p˺^{33}/</td><td>夾 /kep˺^{33}/ (literary pronunciation)</td></tr>
<tr><td>坺 /pʰɛ:t˺^{22}/</td><td>坺 /pet˺^{22}/ (literary pronunciation)</td></tr>
<tr><td>惜 /sɛ:k˺^{2}/</td><td>惜 /sek˺^{5}/ (literary pronunciation)</td></tr>
<tr><td>江 /kɔ:ŋ^{55}/</td><td>公 /koŋ^{55}/</td></tr>
<tr><td>/ɔ:/ - /o/</td><td>落 /lɔ:k˺^{2}/</td><td>六 /lok˺^{2}/</td></tr>
</table>

Table 2. Vietnamese #CV, #(C)VT (12 monothrongs with 6 pairs of vowel length contrast)

<table>
<tr><th rowspan="2">Minimal pairs</th><th colspan="2">#TV (P.17)</th><th colspan="2">#DV(P.17)</th><th colspan="4">#VT (#VVT) (p.21)</th></tr>
<tr><th>long vowel</th><th>short vowel</th><th>long vowel</th><th>short vowel</th><th colspan="2">long vowel</th><th colspan="2">short vowel</th></tr>
<tr><td rowspan="4">/i:/ - /i/</td><td rowspan="4">/ti:/
/ki:/</td><td rowspan="4">/ti/
/ki/</td><td rowspan="4">/bi:/
/di:/</td><td rowspan="4"></td><td colspan="2">/i:ʊ/</td><td colspan="2">/ie/</td></tr>
<tr><td colspan="2">(/uik/, /uin)</td><td>/iem/
/ien/</td><td>/
/iet/</td></tr>
<tr><td>/i:p/
/i:t/
/i:k/</td><td>/i:m/
/i:n/, /i:n/
/</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>/i:em/
/i:en/
/</td><td>/i:ep/
/
/i:eŋ/</td></tr>
<tr><td rowspan="2">/ɛ/ - /e/</td><td rowspan="2">/tɛ/
/kɛ/</td><td rowspan="2">/te/
/ke/</td><td rowspan="2">/bɛ/
/dɛ/</td><td rowspan="2">/be/
/de/</td><td>(/uek/)</td><td>(uen)</td><td colspan="2">/eʊ/</td></tr>
<tr><td>/ɛp/
/ɛt/</td><td>/ɛm/
/ɛn/</td><td>/ep/
/et/</td><td>/em/
/en/, /en/</td></tr>
</table>

					/ɛk/	/ɛŋ/	/ek̺/	/
/əː/ - /ə/	/təː/ /kəː/		/bəː/ /dəː/ /gəː/		/əːɪː/		/əɪ/, /əʊ/, /uəʊ/	
					/əːp/ /əːt/	/əːm/ /əːn/	/əp/ /ət/ /ək/	/əm/ /ən/ /əŋ/
/ɑː/ - /ɑ/	/tɑː/ /kɑː/		/bɑː/ /dɑː/ /gɑː/		/ɑːɪː/, /ɑːɪ/, /ɑːʊ/, /uɑːʊ/ /ɔɑːɪː/, /ɔɑːɪ/ /uɑːɪː/,/uɑːɪ/			
					/ɑːp/ /ɑːt/ /ɑːk/, /ɑːk̺/	/ɑːm/ /ɑːn/, /ɑːn̺/ /ɑːŋ/	/ɑp/ /ɑt/ /ɑk/	/ɑm/ /ɑn/ /ɑŋ/
/ɔ/ - /o/	/tɔ/ /kɔ/	/to/ /ko/	/bɔ/ /dɔ/ /gɔ/	/bo/ /do/ /go/	/ɔɪː/		/oɪː/	
					/ɔp/ /ɔt/ /ɔk/	/ɔm/ /ɔn/ /ɔŋ/	/op/ /ot/ /ok/	/om/ /on/ /oŋ/
/u/ - /ɯ/	/tu/ /ku/	/tɯ/ /kɯ/	/bu/ /du/ /gu/	/bɯ/ /dɯ/ /gɯ/	/uɪ/, /uɪː/, /ue/		/ɯɪː/, /ɯʊ/	
					/up/ /ut/ /uk/	/um/ /un/ /uŋ/	/ /ɯt/ /ɯk/	/ɯm/ / /ɯŋ/
	P.13-14 reannotate	reannotate	reannotate	reannotate	reannotate	reannotate	reanno tate	reannotate

Data (sample used) of #VT (#VVT) in VietnameseØ

Minimal pairs	Long vowel	Short vowel
/iː/ - /i/	/iːe/: liːen^{213}, diːen^{55}, tiːen^{213}, liːeŋ^{55}, liːem^{35}, diːep^{21} /iːʊ/: ʨiːʊ^{21}, tiːʊ^{55}, liːʊ^{35}, liːʊ^{55}, diːʊ^{55} (/uik̺/): luik̺^{35}, huik̺^{35} (/uin̺/): huin̺^{21}, kuin̺^{21}	/ie/: iet^{35}, ien^{35}, iem^{25}, ien^{55}
/ɛ/ - /e/	(/uek̺/): tuek̺^{35}, kuek̺^{35} (/uen̺/): suen̺^{21}	/eʊ/: keʊ^{55}, neʊ^{35}, teʊ^{35}, keʊ^{213}, deʊ^{213}
/əː/ - /ə/	/əːɪː/: ʦəːɪː^{213}, təːɪː ^{35}, dəːɪː^{55}, ʃəːɪː^{55}, ləːɪː^{55}	/əɪ/: bəɪ^{213}, bəɪ^{35}, ʃəɪ^{35}, təɪ^{25}, jəɪ^{55} /əʊ/: kəʊ^{55}, bəʊ^{213}, kəʊ^{35}, səʊ^{35}, nəʊ^{35} /uəʊ/: /Øuəʊ^{35}/
/ɑː/ - /ɑ/	/ɑːɪː/: tɑːɪː^{55}, ʃɑːɪː^{55}, mɑːɪː^{55}, kɑːɪː^{55}, lɑːɪː^{55} /ɑːɪ/: kɑːɪ^{55}, ʦɑːɪ^{35}, ʃɑːɪ^{55}, ʦɑːɪ^{21}, tɑːɪ^{55} /ɑːʊ/: kɑːʊ^{55}, ʃ ɑːʊ^{35}, ɳɑːʊ^{55}, tɑːʊ^{55}, lɑːʊ^{55} /uɑːʊ/: kuɑːʊ^{21} /ɔɑːɪː/: tɔɑːɪː^{21}, tɔɑːɪː^{35}, kɔɑːɪː^{55}, ʃɔɑːɪː^{35} /ɔɑːɪ/: sɔɑːɪ^{55}, lɔɑːɪ^{55}, hɔɑːɪ^{55}	

	/uɑ:ɪ:/: Øuɑ:ɪ:^{55} /uɑ:ɪ/: Øuɑ:ɪ^{55}	
/ɔ/ - /o/	/ɔɪ:/: tɔɪ:^{2314}, kɔɪ:^{55}, ʃɔɪ:^{55}, tɔɪ:^{2314}, hɔɪ:^{35}, ʃɔɪ:^{35}	/oɪ:/: ʦoɪ:^{55}, tʰoɪ:^{55}, toɪ:^{55}, soɪ:^{55}, ʃoɪ:^{55}
/u/ - /ɯ/	/uɪ/: tuɪ^{35}, suɪ^{35}, tuɪ^{55}, luɪ^{55}, kuɪ^{55}, ʦ̪uɪ^{35} /uɪ:/: tui^{55}, ʃui^{55}, tui^{25}, lui^{25}, gui^{55} /ue/: hue^{21}, tue^{35}, tue^{35}, hue^{35}, sue^{35}	/ɯɪ:/: gɯɪ:^{2314}, ŋɯɪ:^{2314}, hɯɪ:^{2314} /ɯʊ/: kɯʊ^{35}, kɯʊ^{21}, kɯʊ^{35}, ʃɯʊ^{2314}, tɯʊ^{2314}

Table 3. Thai #TʰV, #TV, #DV, #(C)VD (18 monothrongs in 9 pairs of vowel length contrast)

Minimal pairs	#TʰV, #TV		#DV	#(C)VD		
	Long vowel	Short vowel	Short vowel	Long vowel	Short vowel	
/a:/ - /a/	/pʰa :^{45}/ /tʰa:^{45}/ /kʰa:^{45}/	/pʰa^{45}/, /pa^{21}/ /tʰa^{45}/, /ta^{21}/ /kʰa^{45}/, /ka^{33}/	/ba^{21}/ /da^{21}/ /ga^{21}/	/a:e$^{33/45}$/ /a:ʊ^{55}/	- -	/aɪ^{21}/ /aʊ^{21}/
				/a:m$^{14/21/41}$/ /a:n$^{14/21/33/41/45}$/ /a:ŋ$^{14/21/41}$/	/ab^{21}/ /ad^{21}/ /ag^{21}/	/am^{21}/ /an^{21}/ /aŋ^{21}/
/e:/ - /e/	/pʰe:^{45}/ /tʰe:^{45}/ /kʰe:^{45}/	/pʰe^{45}/, /pe^{21}/ /tʰe^{45}/, /te^{21}/ /kʰe^{45}/, /ke^{33}/	/be^{21}/ /de^{21}/ /ge^{21}/	- /e:ʊ^{55}/	- -	- /eʊ^{21}/
				/e:m$^{14/21/41}$/ /e:n$^{14/21/33/41/45}$/ /e:ŋ$^{14/21/41}$/	/eb^{21}/ /ed^{21}/ /eg^{21}/	/em^{21}/ /en^{21}/ /eŋ^{21}/
/ɯ:/ - /ɯ/	/pʰɯ:^{45}/ /tʰɯ:^{45}/ /kʰɯ:^{45}/	/pʰɯ^{45}/, /pɯ^{21}/ /tʰɯ^{45}/, /tɯ^{21}/ /kʰɯ^{45}/, /kɯ^{33}/	/bɯ^{21}/ /dɯ^{21}/ /gɯ^{21}/	/ɯ:m$^{14/21/41}$/ /ɯ:n$^{14/21/33/41/45}$/ /ɯ:ŋ$^{14/21/41}$/	/ɯb^{21}/ /ɯd^{21}/ /ɯg^{21}/	/ɯm^{21}/ /ɯn^{21}/ /ɯŋ^{21}/
/u:/ - /u/	/pʰu:^{45}/ /tʰu:^{45}/ /kʰu:^{45}/	/pʰu^{45}/, /pu^{21}/ /tʰu^{45}/, /tu^{21}/ /kʰu^{45}/, /ku^{33}/	/bu^{21}/ /du^{21}/ /gu^{21}/	/u:e$^{33/45}$/ -	- -	- -
				/u:m$^{14/21/41}$/ /u:n$^{14/21/33/41/45}$/ /u:ŋ$^{14/21/41}$/	/ub^{21}/ /ud^{21}/ /ug^{21}/	/um^{21}/ /un^{21}/ /uŋ^{21}/
/ɛ:/ - /ɛ/	/pʰɛ:^{45}/ /tʰɛ:^{45}/ /kʰɛ:/	/pʰɛ^{45}/, /pɛ^{21}/ /tʰɛ^{45}/, /tɛ^{21}/ /kʰɛ^{45}/, /kɛ^{33}/	/bɛ^{21}/ /dɛ^{21}/ /gɛ^{21}/	/ɛ:e$^{33/45}$/ /ɛ:ʊ^{55}/		
				/ɛ:m$^{14/21/41}$/ /ɛ:n$^{14/21/33/41/45}$/ /ɛ:ŋ$^{14/21/41}$/		
/æ:/ - /æ/	/pʰæ:^{45}/ /tʰæ:^{45}/ /kʰæ:^{45}/	/pʰæ^{45}/, /pæ^{21}/ /tʰæ^{45}/, /tæ^{21}/ /kʰæ^{45}/, /kæ^{33}/	/bæ^{21}/ /dæ^{21}/ /gæ^{21}/	- /æ:ʊ^{55}/		
				/æ:m$^{14/21/41}$/ /æ:n$^{14/21/33/41/45}$/ /æ:ŋ$^{14/21/41}$/		

<table>
<tr><td rowspan="2">/ɔ:/ - /ɔ/</td><td rowspan="2">/pʰɔ:^{45}/
/tʰɔ:^{45}/
/kʰɔ:^{45}/</td><td rowspan="2">/pʰɔ^{45}/, /pɔ^{21}/
/tʰɔ^{45}/, /tɔ^{21}/
/kʰɔ^{45}/, /kɔ^{33}/</td><td rowspan="2">/bɔ^{21}/
/dɔ^{21}/
/gɔ^{21}/</td><td>/ɔ:e$^{33/45}$/
/ɔ:u/</td><td>-
-</td><td>-
-</td></tr>
<tr><td>/ɔ:m$^{14/21/41}$/
/ɔ:n$^{14/21/33/41/45}$/
/ɔ:ŋ$^{14/21/41}$/</td><td>/ɔb^{21}/
/ɔd^{21}/
/ɔg^{21}/</td><td>/ɔm^{21}/
/ɔn^{21}/
/ɔŋ^{21}/</td></tr>
<tr><td rowspan="2">/o:/ - /o/</td><td rowspan="2">/pʰo:^{45}/
/tʰo:^{45}/
/kʰo:^{45}/</td><td rowspan="2">/pʰo^{45}/, /po^{21}/
/tʰo^{45}/, /to^{21}/
/kʰo^{45}/, /ko^{33}/</td><td rowspan="2">/bo^{21}/
/do^{21}/
/go^{21}/</td><td>/o:e$^{33/45}$/
-</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>/o:m$^{14/21/41}$/
/o:n$^{14/21/33/41/45}$/
/o:ŋ$^{14/21/41}$/</td></tr>
<tr><td rowspan="2">/ɤ:/ - /ɤ/</td><td rowspan="2">/pʰɤ:^{45}/
/tʰɤ:^{45}/
/kʰɤ:^{45}/</td><td rowspan="2">/pʰɤ^{45}/, /pɤ^{21}/
/tʰɤ^{45}/, /tɤ^{21}/
/kʰɤ^{45}/, /kɤ^{33}/</td><td rowspan="2">/bɤ^{21}/
/dɤ^{21}/
/gɤ^{21}/</td><td>/ɤ:e$^{33/45}$/
-</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>/ɤ:m$^{14/21/41}$/
/ɤ:n$^{14/21/33/41/45}$/
/ɤ:ŋ$^{14/21/41}$/</td></tr>
<tr><td></td><td>(一) p.18</td><td>(二) p.10</td><td>(二) p.10</td><td>(一) p.44</td><td colspan="2">(二) p.25</td></tr>
<tr><td></td><td>reannotate</td><td>reannotate</td><td>reannotate</td><td>reannotate</td><td colspan="2"></td></tr>
</table>

Script name: TB-Obstruent analysis.praat

Form 2 is used to provide the labels you have used to code the individual
variants that you want the script to measure. I use the following pattern:
Pre-phase labels:
-pa: preaspirated
-pv: prevoiced
-pvl: prevoiceless
-pta: postaspirated
-ptv: postvoiced (-m, -n, -ŋ)
-ptvl: postvoiceless

Release phase labels:

- af: affricated

- as: aspirated

- ej: ejective

- el: elided

- fr: partly or fully fricated

- gs: glottalled/glottalised

- ua: unaspirated

- ur: unreleased

用生成對抗網路（GANs）來輸入音檔並生成新的音檔，主要流程如下：

1. 資料準備

- 收集音檔資料集（如語音、音樂等）。
- 將音檔轉換為適合神經網路處理的格式，常見做法有：
  - 轉為梅爾頻譜圖（Mel-spectrogram）
  - 轉為短時傅立葉轉換（STFT）頻譜
  - 直接用原始波形（Raw waveform）

2. 建立 GAN 模型

- GAN 包含生成器（Generator）和判別器（Discriminator）。
- 針對音訊資料，常見的 GAN 架構有：
  - WaveGAN：直接生成原始音訊波形
  - SpecGAN：生成頻譜圖，再用反轉換還原音訊
  - MelGAN、GAN-TTS 等：用於語音合成

3. 訓練模型

- 將音訊資料（如頻譜圖）輸入判別器，讓生成器學習產生類似真實音訊的資料。
- 反覆訓練直到生成器能產生高品質的音訊。

4. 生成音訊

- 訓練完成後，輸入隨機噪聲或特定條件（如語者特徵、音樂風格等）給生成器。
- 生成器輸出音訊資料（如頻譜圖或波形）。
- 若輸出為頻譜圖，需用 Griffin-Lim 等演算法將其還原為音訊波形。

簡單範例流程（以 SpecGAN 為例）：

1. 將音檔轉為梅爾頻譜圖（可用 librosa 庫）。

2. 用頻譜圖訓練 GAN 模型（可用 TensorFlow 或 PyTorch）。
3. 用訓練好的生成器產生新的頻譜圖。
4. 用 librosa 的 inverse 方法將頻譜圖還原成音檔（.wav）。

參考資源：

- WaveGAN: https://github.com/chrisdonahue/wavegan
- SpecGAN: https://github.com/chrisdonahue/specgan
- MelGAN: https://github.com/descriptinc/melgan-neurips

當你提到「GANs 打開音檔的 code」，我理解你可能想用 GANs（生成對抗網絡）來處理音訊檔案，或是想用 Python 讀取音檔作為 GANs 的輸入。這裡我先提供如何用 Python 打開音檔（如 .wav 檔），並將其轉為適合 GANs 處理的格式。
常用的音訊處理套件有 librosa 和 scipy。以下是範例程式碼：

1. 安裝必要套件：

```bash
pip install librosa numpy
```

2. 讀取音檔並轉為 numpy array：

```python
import librosa
import numpy as np

# 讀取音檔
audio_path = 'your_audio_file.wav'
y, sr = librosa.load(audio_path, sr=None)  # y 是音訊資料, sr 是取樣率

print("音訊資料 shape:", y.shape)
print("取樣率:", sr)
```

3. 若要將音訊轉為 GAN 常用的頻譜圖（如梅爾頻譜圖）：

```python
import librosa.display
```

```
import matplotlib.pyplot as plt


# 產生梅爾頻譜圖
S = librosa.feature.melspectrogram(y=y, sr=sr, n_mels=128)
S_dB = librosa.power_to_db(S, ref=np.max)


plt.figure(figsize=(10, 4))
librosa.display.specshow(S_dB, sr=sr, x_axis='time', y_axis='mel')
plt.colorbar(format='%+2.0f dB')
plt.title('Mel-frequency spectrogram')
plt.tight_layout()
plt.show()
```

這樣你就可以把音檔轉成 GANs 可以處理的 numpy array 或頻譜圖了。

如果你想要更進一步，像是用 GANs 生成音訊，歡迎再補充你的需求！