# Oral Movement Similarities between [i] vs. [Λ] Word Articulation and Emotional Expressions Explain the Gleam-Glum Effect

by

Shin-Phing Yu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2023 by the
Graduate Supervisory Committee:

Michael K. McBeath, Chair
Arthur M. Glenberg
Aurel Coza
Marco Santello
Greg O. Stone

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

This project investigates the gleam-glum effect, a well-replicated phonetic emotion association in which words with the [i] vowel-sound (as in “gleam”) are judged more emotionally positive than words with the [Λ] vowel-sound (as in “glum”). The effect is observed across different modalities and languages and is moderated by mouth movements relevant to word production. This research presents and tests an articulatory explanation for this association in three experiments. Experiment 1 supported the articulatory explanation by comparing recordings of 71 participants completing an emotional recall task and a word read-aloud task, showing that oral movements were more similar between positive emotional expressions and [i] articulation, and negative emotional expressions and [Λ] articulation. Experiment 2 partially supported the explanation with 98 YouTube recordings of natural speech. In Experiment 3, 149 participants judged emotions expressed by a speaker during [i] and [Λ] articulation. Contradicting the robust phonetic emotion association, participants judged more frequently that the speaker’s [Λ] articulatory movements were positive emotional expressions and [i] articulatory movements were negative emotional expressions. This is likely due to other visual emotional cues not related to oral movements and the order of word lists read by the speaker. Findings from the current project overall support an articulatory explanation for the gleam-glum effect, which has major implications for language and communication.

DEDICATION

To my family and friends.

# ACKNOWLEDGMENTS

I would like to thank my advisors, Dr. Michael K. McBeath and Dr. Arthur M. Glenberg, for encouraging me and supporting me in exploring my potential, enabling me to step out of my comfort zone during my training in this program. I would also like to thank the committee members for their patience, guidance, and support; especially Dr. Aurel Coza and his lab members for software support. I would like to thank the capstone team members, especially Matthew Watson, for major contributions to this project. Last but not least, my family and friends, especially my husband Andrew, I thank you for your care and empowerment, which gave me buffs to obtain this achievement.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

## SAY "CHEESE": MOUTH MOVEMENTS DURING [i] VS [Λ] PRODUCTION ARE SIMILAR TO POSITIVE VS NEGATIVE EMOTIONAL EXPRESSIONS

Saying “cheese” is a common technique used to force a grin when taking pictures. It was first referenced in a newspaper article quoting U.S. ambassador Joseph E. Davis (“Need to Put on A Smile”, 1943), who remarked it as a technique to trigger an automatic smile, a political technique of appearing pleasant. This is because the long [i] vowel sound forced the corners of the lips to stretch upward. Besides its photogenic use, saying “cheese” also seems to provide linguistic insights into mouth movement-induced phonetic emotion associations.

Several phonetic emotion associations have been documented across multiple languages (Adelman et al., 2018; Auracher et al., 2010; Slavova, 2019; Whissell, 1999). For example, in a study where German participants read and rated cartoons, participants who repeatedly pronounced the [i] vowel-sound rated the cartoons funnier compared to participants who repeatedly pronounced [o] (as in German word “*tod*” which means “dead”) (Rummer et al., 2014). In another study by Rummer and Schweppe (2019), faces with positive facial expressions were given names with [i] more frequently than those with neutral or negative facial expressions, and faces with negative facial expressions were given names with [o] more frequently than those with positive facial expressions.

I recently uncovered similar phonetic emotion associations in English and Mandarin called the *gleam-glum* effect, where participants judge words and pseudowords with the [i] vowel-sound as more emotionally positive compared to those with the [Λ] vowel-sound (as in “glum”) (Yu et al., 2020; Yu et al., 2021a; Yu et al., 2021b; McBeath et al., 2021; McBeath et al., 2019). This phonetic emotion association is highly robust. It has been replicated across sensory modalities (i.e., written and aural presentation of words) and judgment methodology (i.e., rate words and pseudowords along a valence scale, match pseudowords with happy versus sad illustrations, and match pseudowords with English words to assign meaning). Furthermore, the phonetic emotion association encompasses the entire English lexicon of all 3,329 English words (Warriner et al., 2013), regardless of the number of syllables in a word, and is generalizable to Mandarin PinYin (Yu et al., 2021a). See Figure 1 for a chart summarizing the findings described above.

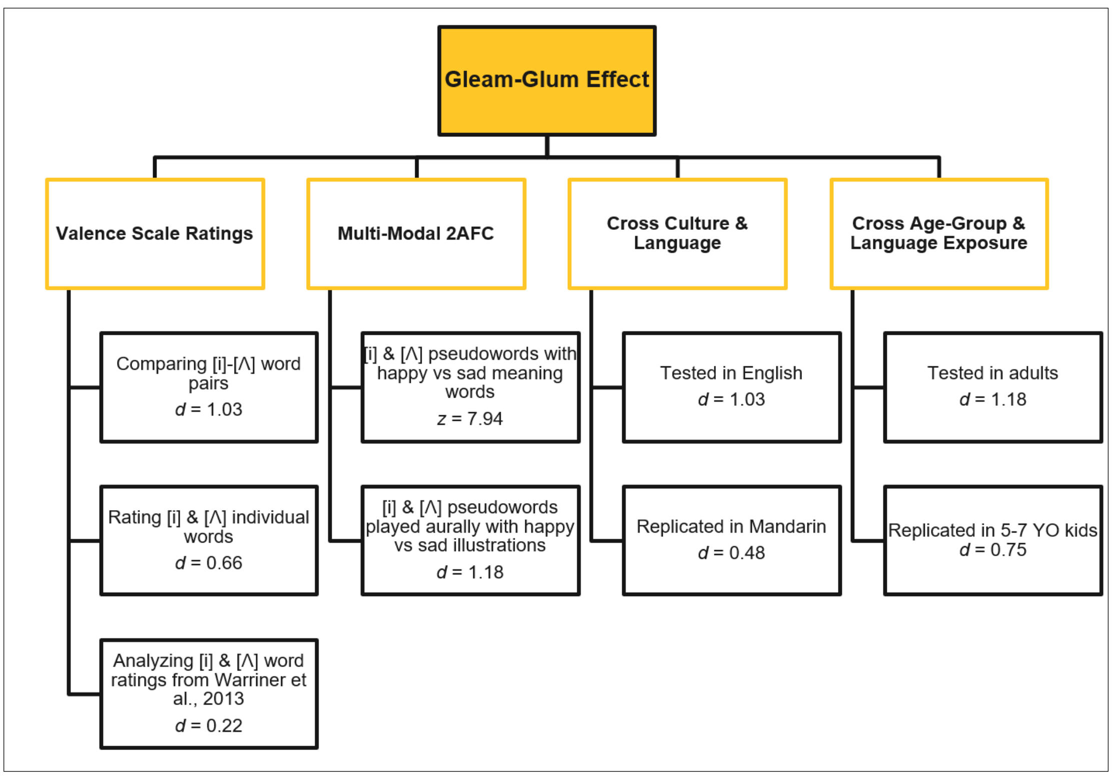

Figure 1. Chart summarizing main gleam-glum effect findings.

Currently, there is no agreed-upon explanation of the mechanisms underlying phonetic emotion associations ( Sidhu & Pexman, 2018). Some studies have supported an acoustic explanation (Aryani et al., 2018; Kawahara & Shinohara, 2012). For example, an acoustic explanation for the *gleam-glum* effect suggests that [i] is typically heard and said with a higher pitch than [Λ], and because a high pitch is associated with a more positive valence than a lower pitch, [i] is judged as more positive than [Λ]. Other studies support an articulatory explanation (Garrido et al., 2021; Körner & Rummer, 2021; Rummer et al., 2014). Specifically, regarding the *gleam-glum* effect, an articulatory explanation could be that the zygomaticus major muscle, whose contraction is prototypical of happy expressions (Ekman et al., 2002b), is likely contracted when articulating [i], and the orbicularis oris muscle, whose contraction is prototypical of angry expressions (Ekman et al., 2002b), is likely contracted when articulating [Λ]. A previous finding supports the articulatory explanation for the *gleam-glum* effect, showing that facial movements moderate *gleam-glum* effect size. Participants chewing gum during word rating tasks showed an attenuated *gleam-glum* effect compared to participants reading words aloud during word rating tasks (Yu et al., 2021a). Additionally, another study showed that valence ratings for vowel sound [y] (as in the first syllable of German word "über") relative to [i] and [o] support the articulatory explanation over the auditory one (Körner & Rummer, 2021). Auditorily, the vowel sound [y] is similar to [i] in pitch height, but articulatorily, it is similar to the vowel sound [o] in its contraction of the orbicularis oris. Supporting the articulatory explanation, participants across the four experiments judged [i] to be more positive than both [y] and [o], and [y] was not judged to be more positive than [o]. Nonetheless, acoustic and articulatory explanations are likely interdependent rather than mutually exclusive (e.g., Arias et al., 2018; Sidhu & Pexman, 2018; Whissell, 2000). This debate is beyond the scope of the current project, and the focus on the articulatory explanation of the *gleam-glum* effect does not imply a stance against the acoustic explanation. In fact, a recent study from the lab investigated the differences in acoustic properties between [i] and [Λ] (Patten & McBeath, 2020).

The focus of the current project is to test the articulatory explanation of the *gleam-glum* effect. Mouth movements during articulation and emotional expressions are compared objectively (by analyzing oral landmark movement similarities) and subjectively (by collecting participant judgments). In Experiments 1 and 2, I test the hypothesis (H1) that mouth movements when articulating [i] words will be more similar to those of positive emotional expressions, while mouth movements when articulating [Λ] words will be more similar to those of negative emotional expressions. In Experiment 3, I test the hypothesis (H2) that that participants will judge the speaker as expressing a positive emotion more frequently when the speaker is saying an [i] word, while participants will judge the speaker as expressing a negative emotion more frequently when the speaker is saying an [Λ] word.

**Emotion Models**

The current project investigates oral movements during expressions of Darwinian basic emotions: happiness, sadness, anger, fear, surprise, and disgust (1890/2009). However, these emotions are commonly referred to as Ekman's basic emotions. The decision to use the Darwinian emotional model is primarily because these emotions are among the most frequently investigated (Kreibig, 2010). Additionally, the facial expressions of the six emotions have been shown to be highly recognizable across a large number of cultures (Ekman & Friesen, 1971; Ekman et al., 1987; Ekman et al., 1969; Takarae et al., 2021). Alternative emotion models have been proposed, including Russell's Circumplex model (1980), which classifies emotions along the valence, arousal, and dominance dimensions, and Barrett's theory of constructed emotion (Barrett, 2017), where contextual cues, physiological cues, and past experiences are combined to predict the current emotion experienced. It is beyond the scope of this project to determine the best model for understanding emotions. In fact, other emotional models have been referenced in previous studies on the *gleam-glum* effect. For example, emotional ratings have been along the valence dimension of Russell's circumplex model (McBeath et al., 2019). The choice of emotional model depends on the specific needs of each study.

**Random Forest Models**

The objective approach used in Experiments 1 and 2 to assess similarities in oral movements relies on informative oral landmarks that differ systematically in their movements during emotional expressions and articulation of [i] and [Λ] words. The 10 oral landmarks used in this experiment were extracted from participant recordings at a frequency of approximately 30 Hz using the Affectiva facial expression recognition engine via iMotions Biometric Research Platform 9.1 software (iMotions, 2021). Figure 2 (taken from Toit et al., 2022) shows the 34 landmarks identified by the software for an entire face. Of interest to the current project are the 10 oral landmarks around the mouth (i.e., points 21 to 30).

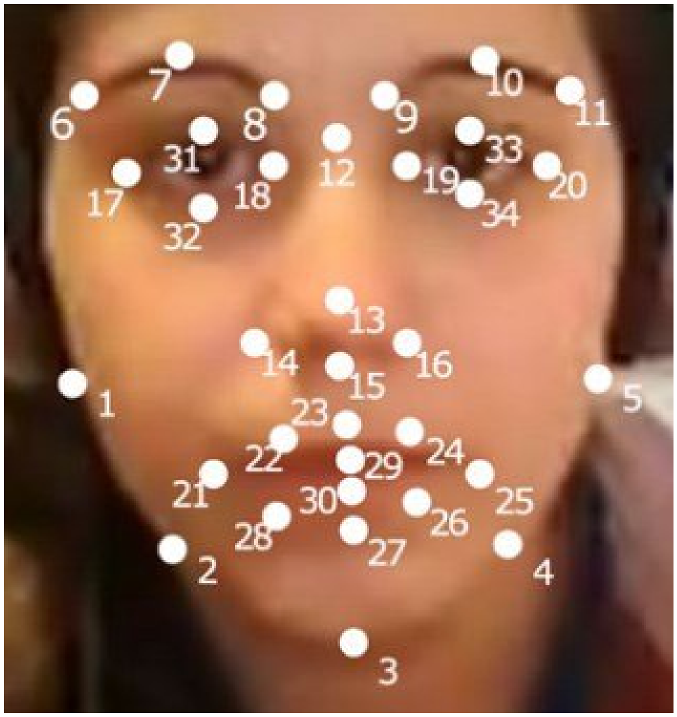

Figure 2. Facial landmarks identified by Affectiva for an entire face.

Previous studies have validated the facial landmarks used by the emotion recognition software, showing that its ability to identify facial expressions using movements of these landmarks is comparable to that of Electromyography (Kulke et al., 2020). In Experiments 1 and 2, the validity of these oral landmarks was further tested using the accuracy of the random forest classification models as an index. The random forest model is a type of predictive classification machine-learning model. It is based on the principle that data points from the same category are likely to meet a similar set of criteria, and data points from different categories differ in the criteria that they meet. The models in Experiments 1 and 2 used the standardized X and Y coordinates of the 10 oral landmarks to predict the eight task types (i.e., six emotional expressions and two word-reading tasks). Therefore, the accuracy of the classification models depends on whether the oral landmarks of interest are informative and differ systematically among different emotional expressions and between [i] and [Λ] articulation. The model is also useful for analyzing feature importance and providing oral landmarks whose movements contributed the most to systematic differences. These movements can then be used to infer oral muscles that contract similarly during emotional expressions and word articulation by referencing the Facial Action Coding System proposed by Ekman et al., (2002b).

Each random forest model consists of a large set of decision trees, and each decision tree has access to only a random subset of input data on which they are trained and tested. These decision trees independently predict the category from which a data point originates, and these predictions are tallied across decision trees to generate the final output of a random forest model. Each decision tree can be considered as nested if-then statements. Namely, each tree's prediction is based on a unique sequence of branches narrowing down the possible categories by checking whether the datapoint meets a specific criterion (i.e., binary "true" or "false" output). For example, a decision tree in Experiment 1 may have a starting node that checks if the Y-coordinate for point 21 is above a certain value, or if it is equal to or below the value. When the output is "true", the tree reaches the outcome category, which it then outputs. When the output is "false", the tree reaches the next node in a unique sequence and repeats until it reaches an output. It is common practice to randomly split the data into two subsets: a subset of data for model training (typically 80% of the data, as is the case for all models in this project) and the rest of the data for model evaluation. During the training phase, the model defines the split points for each continuous predictor (i.e., the X and Y coordinates for each facial landmark). These split points are optimized such that the data from the same predefined categories are the most homogeneous. The model performance is then evaluated using the remaining subset of the data. A model is considered to have adequate performance if its accuracy is higher than the no-information rate, that is, the accuracy rate that can be achieved without a model, based on pure randomness (Kuhn & Johnson, 2013). In the current experiment, the no-information rate for classifying emotions is 1/6 (approximately 17%) because there are six categories of emotions, and the no-information rate for classifying [i] versus [Λ] words is 1/2 (50%) because there are two categories of words. All code used for analyses in the current project can be found on https://github.com/mjwats10/Gleam-Glum-Emotional-Valence.

This project tests an articulatory explanation of the *gleam-glum* effect. The hypothesis (H1) is that mouth movements when articulating [i] words will be more similar to those of positive emotional expressions, whereas mouth movements when articulating [Λ] words will be more similar to those of negative emotional expressions. In Experiment 1, the participants recorded themselves completing an emotional recall task and a word read-aloud task. Oral landmark movement similarities were assessed objectively by analyzing the Euclidean distances among eight clusters, representing the six emotional expressions and two types of words read aloud. Specifically, I make four predictions. I predict (P1) that the Euclidean distance will be shorter between the happy expression (the only positive emotion from the Darwinian basic emotions) cluster mean and the [i] articulation cluster mean than the [Λ] articulation cluster. Similarly, I predict (P2) that the Euclidean distance will be shorter between the cluster mean of a negative emotion and the [Λ] articulation cluster mean than the [i] articulation cluster. I also predict (P3) that within word-type, the Euclidean distance will be shorter between the [i] articulation cluster mean and the cluster means of positive and neutral emotions than cluster mean to negative emotions, and (P4) the Euclidean distance will be shorter between the [Λ] articulation cluster mean and the cluster means of negative emotions than the cluster means of positive and neutral emotions. Experiment 2 tested the same hypothesis using naturalistic data. In Experiment 2, recordings from the Experiment 1 emotional recall task were reused, while recordings of the word read-aloud task were replaced with YouTube video segments of [i] and [Λ] words being said. Oral landmark movement similarities were again objectively assessed, and the same predictions were made. Experiment 3 tested whether (subjective) human judgment also showed oral landmark movement similarities predicted in the first two experiments. In Experiment 3, participants sorted a stack of GIF moving images of a speaker saying [i] and [Λ] words or pseudowords, by which of the six basic emotions the speaker seems to be expressing. I predict that participants will judge the speaker as expressing happiness more frequently when the speaker is saying an [i] word, while participants will judge the speaker as expressing a negative emotion more frequently when the speaker is saying an [Λ] word. Thus, the current project is the first to directly test the articulatory explanation of the *gleam-glum* effect by providing objective and subjective comparisons of mouth movements during emotional expressions and word articulation.

# CHAPTER 2

# EXPERIMENT 1

Experiment 1 objectively compares mouth movements during [i] and [Λ] articulation with those during positive and negative emotional expressions. The hypothesis (H1) is that oral movements when articulating [i] words will be more similar to those of happy expressions, whereas oral movements when articulating [Λ] words will be more similar to those of negative emotional expressions.

## Methods

The Arizona State University Institutional Review Board approved all Experiment 1 protocols. To test the predicted oral movement similarities between emotional expressions and word articulation, recordings were collected from participants completing an emotional recall task and a word read-aloud task. I then validated the 10 oral landmarks by examining the accuracy of two random forest classification models that use the X and Y coordinates of these oral landmarks as predictors, one that predicts the emotion expressed and another that predicts the type of word articulated. I then calculated the Euclidean distance between the eight cluster means (expression of six types of emotions and articulation of two types of words).

### *Participants*

All participants provided their consent before starting the experiment. The participants were 71 Arizona State University undergraduate students taking an introductory psychology course (ages 18-24 years old, $M_{age}$ = 19.08, *SD* = 1.41; 38 females, 32 males, 1 preferred not to say). Because a data point is generated approximately every 33 milliseconds, an N of 71 generates a large dataset that is sufficient for machine learning analyses (e.g., a 5-second recording generates 151 data points). None of the participants reported any reading, speaking, or hearing disabilities. See Table 1 for the other demographic information.

**Table 1**

*Experiment 1 Participant Demographic Information*

| Demographics | Count for each Demographic Category | | | | | |
|---|---|---|---|---|---|---|
| Racial Ethnicity | White | Hispanic/ Latino/ Spanish Origin | Two or more races | Asian | Native Hawaiian/ Pacific Islandar | Black/ African American |
| | 41 | 11 | 11 | 5 | 2 | 1 |
| Native Language | English | Spanish | Others | | | |
| | 63 | 3 | 5 | | | |
| Languages Spoken | Monolingual | Bilingual | Trilingual | Quadrilingual | | |
| | 49 | 34 | 9 | 8 | | |
| English Spoken | > 10 Years | 5-10 Years | 0-4 Years | | | |
| | 69 | 1 | 1 | | | |

### *Materials*

The participants completed the online experiment using their own electronic devices in the private space of their choice. The Lookback website (2022) guided the participants to check their microphone, webcam, and screen sharing. The website also recorded participants as they completed emotional recall and word read-aloud tasks by following instructions from the Qualtrics website (2022). The recordings were then imported into the iMotions Biometric Research Platform 9.1 software, which extracted X and Y coordinates of 10 oral landmarks approximately every 33 milliseconds using the Affectiva facial expression recognition engine (iMotions, 2022).

**Word-Reading Task.** The word-reading task consisted of 507 words across six blocks:132 monosyllabic [i] words, 167 monosyllabic [Λ] words, 50 monosyllabic [i] pseudowords, 50 monosyllabic [Λ] pseudowords, 54 monosyllabic [æ] words, and 54 monosyllabic [u] words (see Appendix A for a list of all the words).

***Monosyllabic English Words.*** A comprehensive list of English monosyllabic words with either the [i] or [Λ] vowel sounds was obtained from the English Lexicon Project (Balota et al., 2007). To ensure that participants pronounced the words as intended, I excluded words with multiple valid pronunciations based on the recommended criteria proposed by Stone et al. (1997), such as the frequency of a spelling body (i.e., the vowel and any ending consonants of a word) appearing in English words. To reduce task effort and duration, I removed any variations in the words, preserving only their simple present tense. I further excluded words that were names (capitalized in the database) and words that included an apostrophe.

The 54 filler word pairs with [æ] and [u] vowel sounds were randomly selected from a previously used list (McBeath, 2019).

***Monosyllabic Pseudowords.*** The exhaustive monosyllabic pseudoword pairs were generated for use in a previous study (Yu et al., 2020a). Each pseudoword pair shares the same consonant frames, for example, "gleap" and "glup." To create these pairs, we identified all yoked pairs of spelling-bodies in English that include the [i] and [Λ] vowel sounds, such as "eap" and "up" (Ziegler et al., 1997). We then included spelled bodies with only one possible pronunciation (Stone et al., 1997). This was done to ensure that participants pronounced the pseudowords as intended. We then attached each initial consonant to each spelling body in order to create a list of CVC pseudowords. Finally, we eliminated items with possible meanings (e.g., pseudo-homophones “keap” and slang “yeet”). The resulting list included 50 yoked pairs of monosyllabic CVC pseudowords that were identical, except for the central vowel (e.g., “gleap” and “glup”).

**Video Post-processing.** The iMotions Biometric Research Platform 9.1 software (iMotions, 2021) was used to capture the X and Y coordinates of 10 oral landmarks at a frequency of approximately 30 Hz. I standardized the X and Y coordinates from the absolute pixel space to the standard deviation distances from the center of the face. I removed outlier coordinates that may have resulted from irrelevant behavior, such as sudden head movements, by excluding coordinates that were more than 4.18 standard deviations away from the mean distances. This specific value of standard deviation was chosen because it provided an optimal balance between the number of data points lost and the number of outliers accounted for.

### *Procedure*

Participants were recorded as they completed the emotional recall task and the word read-aloud task. The emotional recall task consisted of seven trials. The first trial served as a practice in which the participants described a proud memory. The remaining six trials corresponded to the six basic Darwinian emotions of happiness, sadness, anger, surprise, disgust, and fear (Darwin 1890/2009) and were presented in random order (see Figure 3). Each trial began with reminders for participants to relax, describe memories that they felt comfortable sharing, and progress at their own pace. The emotion literature supports the use of the autobiographical recall method, which is effective in eliciting emotional expressions for all six emotions of interest (Lane et al., 2009; Rainville et al., 2006; Siedlecka & Denson, 2019).

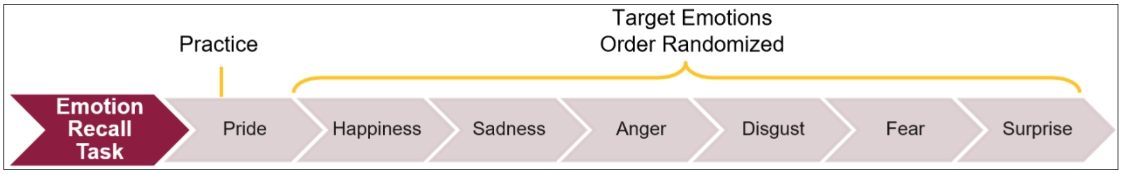


Figure 3. Progression of emotion recall task.

The word-reading task consisted of six blocks with a total of 507 words and pseudowords. The presentation order of words was randomized both across and within blocks (see Figure 4). Participants were instructed to pronounce the words clearly, make as few mistakes as possible, and avoid repeating them. They proceeded at their own pace and were encouraged to take breaks between the blocks.

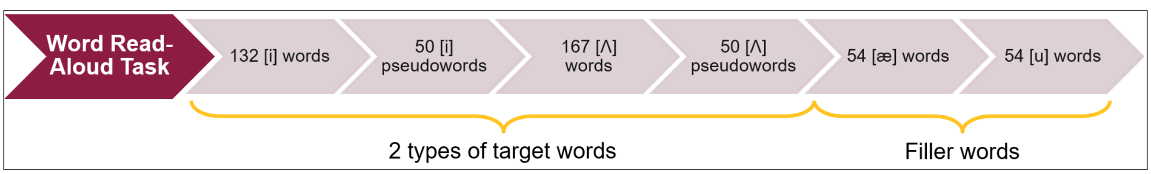


Figure 4. Progression of word read-aloud task.

## Results

### *Validating the Oral Landmarks*

To validate the 10 oral landmarks, the accuracy of two random forest classification models that used the X and Y coordinates of these oral landmarks as predictors was examined. Both models were run with five iterations, where the training and evaluation subsets were randomly resampled. The reported accuracy scores are the averages across these five iterations. The first model predicted the emotions expressed (i.e., happy, sad, anger, fear, disgust, or surprise) with an average accuracy of 60.62%, 95% CI [60.50%, 60.74%]. This accuracy score is considered high given that the no-information prediction rate for the six categories is 17% and that emotion recognition tasks typically use visual cues from entire faces. The second model predicted the type of articulated word (i.e., [i] or [Λ]) with an average accuracy of 80.98%, 95% CI [80.94%, 81.02%]. The accuracy of the two random forest models validated the use of the 10 oral landmarks, showing that they inform systematic oral movement differences among the different emotions expressed and between the two types of words articulated. Furthermore, horizontal stretches (i.e., X-coordinate movements) and vertical stretches (i.e., Y-coordinate movements) of the two corners of the mouth (points 21 and 25) were equally important for distinguishing emotions and distinguishing the word-type (see Table 2). This is consistent with the predictions from an articulatory explanation for the *gleam-glum* effect by possibly showing that the contraction of the zygomaticus major muscle and its antagonistic muscle, the orbicularis oris, are informative for both emotion and word-type classification.

**Table 2**

*The Six Most Important Predictors for the Two Random Forest Classification Models from Experiment 1*

| | Feature Importance | | | | | | |
|---|---|---|---|---|---|---|---|
| Emotion Classification | 25Y | 21X | 21Y | 25X | 30Y | 30X | 24X |
| Word-type Classification | 21X | 25Y | 25X | 28X | 21Y | 24X | 30X |

*Note.* The features are in the order of most to least importance from left to right.

### *Euclidean Distance Among Clusters*

I predict (P1) that the Euclidean distance will be shorter between the happy expression (the only positive emotion from the Darwinian basic emotions) cluster mean and the [i] articulation cluster mean than the [Λ] articulation cluster. Similarly, I predict (P2) that the Euclidean distance will be shorter between the cluster mean of a negative emotion and the [Λ] articulation cluster mean than the [i] articulation cluster. I also predict (P3) that within word-type, the Euclidean distance will be shorter between the [i] articulation cluster mean and the cluster means of positive and neutral emotions than cluster mean to negative emotions, and (P4) the Euclidean distance will be shorter between the [Λ] articulation cluster mean and the cluster means of negative emotions than the cluster means of positive and neutral emotions. (see Figure 5 for the eight clusters presented along two dimensions). Table 3 provides the Euclidean distance between the cluster means. Note that the means are 20-dimensional vectors calculated from each standardized X and Y coordinate of the 10 oral landmarks.

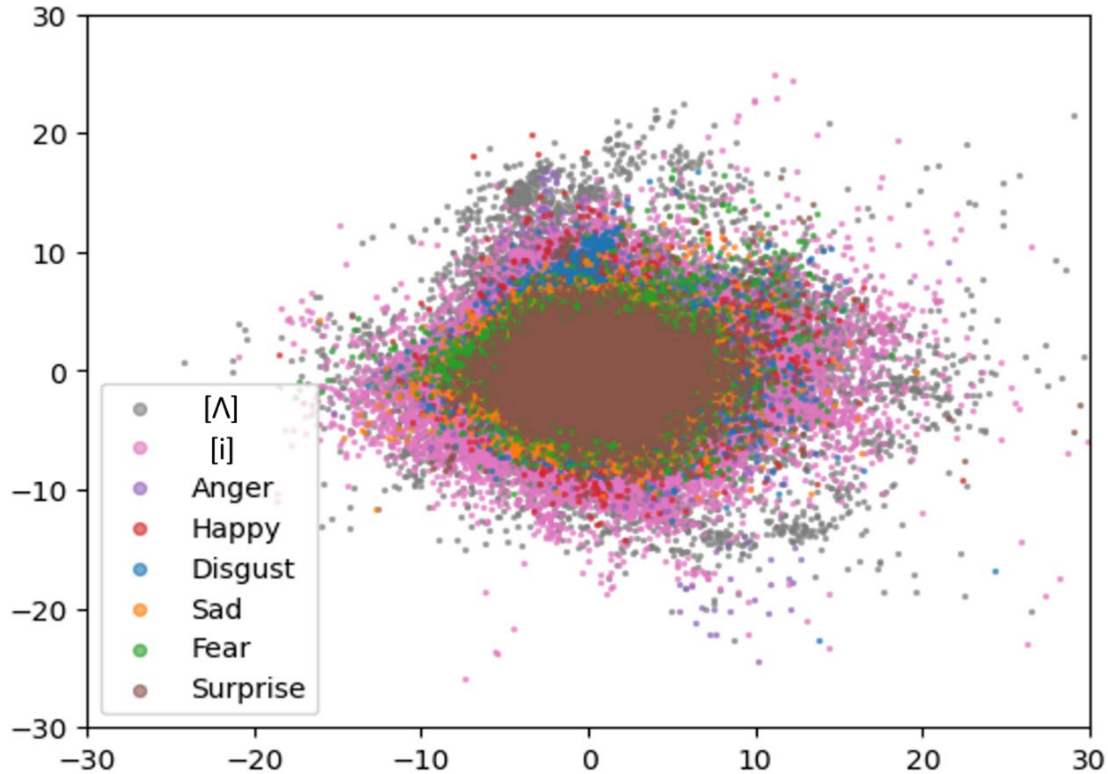


Figure 5. Experiment 1 clusters depicted in two dimensions after dimensionality reduction using principal components analysis.

Supporting P1 and P2, the Euclidean distance was shorter between the cluster means for happy expression oral movements and [i] articulatory oral movements (0.50) compared to [Λ] articulatory oral movements (0.83), and the Euclidean distance was shorter between the cluster means for negative emotional expression oral movements and [Λ] articulatory oral movements compared to [i] articulatory oral movements (see Table 3). Consistent with P3, albeit non-significant, [i] articulatory movements were closer to positive and neutral emotions ($M_{Happy,Surprise}$ = 0.48, *SD* = 0.03) than negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 0.54, *SD* = 0.10), $t(3.93) = 1.17$, $p$ = N.S. Supporting P4, [Λ] articulatory movements were significantly closer to negative emotions ($M_{Sad,Anger,Fear,Disgust}$

= 0.33, *SD* = 0.16) than positive and neutral emotions ($M_{Happy,Surprise}$ = 0.82, *SD* = 0.02), $t(3.17) = 6.04$, $p < 0.01$, 95% CI [0.24, 0.73], $d = 6.78$.

**Table 3**

*Euclidiean Distance between Oral Landmark Cluster Means from Experiment 1*

| | Happy | Sad | Anger | Fear | Surprise | Disgust | [Λ] |
|---|---|---|---|---|---|---|---|
| [i] | 0.50 | 0.68 | 0.52 | 0.51 | 0.46 | 0.45 | 0.43 |
| [Λ] | 0.83 | 0.54 | 0.23 | 0.19 | 0.80 | 0.36 | -- |

*Note.* Because the coordinates were standardized, these values are unitless (distances from the center of the face in standard deviations).

To explore whether movements of the rest of the 24 facial landmarks show consistent patterns compared with oral movements, the same calculations were done including all facial landmarks (see Table 4 for Euclidean distance between the cluster means when all 34 facial landmarks are included).

**Table 4**

*Euclidiean Distance between Whole-Face Landmark Cluster Means from Experiment 1*

| | Happy | Sad | Anger | Fear | Surprise | Disgust | [Λ] |
|---|---|---|---|---|---|---|---|
| [i] | 4.89 | 3.81 | 4.27 | 4.24 | 3.94 | 4.20 | 0.58 |
| [Λ] | 4.79 | 3.80 | 4.16 | 4.20 | 3.91 | 4.15 | -- |

Taking into account whole face visual cues led to mixed results when comparing the articulatory and emotional expression landmark movements. The distance between the [Λ] articulation cluster mean and all emotion cluster means was overall shorter than the [i] articulation cluster mean. Within word type, [i] articulatory movements were non-significantly closer to negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 4.13, *SD* = 0.22) than positive and neutral emotions ($M_{Happy,Surprise}$ = 4.42, *SD* = 0.67), $t(1.10) = -0.59$, $p$ = N.S. Albeit non-significant, as predicted [Λ] articulatory movements were slightly closer to negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 4.08, *SD* = 0.19) than positive and neutral emotions ($M_{Happy,Surprise}$ = 4.35, *SD* = 0.62), $t(1.10) = 0.61$, $p$ = N.S.

The differing results obtained from analyzing only oral movements versus analyzing movements of the entire face suggest that the movements involved in speech production can produce oral expressions whose emotional valence do not match those conveyed by the rest of the face.

**Discussion**

In summary, the findings of Experiment 1 support H1, providing support for the articulatory explanation for the *gleam-glum* effect and its predictions of oral movement similarities between happy versus angry expressions and [i] versus [Λ] articulation. However, the articulatory oral movements analyzed in Experiment 1 may not represent those of natural speech behavior because they were collected from participants reading lists of words aloud. Experiment 2 extended the findings from Experiment 1 to test the articulatory explanation for the *gleam-glum* effect using naturalistic speech collected from YouTube videos.

CHAPTER 3

EXPERIMENT 2

Experiment 2 was an extension of Experiment 1, with a focus on natural speech behavior collected from YouTube recordings. The same hypothesis (H1) was tested in this experiment. Mouth movement similarities was again assessed objectively by analyzing the distances between cluster means.

## Methods

The Arizona State University Institutional Review Board approved all Experiment 2 protocols. To test the predicted oral movement similarities between emotional expressions and word articulation, the recordings of the emotional expression task from Experiment 1 were reused in Experiment 2. However, recordings of the word read-aloud task from Experiment 1 were substituted with 98 YouTube clips of people saying [i] or [Λ] words.

### *Materials*

**YouTube [i] and [Λ] Clips.** To expand the list of target words used in Experiment 1, I used the untrimmed version of the word list from the English Lexicon Project (Balota et al., 2007). A video scraper was built to randomly select YouTube videos that generated a set of pseudorandom YouTube links, resulting in 200 YouTube video search results. These videos were downloaded, and their transcripts were obtained to ensure that speakers said [i] or [Λ] target words. The transcripts also provided timestamps to sentences containing the target words, and clips were generated only with sentences that included the target words. The OpenCV frontal-face classifier (Bradski, 2000) was used to identify videos with 80% or more frames containing faces, and clips without faces were deleted. These clips were then manually trimmed using MoviePy (2017) to preserve the portion containing the target words. Due to technical difficulties during the import of these recordings to iMotions (2021) that led to excessive frame loss and due to time constraints, I elected to analyze 98 clips.

## Results

### ***Validating the Oral Landmarks***

Because the emotional expression data from Experiment 1 were reused, it was unnecessary to rerun the random forest model for emotion classification in Experiment 2. The random forest model for word-type classification again validated the 10 oral landmarks, achieving an average prediction accuracy of 98.61%, 95% CI[97.45%, 99.77%]. Regarding feature importance, Experiment 2 reproduced some of the Experiment 1 findings, showing that horizontal stretches (i.e., X coordinate movements) of the two corners of the mouth (points 21 and 25) were important for distinguishing the word-type in naturalistic speech (see Table 5). However, vertical stretches did not contribute much to the [i] and [Λ] oral movement differences. This is a potential signal that the [i] and [Λ] YouTube recordings differ systematically in other unintended ways that may affect the findings. Nonetheless, the feature importance here is still consistent with predictions from an articulatory explanation of the *gleam-glum* effect.

**Table 5**

*The Six Most Important Predictors for the Two Random Forest Classification Models from Experiment 2*

| | Feature Importance | | | | | | |
|---|---|---|---|---|---|---|---|
| Emotion Classification | 25Y | 21X | 21Y | 25X | 30Y | 30X | 24X |
| Word-type Classification | 28X | 25X | 21X | 26X | 24X | 22Y | 24Y |

*Note.* The features are in the order of most to least importance from left to right.

### *Euclidean Distance Among Clusters*

The predicted Euclidean distances between the cluster means remain (see Figure 6 for the visual presentation of the clusters and Table 6 for the list of the cluster means). Consistent with P1, the Euclidean distance was shorter between the cluster means for happy expression oral movements and [i] articulatory oral movements (0.86) compared to [Λ] articulatory oral movements (3.40). However, no definitive conclusions can be drawn about P1 and P2 as the distance between the [i] articulation cluster mean and all emotion cluster means was overall shorter than the [Λ] articulation cluster mean. Consistent with P3, albeit non-significant, [i] articulatory movements were closer to positive and neutral emotions ($M_{Happy,Surprise}$ = 0.73, $SD$ = 0.18) than negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 1.15, $SD$ = 0.22), $t$(2.49) = 2.53, $p$ = N.S. Supporting P4, [Λ] articulatory movements are on average closer to negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 2.90, $SD$ = 0.21) than positive and neutral emotions ($M_{Happy,Surprise}$ = 3.47, $SD$ = 0.10) ), $t$(3.95) = 4.44, $p < 0.05$, 95% CI [0.21, 0.92], $d$ = 4.47.

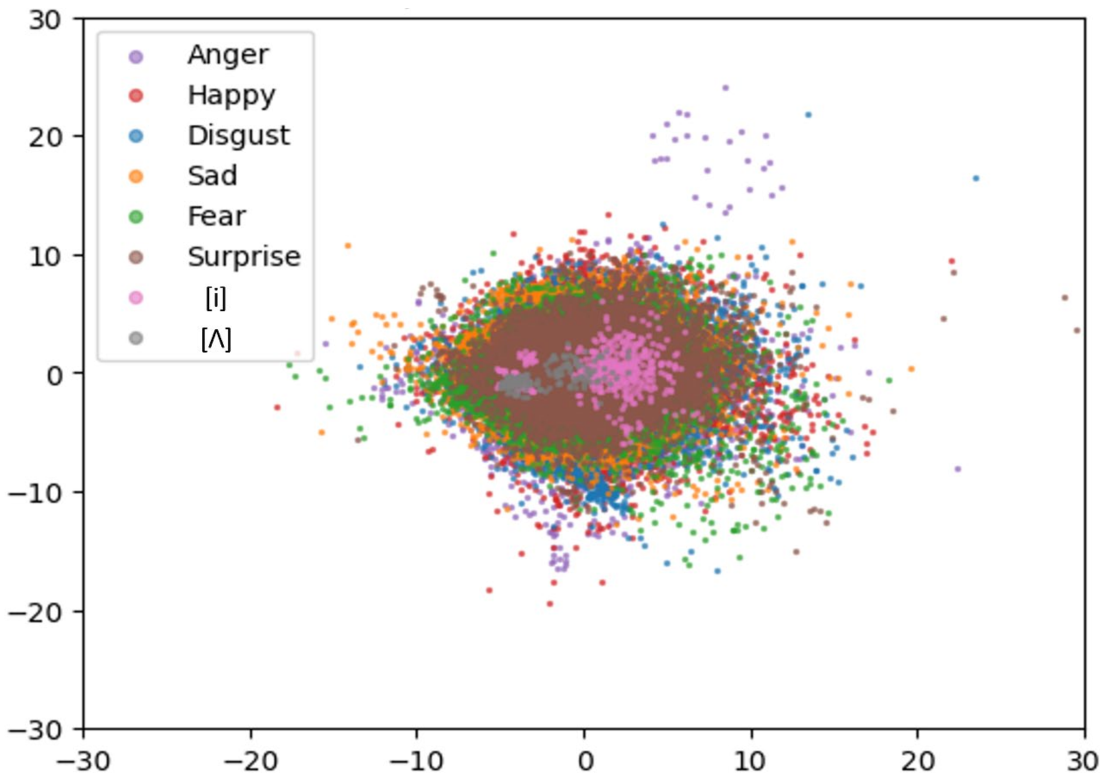

Figure 6. Experiment 2 clusters depicted in two dimensions after dimensionality reduction using principal components analysis.

**Table 6**

*Euclidiean Distance between Oral Landmark Cluster Means from Experiment 2*

| | Happy | Sad | Anger | Fear | Surprise | Disgust | [Λ] |
|---|---|---|---|---|---|---|---|
| [i] | 0.86 | 1.45 | 1.08 | 1.16 | 0.60 | 0.93 | 3.99 |
| [Λ] | 3.40 | 2.63 | 2.98 | 2.87 | 3.54 | 3.13 | -- |

To explore whether movements of the rest of the 24 facial landmarks show consistent patterns compared with oral movements, the same calculations were done including all facial landmarks (see Table 7 for Euclidean distance between the cluster means when all 34 facial landmarks are included).

**Table 7**

*Euclidiean Distance between Whole-Face Landmark Cluster Means from Experiment 2*

| | Happy | Sad | Anger | Fear | Surprise | Disgust | [Λ] |
|---|---|---|---|---|---|---|---|
| [i] | 5.87 | 4.57 | 5.01 | 5.02 | 4.49 | 4.92 | 8.24 |
| [Λ] | 4.82 | 5.67 | 5.54 | 5.57 | 6.17 | 5.68 | -- |

Similar to results from Experiment 1, when taking into account whole face visual cues, the outcome of comparing the articulatory and emotional expression landmark movements becomes washed out or even reversed. The Euclidean distance is shorter between the cluster means for happy expression oral movements and [Λ] articulatory oral movements (4.82) compared to [i] articulatory oral movements (5.87), and the Euclidean distance is shorter between the cluster means for negative emotional expression oral movements and [i] articulatory oral movements compared to [Λ] articulatory oral movements (see Table 7). Within word type, [i] articulatory movements were not significantly closer to negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 4.88, $SD$ = 0.21) than positive and neutral emotions ($M_{Happy,Surprise}$ = 5.18, $SD$ = 0.98), $t(1.05) = -0.43$, $p$ = N.S. Similarly, [Λ] articulatory movements were not significantly closer to positive and neutral emotions ($M_{Happy,Surprise}$ = 5.50, $SD$ = 0.95) than negative emotions ($M_{Sad,Anger,Fear,Disgust}$ = 5.62, $SD$ = 0.07), $t(1.01) = -0.18$, $p$ = N.S.

**Discussion**

In summary, Experiment 2 provided mixed results that partially support H1. These mixed results may be attributed to distinctions between [i] and [Λ] YouTube recordings irrelevant to the interests and goals of the current project. However, analyses on [Λ] articulatory movements focusing on oral landmarks support the articulatory explanation.

# CHAPTER 4

# EXPERIMENT 3

Experiment 3 tested whether the subjective judgment of a speaker's facial expressions while reading aloud a list of words would be consistent with the *gleam-glum* effect. I hypothesize (H2) that participants will judge the speaker as expressing happiness more frequently when the speaker is saying an [i] word, while participants will judge the speaker as expressing a negative emotion more frequently when the speaker is saying an [Λ] word.

## Methods

The Arizona State University Institutional Review Board approved all Experiment 3 protocols.

### *Participants*

All participants provided their consent prior to the experiment. A power analysis was conducted to estimate sample size required to find the response frequency difference between [i] and [Λ] word types across six emotion categories using a Pearson's Chi-square test. The analysis assumed a small to medium effect size of 0.35 ($df = 5$, $\alpha = 0.05$, $1-\beta = .8$), and the outcome indicated that a minimum of 105 participants is needed to test the effect. Note that after data collection, I decided to switch my analyses to paired-samples t-tests due to violated assumptions of response independence for a Pearson's Chi-square test of independence. In the end, 149 Arizona State University undergraduates taking an introductory psychology course participated in the experiment (ages 18-39 years old, $M_{age} = 19.01$, $SD = 2.44$; 91 females, 58 males; see Table 8 for other demographic information). Only one participant preferred not to disclose whether they have any reading disabilities, only one participant reported having a speaking disability, and only one participant reported having a hearing disability (these responses are from three separate participants).

**Table 8**

*Demographic Information of Experiment 3 Participants*

| Demographics | Count for each Demographic Category | | | | | | |
|---|---|---|---|---|---|---|---|
| Racial Ethnicity | White | Asian | Hispanic/ Latino/ Spanish Origin | Two or more races | Black/ African American | Native Hawaiian/ Pacific Islandar | American Indian/ Alaskan Native |
| | 85 | 32 | 16 | 10 | 3 | 2 | 1 |
| Native Language | English | Vietnamese | Spanish | Chinese | Cantonese | Telugu | Others |
| | 125 | 7 | 3 | 3 | 2 | 2 | 7 |
| Languages Spoken | Monolingual | Bilingual | Trilingual | | | | |
| | 99 | 37 | 13 | | | | |
| English Spoken | > 10 Years | 5-10 Years | 0-4 Years | | | | |
| | 140 | 4 | 5 | | | | |

***Materials & Procedure***

The emotion judgement task was presented in a card-sorting game on a Qualtrics website survey (2022). Participants were shown a stack of 50 GIFs of a speaker saying [i] and [Λ] words or pseudowords along with six boxes in which the GIFS are sorted (See Figure 7 for a screenshot of the task). These six boxes correspond to the six Darwin basic emotions of happiness, sadness, anger, suprise, disgust, and fear (1890/2009).

The 50 word stimuli shown in the GIFs were randomly chosen from a list of 100, consisting of 50 [i] words and pseudowords and 50 [Λ] words and pseudowords (See Appendix B for the list of 100 word stimuli included in this experiment). This was done to accommodate the limitation of the Qualtrics website survey (2022), which only allowed a maximum of 100 stimuli to be added to a card-sorting formatted question. To prevent website lag, only 50 of the 100 GIFs were randomly chosen for each participant. Participants were asked to focus on the speaker's mouth movements and determine the emotion being expressed. They were instructed to drag each GIF into the appropriate emotion box.

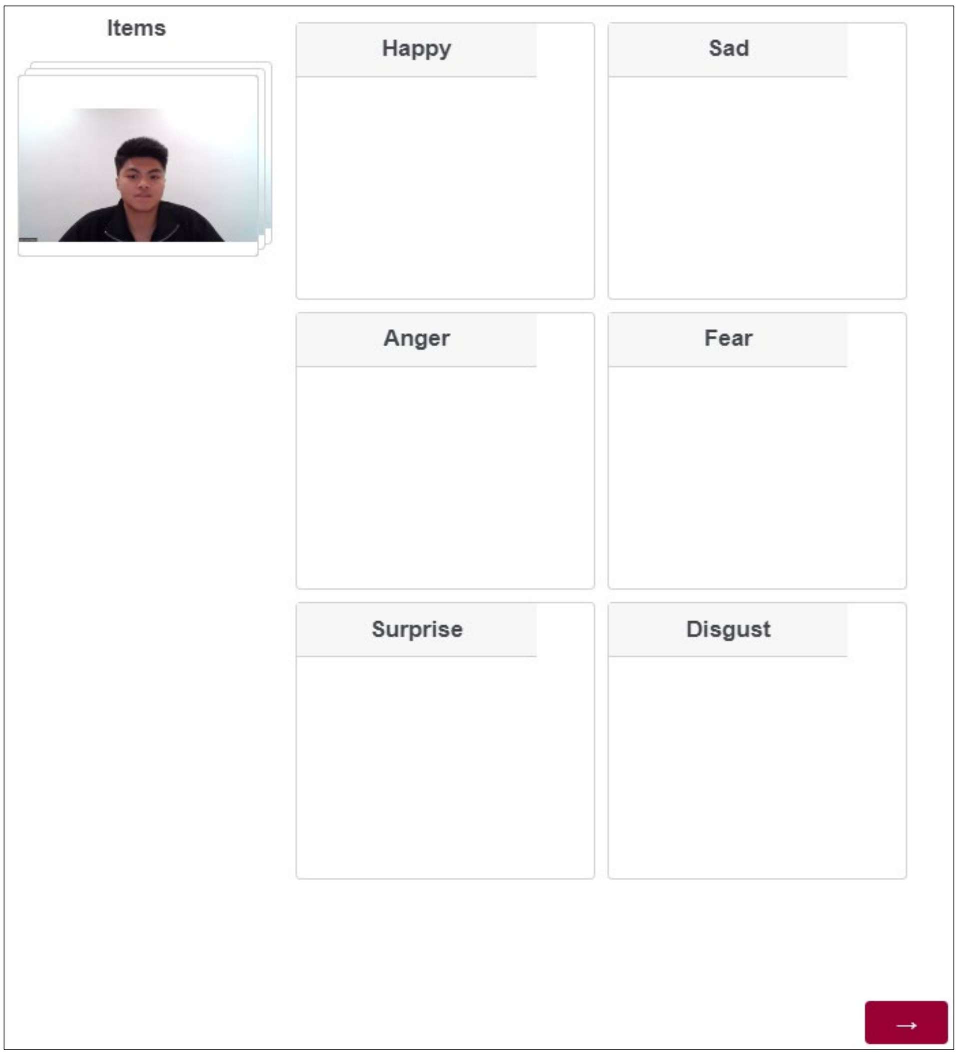

Figure 7. A screenshot of the emotion judgment task.

**Results**

According to paired-samples t-tests comparing the number of [i] and [Λ] GIFS sorted into the six emotion boxes, participant judgments contradicted predictions by judging the speaker as expressing happiness significantly more times when the speaker was saying an [Λ] word than when the speaker was saying an [i] word, and by overall judging the speaker as expressing negative emotions significantly more times when the speaker was saying an [i] word compared to when saying an [Λ] word (see Table 9 for all Experiment 3 paired-samples t-test results).

**Table 9**

*Experiment 3 Paired-Samples T-Test Results*

| Emotion | [i] Descriptives | | [Λ] Descriptives | | Paired-Samples T-test Result |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Happy | 3.70 | 2.62 | 7.37 | 2.96 | $t(148) = -12.2^*$, 95% CI [-4.27, -3.08], $d = -2.01$ |
| Sad | 5.56 | 2.50 | 4.34 | 2.08 | $t(148) = -4.91^*$, 95% CI [-1.70, -0.73], $d = -0.81$ |
| Anger | 5.85 | 2.68 | 4.44 | 1.92 | $t(148) = -5.38^*$, 95% CI [-1.93, -0.89], $d = -0.88$ |
| Fear | 3.53 | 1.93 | 2.57 | 1.75 | $t(148) = -4.57^*$, 95% CI [-1.37, -0.54], $d = -0.75$ |
| Surprise | 3.09 | 1.76 | 5.56 | 2.50 | $t(148) = -0.22$, 95% CI [-0.46, 0.37], $d = -0.04$ |
| Disgust | 3.26 | 1.72 | 3.15 | 1.57 | $t(148) = -0.59$, 95% CI [-0.47, 0.25], $d = -0.10$ |

*Note.* Significance alpha level is Bonferroni corrected to account for the 6 pair-wise comparisons ($^*p < 0.0083$).

**Discussion**

Findings from Experiment 3 not only contradicted the articulatory explanation, but also contradicted the robust *gleam-glum* effect. Participant judgments were likely affected by other visual cues of the face and not just oral movements, despite instructions to focus on oral movements. This explanation is consistent with Experiment 1 and 2 results, where the outcome of comparing articulatory movements with emotional expressions differed when the analyses only looked at movements of oral landmarks versus whole-face landmarks. The speaker's expressions were also likely affected by the counterbalanced order of the word stimuli. The speaker began by reading the entire list of 167 [Λ] English monosyllabic words, which likely led to fatigue, causing more negative emotional expressions when reading of the subsequent lists of words, including the [i] ones.

# CHAPTER 5

# GENERAL DISCUSSION

The *gleam-glum* effect is a robust phonetic emotion association that is multimodal, encompasses entire language lexicons, and potentially universal (Yu et al., 2020; Yu et al., 2021a; Yu et al., 2021b; McBeath et al., 2021; McBeath et al., 2019). Such robust phonetic emotion association suggests an underlying cognitive mechanism that concurrently engages speech and emotion processes. One potential explanation is regarding articulatory oral movements, namely that facial muscles that typically contract during positive emotional expressions (e.g., the zygomaticus major muscle), is likely contracted when articulating [i], and facial muscles that typically contract during negative emotional expressions (e.g., the orbicularis oris muscle), is likely contracted when articulating [Λ].

To test this articulatory explanation, three experiments were conducted, comparing oral movements during [i] versus [Λ] word articulation and emotion expressions. Experiments 1 and 2 objectively assessed the oral movement similarities by analyzing the X and Y coordinates of the oral landmarks. Experiment 1 supported the articulatory explanation, [i] articulatory oral movements is more similar to happy expressions compared to [Λ] articulatory oral movements, and [Λ] articulatory oral movements were more similar to all negative emotional expressions compared to [i] articulatory oral movements. Experiment 1 results further showed that overall [i] articulatory oral movements were non-significantly more similar to positive and neutral emotional expressions than negative emotional expressions, and [Λ] articulatory oral movements were significantly more similar to negative emotional expressions than positive and neutral emotional expressions. Moreover, random forest classification models showed that vertical and horizontal stretches of the two corners of the mouth were equally important for distinguishing emotions and distinguishing the word-type. Experiment 2 provided mixed results, partially supporting the articulatory explanation in natural speech. The results showed that across all emotions, the cluster means were shorter in distance from the [i] articulation cluster mean compared to the [Λ] articulation cluster mean. However, predictions were overall supported by within word-type analyses where [i] articulatory oral movements were non-significantly more similar to positive and neutral emotional expressions than negative emotional expressions, and [Λ] articulatory oral movements were significantly more similar to negative emotional expressions than positive and neutral emotional expressions. Random forest classification showed that the primary difference between [i] and [Λ] articulation from YouTube segments was the horizontal oral movements across various landmarks, and the vertical movement differences contributed very little to the classification. This suggests irrelevant systematic differences between [i] and [Λ] YouTube segments, potentially explaining the mixed results. Experiment 3 assessed the oral movement similarities using a subjective judgment task. Contrary to predictions, participants judged the speaker as expressing happiness more when the speaker was saying an [Λ] word compared to when the speaker was saying an [i] word, and participants judged the speaker as expressing negative emotions more when the speaker was saying an [i] compared to when the speaker was saying an [Λ] word. According to analyses using all 34 facial landmarks in Experiment 1 and Experiment 2, this is very likely due to participants having access to non-oral facial movements. Experiment 3 results can further be explained using the counterbalanced order of the word stimuli.

Considering findings focusing on oral-only movements, the project provides some support for the articulatory explanation for the *gleam-glum* effect. Taken together with previous work showing that oral movements moderate the *gleam-glum* effect size (Yu et al., 2021a) and other studies supporting an articulatory explanation for similar phonetic emotion associations (Körner & Rummer, 2021), the findings have major implications for language and communication. Adelman et al., (2018) had previously suggested that the consistency between the overall definition of a word and the associations from its phonemes is likely to give the word a higher survival advantage (Adelman et al., 2018) because it facilitates word learning and effective communication. If articulatory movements lead to associations, findings reported here are consistent with a potentially embodied mechanisms that affects human language evolution universally.

## REFERENCES


Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, *175*, 122–130. https://doi.org/10.1016/j.cognition.2018.02.007

Arias, P., Belin, P., & Aucouturier, J.-J. (2018). Auditory smiles trigger unconscious facial imitation. *Current Biology*, *28*(14), R782–R783. https://doi.org/10.1016/j.cub.2018.05.084

Aryani, A., Conrad, M., Schmidtke, D. S., & Jacobs, A. M. (2018). Why “piss” is ruder than “pee”? The role of sound in affective meaning making. *PLOS ONE*, *13*(6). https://doi.org/10.1371/journal.pone.0198430

Auracher, J., Albers, S., Zhai, Y., Gareeva, G., & Stavniychuk, T. (2010). P Is for Happiness, N Is for Sadness: Universals in Sound Iconicity to Detect Emotions in Poetry. *Discourse Processes*, *48*(1), 1–25. https://doi.org/10.1080/01638531003674894

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt: Boston, NY.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Darwin, C. (2009). *The Expression of the Emotions in Man and Animals* (2nd ed.). F. Darwin, Ed. Cambridge University Press: New York (Original work published 1890).

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi-org.ezproxy1.lib.asu.edu/10.1037/h0030377

Ekman, P., Friesen, W. V., & Hager, J. C. (2002b). *Facial action coding system: An investigator’s guide*. Salt Lake City: Research Nexus.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712–717. https://doi-org.ezproxy1.lib.asu.edu/10.1037/0022-3514.53.4.712

Ekman, P., Sorenson, R. E., & Friesen, W. V., 1969. "Pan-Cultural Elements in Facial Displays of Emotion." *Science, 164* (3875): 86 – 88.

Garrido, M., Godinho, S., & Godinho, S. (2021). When vowels make us smile: The influence of articulatory feedback in judgments of warmth and competence. *Cognition & Emotion*, *35*(5), 1–7. https://doi.org/10.1080/02699931.2021.1900076

iMotions (2022). *Affectiva iMotions Biometric Research Platform.* https://imotions.com/affectiva/

Kawahara, S., & Shinohara, K. (2012). A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation. *Cognitive Science*, *34*(34).

Körner, A., & Rummer, R. (2021). Articulation contributes to valence sound symbolism. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001124

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Kulke, L., Feyerabend, D., & Schacht, A. (2020). A Comparison of the Affectiva iMotions Facial Expression Analysis Software With EMG for Identifying Facial Expressions of Emotion. *Frontiers in Psychology, 11*, 329. https://doi.org/10.3389/fpsyg.2020.00329

Lane, R., Mcrae, K., Reiman, E., Chen, K., Ahern, G., & Thayer, J. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage*, *44*(1), 213–222. https://doi.org/10.1016/j.neuroimage.2008.07.056

Lookback (2022). https://www.lookback.com

McBeath, M. K., Yu, C. S. P., Barnes, H. & Benitez, V. (2021, May). *Visual-emotional association with vowel phonemes: Support for the Gleam-Glum Effect when paired with visual imagery*. 21st Annual Vision Sciences Society Meeting, Virtual.

McBeath, M.K., Yu, C.S.P., Greenstein, E., Lobato, T., Patten, K.J., & Glenberg, A.M. (2019). Vowel phoneme space maps onto Russell's circumplex model of emotion, *Abstracts of the Psychonomic Society. 24*, 42. (Ms.No.191,p.42).

MoviePy (2017). https://moviepy.readthedocs.io/en/latest/index.html

Need to put on a smile? Here's how: say "cheese". (1943, October). *The Big Spring Daily Herald.*

Patten, K. J. & McBeath, M. K. (2020). The difference between shrieks and shrugs: Spectral envelope correlates with changes in pitch and loudness. *Proceedings of the 2nd International Conference on Timbre (Timbre 2020)*. Thessaloniki, Greece.

Qualtrics (2022). https://www.qualtrics.com

Rainville, P., Bechara, A., Naqvi, N., & Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, *61*(1), 5–18. https://doi.org/10.1016/j.ijpsycho.2005.10.024

Rummer, R., & Schweppe, J. (2019). Talking emotions: Vowel selection in fictional names depends on the emotional valence of the to-be-named faces and objects. *Cognition and Emotion, 33*(3), 404–416. https://doi.org/10.1080/02699931.2018.1456406

Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion, 14*(2), 246–250. https://doi.org/10.1037/a0035752

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology,39*(6), 1161-1178.

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review, 25*(5), 1619–1643. https://doi.org/10.3758/s13423-017-1361-1

Siedlecka, E., & Denson, T. F. (2019). Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review*, *11*(1), 87–97. https://doi.org/10.1177/1754073917749016

Slavova, V. (2019). Towards emotion recognition in texts—A sound-symbolic experiment. *International Journal of Cognitive Research in Science, Engineering and Education, 7*(2), 41–51. https://doi.org/10.5937/ijcrsee1902041s

Stone, G. O., Vanhoy, M., & Orden, G. C. V. (1997). Perception Is a Two-Way Street: Feedforward and Feedback Phonology in Visual Word Recognition. *Journal of Memory and Language*, *36*(3), 337–359. https://doi.org/10.1006/jmla.1996.2487

Takarae, Y., McBeath, M. K., & Krynen, R. C. (2021). Perception of Dynamic Point Light Facial Expression. *The American Journal of Psychology*, *134*(4), 373–384. https://doi.org/10.5406/amerjpsyc.134.4.0373

Toit, T. du, Kruger, H., Drevin, L., & Maree, N. (2022). Deep Learning Affective Computing to Elicit Sentiment Towards Information Security Policies. *Advances in Science, Technology and Engineering Systems Journal*, *7*(3), 152–160. https://doi.org/10.25046/aj070317

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Whissell, C. (1999). Phonosymbolism and the emotional nature of sounds: Evidence of the preferential use of particular phonemes in texts of differing emotional tone. *Perceptual and Motor Skills, 89*(1), 19–48. https://doi.org/10.2466/pms.1999.89.1.19

Whissell, C. (2000). Phonoemotional profiling: A description of the emotional flavour of English texts on the basis of the phonemes employed in them. *Perceptual and Motor Skills, 91*(2), 617–648. https://doi.org/10.2466/pms.2000.91.2.617

Yu, C. S.-P., McBeath, M. K., & Glenberg, A. M. (2021a). The gleam-glum effect: /I:/ versus /λ/ phonemes generically carry emotional valence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001017

Yu, C. S. P., McBeath, M. K., & Glenberg, A. M. (2021b). Phonemes Convey Embodied Emotion. In M. D. Robinson & L. E. Thomas (Eds.), *Handbook of Embodied Psychology* (pp. 221–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-78471-3_10

Yu, C. S. P., Stone, G., McBeath, M., Benitez, V. (2020, November 19-22). *Bouba-Kiki Effect Applies to Meaning as Well as Shape: /i:/ vs /^/ Phonemes Generically Carry Affective Valence* [Poster session]. Psychonomic Society 61st Annual Meeting, Virtual.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for OUGH and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments & Computers*, *29*(4), 600-618.

# APPENDIX A

# LIST OF WORDS FOR EXPERIMENT 1 WORD READ-ALOUD TASK

[Λ] words:

bluff brunch buck budge buff bug bum bump bun bunch bung bunk bus bust chub chuck chug chum chump chunk club cluck clump clung clunk crunch crust cub cuff cup does drub drudge drug drum drunk dub duck duff dug dump dun dung dunk dust fluff flung flunk frump fuck fudge fug fun funk glum grub grudge gruff gum gun gunk gust hub huff hug hum hump hunch hunk judge jug jump junk just luck lug lump lunch lung lust much muck muff mug mum munch must nub nudge nun pluck plug plum plump plunk plus pub puck puff pug pump pun punch punk pup pus rub ruck ruff rug rum rump run rung rust scrub scruff scrum scrunch scuff scum shrub shrug shuck shun skunk sludge slug slum slump smudge smug snub snuff snug spunk strum stub stuff stump stun sub such suck sum sump sun sup thrum thrust thug thump thus tonne truck trudge trump trunk trust tub tuck tuft tug tun tup up us

[Λ] pseudowords:

blum blun brup druch druck drun fluch flum fruch frun frup gulch gluck glup gruch yuch kuch clum crun pluch plun plup prup schuch scruck scrup sluch smuch smuck smum smun smup snum snun spluch spluck splum splup spruch spruck sprum sprup struch thruch truch thrun trun zuch zuck zum

[i] words:

beach beam beech been beep beet bleach bleed bleep breach bream breech

breed breve cheap cheek cheep cleave cream creed creek creel creep deal deed

deem deep dream each eave eel eve feed feel feet fleet free

gene gleam greave grebe greed green greet heal heap heave heed heel

jeep keel keen keep kneed kneel leach leap leave leech leek meal

meed meek meet need peach peal peed peek peel peep preach

preen queen reach real ream reap reed reek reel scene scheme cream

screech screed screen seal seam seed seek seem seen seep sheave sheen

sheep sheet skeet sleek sleep sleet speech speed spleen squeal steal steam

steed steel steep stream street sweep sweet teach teal team teed teem teen

theme tweed tweet veal weal weave weed week weep wheel zeal

[i] pseudowords:

bleem bleen breap dreach dreek dreen fleach fleem freach freen freap gleech

gleek gleap greech yeach keach cleem crene pleech plene pleap preep scheach

screek screep sleech smeachsmeek smeem smeen smeap sneme sneen spleachspleek

spleem spleep spreach spreekspream spreap streech threech treach threne treen zeech zeek

zeem

[æ] words:

apse as ban bast bat bath brad brass brat can cap cat crap

dad dam dan dash fad flak flat gaff gal gas gnat gram half ham hap hat jack lack lad lamb lap lass mad man

mass mat match pal pap ram rat sap scat slap snap span

tab tan tat trap wrath

[u] words:

oops ooze boon boost boot booth brood bruce brute coon coop coot croup

dude doom dune douche food fluke flute goof ghoul goose newt

groom hoof whom hoop hoot juke luke lewd loom loop loose mood

moon moose moot mooch pool poop room root soup scoot sloop snoop

spoon tube tune toot troop ruth

APPENDIX B

THE LIST OF 100 WORD STIMULI IN EXPERIMENT 3

[i] words and pseudowords:

beam bleed bleem bream cheap cleave creel creep deem deep dreach dreen feed feel fleem freap free gleam grebe greed heal heel keel keen keep kneed leach leek preach preen reek reel screech sheen skeet sleet smeach smeek smeen speech spreach steed sweet teem treach veal weal week weep zeech

[Λ] words and pseudowords:

brunch yuch lug us crun drunk tup brup suck plug plun flung snug much must run dung cluck hug gluch truck drub smum snuff grub stub sprup lust smuck snum pluch jug zuck puck hub gum dug luck funk clump duck drudge grudge drun kuch thrum fun thug muff druch

# APPENDIX C

# IRB APPROVAL DOCUMENT

EXEMPTION GRANTED

Michael McBeath
CLAS-NS: Psychology
480/965-8930

Michael.McBeath@asu.edu

Dear Michael McBeath:

On 10/17/2022 the ASU IRB reviewed the following protocol:

| | |
|---|---|
| Type of Review: | Initial Study |
| Title: | Communicative Facial Movements |
| Investigator: | Michael McBeath |
| IRB ID: | STUDY00016617 |
| Funding: | None |
| Grant Title: | None |
| Grant ID: | None |
| Documents Reviewed: | • IRB_Social_Behavioral_Protocol_FacialMovements_10-12-2022.docx, Category: IRB Protocol;<br>• Perceived Verbal Communication Consent_Form_10-10-2022.pdf, Category: Consent Form;<br>• Perceived Verbal Communication SONA Study Description_recruitment_methods_10-11-2022.pdf, Category: Recruitment Materials;<br>• supporting_documents_10-11-21.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions);<br>• Verbal Communication Consent_Form_10-10-2022.pdf, Category: Consent Form;<br>• Verbal Communication SONA Study Description_recruitment_methods_10-11-2022.pdf, Category: Recruitment Materials;<br>• YouTube Video Segments, Category: Other; |

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2)(i) Tests, surveys, interviews, or observation (non-identifiable), (3)(i)(A) Benign behavioral interventions (non-identifiable), (3)(i)(B) Benign behavioral interventions (low risk) on 10/17/2022.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Shin-Phing Yu
Shin-Phing Yu