

AdaptiCon Vowel Experiment Sketch
Janitta Wong & Samantha Gryz
Conversational Adaptation Tasks
June 2023

Introduction

Phonetic convergence, or adaptation, is defined as “an increase in segmental and suprasegmental similarity of the speech of one talker to another” (Pardo, 2006). Past research has found that adaptation in conversation involves a few factors, including conversational roles, gender, dominance in conversational, dialect formation, temporal factors, and social relationships (Labov 1976, 1984; Giles, Coupland and Coupland, 1991). The proposed English vowel task study is part of a collaborative series of experiments focused on conversational adaptation by Simon Fraser University’s (SFU) Language and Brain Lab (LABLab).

The present study specifically considers English vowel articulation in a natural conversational setting. Complimenting experiments consider phonetic adaptation in Cantonese tones, and conversational adaptation between computer-based AI and human participants.

Hypotheses

The proposed English vowel task considers a focused research question: What is the degree to which talkers adjust their English vowel articulations in natural conversation with two specific variants: (1) high and low frequency words and (2) different noise conditions. Based on previous findings from Kim et al. (2011) and Munson and Solomon (2004), we predict that the extent of vowel modifications will be greater in higher frequency words and in noisier conditions. More specifically, we predict that greater vowel adaptation will occur in the beginning when compared with the end of conversation target words, between interlocutors with greater noise levels, or with different L1 backgrounds), as found by Kim et al. (2011). Furthermore, we predict that higher frequency words, defined as words which occur more frequently in a specific language, will result in productions with longer vowel durations and a contracted vowel space, when compared to lower frequency words, as noted by Munson (2004).

Method

Participants

The proposed English vowel task study participants will involve native and non-native English speakers from Simon Fraser University. All participants will be recruited based on set inclusion criteria and be right-handed and between 18 to 35 years old. Native English speakers will refer to participants who grew up in a dominant English-speaking environment, regardless of their heritage language. Prior to the study, participants will be confirmed to have no visual, hearing or speech impairments.

Stimuli

The English vowel stimuli will be acquired in the pre- and post-tasks at the word level from both semi-scripted and unscripted natural conversational sentences between pairs of speakers. Stimuli criteria require that words be monosyllabic, CVC, easily visualized objects and of high and low frequency (for target and non-target words, respectively). Object words must target different vowels using minimal or pseudo-minimal pairs and found in similar consonantal contexts. In addition, vowels will include specific phonetic contrasts of tense (/i/, /u/) and lax types (/ɪ/, /ʊ/, /ɛ/, /æ/, /ɑ/) in minimal or near-minimal pairs of object nouns.

Currently, targets have been updated to include seven (7) ‘flapping’ pairs, focused on a t/d contrast. Additionally, to add to a more robust experimental landscape, English vowel ‘filler’ words have been included; nine (9) pairs of words, focused on minimal pairs: a-ʌ, i-ɪ and u-ʊ. See Appendix A for a complete list of current stimuli. Stimuli and task landscapes are created using an online resource, Irasutoya (<https://www.irasutoya.com/>).

Materials

During the pre- and post-tasks, a recording booth with a computer monitor in a separate room will be used. Participants will be provided with a headset and microphone. Materials are similar or identical in all setups.

General equipment used in the study will include:

Audio:

1. Rhode Wireless Go II with lapel mics; one per each participant
2. NT2-A, set to Cardiod mode; one facing each participant
3. Recorder: TASCAM DR70D

Computer:

1. Mac Mini with 2 virtual machines; one per each participant (separate controllers and displays)
2. 2 x VESA monitors; VESA to C-Stand mounts, C-stands

Video Recording:

1. 2 cameras recording participants faces
 2. 2 camera outputs to ATEM mini pro
 3. 2 monitors using HDMI split
 4. HDMI recorder (Atomos Ninja V) recording outputs of the ATEM mini pro
- (a) In Experiment 1: Guess What card game, each participant will be provided with the same set of **32 character cards** with target and filler word objects and a physical copy of the game instructions that includes examples of carrier phrases. The participants will share a set of mystery character cards placed in the middle of the table. See Appendix B.

- (b) In Experiment 2: Diapix elicitation task, each participant will have a computer monitor placed in front of them on a clamped arm to adjust the height. The monitor will present a similar scene (picture A or B) with some differences based on target words, as well as on-screen prompted questions (the 5 Ws) for participants to discuss their own Diapix scenes. A total of 3 different sets of pictures will be presented. See Appendix C.

Procedure

Experiment 1: Guess What card game

Participants will first complete pre-task training individually for baseline measurements. Then, each participant will be given a printed list of target and non-target object words and asked to read aloud in plain speech. Target words will be recorded and isolated for comparisons in the post-task. In order to encourage collaborative play, participants will be timed and encouraged to solve the ‘mystery character card’

The pair of participants will set up their 32 cards facing up in a 4 x 8 layout. They will each pick up a mystery character card with target words from the pre-task list from the shared deck without letting their partners see it and keep it facing down for the rest of the game. The participants will take turns guessing the object on their partner’s mystery card until one guesses correctly. A physical copy of the game instructions will include examples of carrier phrases for the participants to encourage and elicit word repetitions. The participants will play three rounds of the game, each time in 3 different noise conditions adopted by Gilbert (2014). Noise conditions will be randomized and played through the participants’ headsets, including quiet speech, 4-6 people babbling and mixed noise at various decibel levels.

After completing three rounds of the game, each participant will complete a post-task to compare with their baseline measurements. Participants will be given the same list of target and non-target object words from the pre-task training that they will be asked to read aloud in plain speech. Comparative analyses of pre- and post-task recordings will be conducted with the target word vowels for visual and acoustic cues.

Experiment 2: Diapix elicitation task

Participants will complete the same pre- and post-task procedures as Experiment 1. The only difference will be that participants will be provided with a different list of target and non-target words to read aloud in plain speech. Similarly, speech recordings will be analyzed and compared for visual and acoustic cues of target word vowels between pre- and post-task productions. In order to encourage collaborative play, participants will be timed and encouraged to solve the ‘mystery object’ collaboratively.

The pair of participants will be presented with a similar picture, either picture A or B, that will have some differences based on the target words from the pre-task. Participants will be directed on-screen to focus on specific objects in the scene. Partners will take turns questioning and answering about the target objects until all the locations have been discovered. Prompted questions will appear on the screen (WHERE questions with either visual representation or

written, ie. Where is the pill?) to force breaks to encourage eye contact and natural conversation about their pictures. Likewise, in Experiment 1, the participants will complete the task 3 times with different sets of pictures in Gilbert's (2014) 3 noise conditions: quiet speech, 4-6 people babbling, and mixed noise at various decibel levels, in randomized order.

Measurements

Audio-only

The proposed task will involve pre-task baseline training for both Experiment 1 and Experiment 2, as well as post-task baseline comparisons (see Table 1). Pre-task procedures for both experiments will ask participants to read a provided list of target and non-target words on a computer screen. Participants will be asked to read the words aloud in plain speech, and the target words will be isolated for comparisons to the same target words in the post-task. The post-task procedures involve the same set of target and non-target words in random order, read aloud in plain speech. Speech recordings of the pre-and post-task will be analyzed for visual and acoustic cues of target word vowels. Comparisons will be made between pre- and post-task productions.

Audio-visual

A phonetic convergence assessment, adopted from Kim et al. (2011) will be conducted post-study with the conversational recordings between each pair of participants. From Experiments 1 and 2, the early and late portions of the target word recordings will be used, specifically speakers' first and last target word utterances. Acoustic analyses will be used for comparisons between the target utterances with video recording observations.

Discussion

As of December 2022, we are in the proposal stage of the present experiment. We believe that these tasks will create opportunities for both natural conversation, and slightly structured conversation, as well as analyzing adaptation in terms of word frequency and noise. We believe that the simplicity and procedure of the proposed tasks will allow for natural face-to-face interactions, straightforward understanding, and familiarity with childhood games. Limitations of the proposed tasks might include the consistent location of target words in participant productions, and the frequency of repetitions may need to be adjusted.

As of May 2023, via pilots in February, we have piloted the AdaptiCon vowel task. Feedback from this pilot includes:

- General instructions should include the following details: speak as naturally and clearly as possible, clarify with your partner if necessary
- For the Diapix/Spot the Difference task, instructions must include prompted breaks to encourage eye contact and conversation
- Guess What cards should include one object per card, 1 set per participant
- Use of an English spoken-only database to consider word frequency

- Include minimal pair ‘filler words’ to contrast target list; helps to ensure participants can’t easily define the experiment motivation

This feedback will be adapted and included in future pilots, Summer 2023.

Table 1

Experiment 1	BREAK FOR PARTICIPANTS	Experiment 2
Pre-Task: Baseline Training		Pre-Task: Baseline Training
Task 1: Guess What Card Game		Task 2: Diapix Elicitation Task
Post-Task: Baseline Comparison		Post-Task: Baseline Comparison

Acknowledgments

The present task development is a collaborative effort of SFU’s Language and Brain Lab.

References

Bailly, G., & Lelong, A. (2010). Speech dominoes and phonetic convergence. *Interspeech 2010 - 11th Annual Conference of the International Speech Communication Association*, Sep 2010, Makuhari, Japan.

Biro, T., Toscano, J. C., & Viswanathan, N. (2022). The influence of task engagement of phonetic convergence. *Speech Communication*, 138, 50-66.

<https://doi.org/10.1016/j.specom.2022.02.002>

Kim, M., Horton, W. S., & Bradlow. A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Lab Phonology*, 2(1), 125-156.

<http://doi.org/10.1515/labphon.2011.004>

Lee, Y., Gordon Danner, S., Parrell, B., Lee, S., Goldstein, L., & Byrd, D. (2018). Articulatory, acoustic, and prosodic accommodation in a cooperative maze navigation task. *Plos one*, 13(8), <https://doi.org/10.1371/journal.pone.0201444>

Mukherjee, S., Badino, L., Hilt, P. M., Tomassini, A., Inuggi, A., Fadiga, L., Nguyen, N., &

D'Ausilio, A. (2019). The neural oscillatory markers of phonetic convergence during verbal interaction. *Human brain mapping*, 40(1), 187–201. <https://doi.org/10.1002/hbm.24364>

Munson, B., & Solomon, N. (2004). The effect of phonological neighbourhood density on vowel articulation. *Journal of Speech Language and Hearing Research*, 47(5), 1048-58.
[https://doi.org/10.1044/1092-4388\(2004/078\)](https://doi.org/10.1044/1092-4388(2004/078))

Pardo, J. S., Urmache, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception & Psychophysics*, 79, 637-659.
<https://doi.org/10.3758/s13414-016-1226-0>

Appendix A

Target words:

- Batter / badder
- Butter / budder
- Latter / ladder
- Litre / leader
- Metal / medal
- Petal / pedal
- Writer / rider

Filler words:

- Sheep / ship
- Pill / peel
- Pick / peak
- Mill / meal
- Pitch / peach
- Chick / cheek
- Suit / soot
- Poppy / puppy

- Cud / cod
- Bin / bean

Appendix B

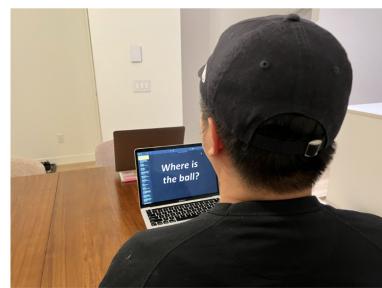
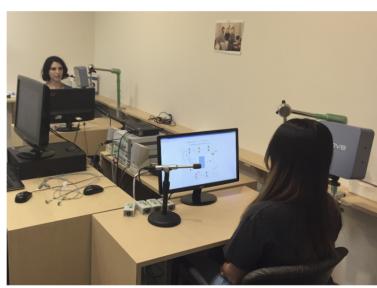
Demo 1: Guess What Game



20

Appendix C

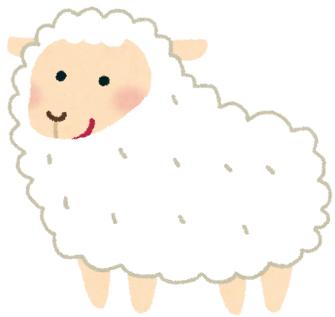
Demo 2: Diapix Task

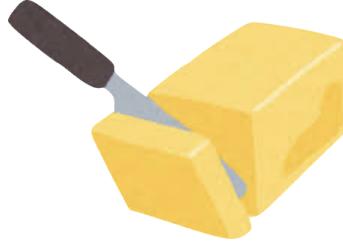
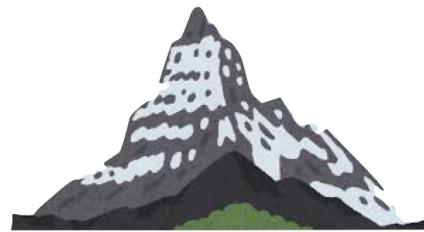


21

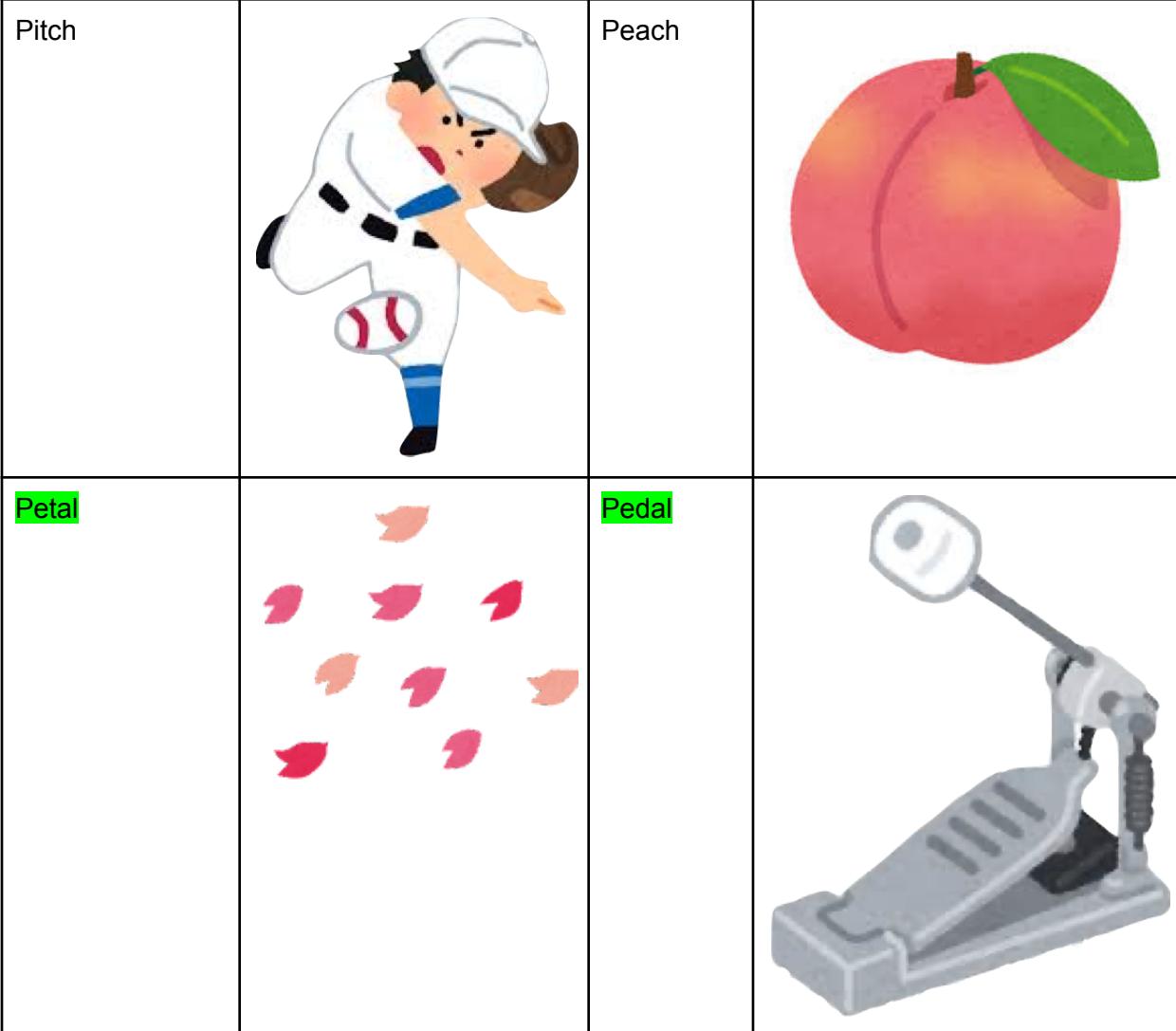
Appendix D

Stimuli Familiarization for Participants

Stimuli			
Group 1	Image #1	Group 2	Image #2
Sheep		Ship	
Pill		Peel	
Batter		Badder	

Butter		Budder	
Pick		Peak	
Latter		Ladder	

Litre		Leader	
Metal		Medal	
Mill		Meal	



Chick



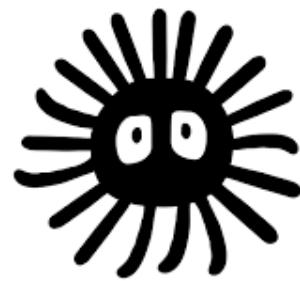
Cheek



Suit



Soot



Poppy



Puppy



Writer



Rider



Bin



Bean

