# Speech-to-Song Illusion in English Monolinguals and Spanish-English Bilinguals

**Lyzzette Melgoza**, Dylan Richardson, Alejandra Santoyo, Kristina Backer, Ph.D.
University of California, Merced

SUMMER UNDERGRADUATE RESEARCH FELLOWSHIP

## Abstract


- The Speech-to-Song illusion is when a spoken phrase repeated in close succession will eventually be heard as sung.
- The goal of this present study is to replicate and extend the research done by Margulis el al. 2015 and focus on English monolinguals and Spanish-English bilinguals. The languages used in this present study are English and Spanish.

- **Research question:** Is the Speech-to-Song illusion for Spanish words perceived more in English monolinguals than Spanish-English bilinguals?
- **Hypothesis**: The English monolingual speakers will perceive the Speech-to-Song illusion for Spanish words more often and greater than the bilingual Spanish-English speakers.


## Introduction

- Illusion was first presented by Deutsch et al. 2011 where they demonstrated that a speech segment after repetitions could be transformed from speech to song.
- There has been a small group of papers published related to the speech–to-song illusion, 14 published papers in total.
- Margulis et al. 2015, discovered that spoken languages that were difficult to pronounce relative to the listener's native language were more likely to perceive the illusion.
- The different languages that the Margulis et al. 2015 research paper contained were: English, Catalan, Portuguese, French, Croatian, Hindi, and Irish.

## Participant Inclusion Criteria

- Must be U.S adults 18 years or older.
- English Monolinguals and Spanish-English Bilinguals.
- Must have internet and headset access.
- Target Participant number: 30

## Methods and Materials

- Composed a list of 20 words on a Spreadsheet based on the frequencies (per million) on CLEARPOND database (Marian et al., 2012 ), for both English and Spanish. Provided is an example of high frequency Spanish words from database.

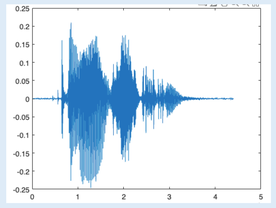
| Word | PhoWord | Length (Ortho) | Length (Phono) | Freq (per-million) | OTAN | PTAN |
|---|---|---|---|---|---|---|
| clase | k.l.a.s.e | 5 | 5 | 278.6298 | 3 | 3 |
| manos | m.a.n.o.s | 5 | 5 | 257.3077 | 10 | 11 |
| cerca | T.E.4.k.a | 5 | 5 | 256.3462 | 6 | 6 |

**Figure 1**. Demonstrates the search results for the words clase, manos, and cerca. Length(Ortho) is the length of how the word is spelt, and Length(Phono) is how many different phonological sounds there are in the word. For this study we focused on Length(Phono) and Freq(per-million) column.

### Procedure

- Both English and Spanish word lists were nouns and had 4-7 phonological differences per word.
- The derived word bank was audio recorded by a female Spanish-English bilingual speaker to create neutral stimuli. Raw audio files were edited and RMS-normalized using MATLAB.

2a.



2b.

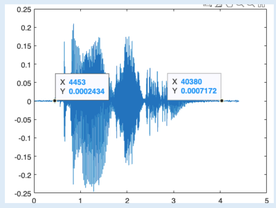
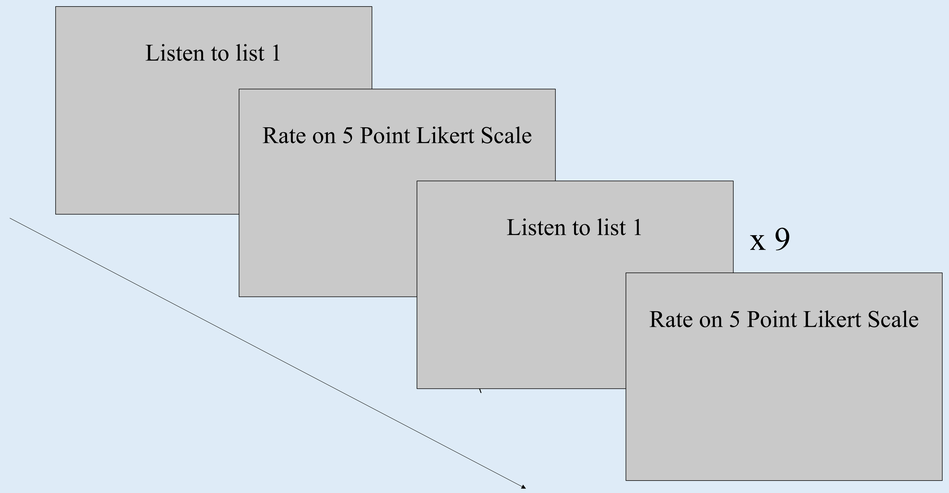



**Figure 2a.** Displays the word "clase" before edits were made.
**Figure 2b.** The two x-coordinate plots shown are the "on-set" and "off-set" times, 4453 and 40380 respectively. These numbers were used to trim the original sound recording using the MATLAB code: >>sound(data(onset:offset),fs). Our principal investigator then ran a code to normalize loudness.

- Words were grouped on MATLAB to create a list of 3 words, for a total of 6 lists. Each list contained no pauses in between the words to enhance repetition and illusion effect (Castro et al. 2018).
- Every participant listened to a high frequency list and low frequency list for both English and Spanish.
- A 5-point Likert scale was use to rate the stimuli. 1 being "sounds exactly like speech" and 5 being "sounds exactly like singing".




## Results

- MATLAB was used to combine the Normalized & Trimmed audio recordings to create a continuous stream of 3 words per list.

3.

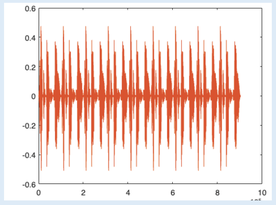


**Figure 3.** Demonstrates the word list "suerte, horas, guerra" repeated 10 times with no pause. This was the final stimuli used.

- This project is on-going as we are still gathering data and participants

## Discussion

- We chose the CLEARPOND (Marian et al., 2012 ) database because it contained both languages used in this study, and we wanted to be consistent with the frequencies. We were considering using Espal corpus, but there was no option for the English language.
- Limitations of this study included having to record stimuli on Apple Voice Memos and not in a sound-attuned booth with a high-quality microphone, like done so in previous studies. This makes the stimuli more susceptible to background noise.
- Another limitation to this study is that due to COVID-19, participants did not complete the experiment in a lab, but from a place they had access to a computer, headphones, and internet.

## Acknowledgements


- Undergraduate Research Opportunity Center
- Principal investigator Dr. Kristina Backer
- Graduate mentors Alejandra Santoyo and Zunaira Iqbal
- Colleague Dylan Richardson


## References


- Castro, N., Mendoza, J., Tampke, E., et al. (2018). An account of the Speech-to-Song Illusion using Node Structure Theory. PLoS ONE 13(6): e0198656. https://doi.org/10.1371/journal.pone.0198656
- Duestch, D., Henthom, T., & Lupids, R,(2011). Illusionary transformation from speech to song. The Journal of the Acoustical Society of America, 129(4), 2245-225.https://doi.org/10.1121/1.3562174
- Margulis E. H., Simchy-Gross R., Black J. L, (2015). Pronunciation difficulty, temporal regularity, and the speech-to-song illusion. Frontiers in Psychology, https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00048/full
- Marian, V., Bartolotti, J., Chabal, S., Shook, A.(2012). CLEARPOND: Cross-Linguistic Easy-Access Resources for Phonological and Orthographic Neighborhood Densities. PLoS ONE 7(8): *e43230.* doi:10.1371/journal.pone.0043230