Do you understand Clonish?

Grandon Goertz[1], Terese Anderson[2], Evan Ashworth[1]
[1]University of New Mexico, US
[2]University of Chicago, US

https://doi.org/10.36505/ExLing-2024/15/0014/000639

Abstract

Artificial intelligence (AI) voice cloning programs produce speech clones that are advertised as being very close to or even identical to naturally spoken language. We hypothesize the clones have measurable differences compared to natural speech, and these differences would be useful in a forensic analysis. This exploratory study compares the differences in the F1-F2 vowel spaces of naturally spoken speech to those of cloned speech produced by an AI cloning program. This investigation uses an innovative analysis program written specifically for this research.

Keywords: Clonish, centroid, cloned speech, vowel space

Introduction

Perceptually, cloned speech often compares to natural spoken speech and formant comparisons may not reveal differences between cloned and natural speech. We theorize that cloned speech, which by definition should be an exact numerical duplicate of naturally produced speech, is in fact different. This difference can be determined by comparing the centroids of the naturally spoken vowels to the centroids of the cloned vowels. The second part of this experiment compares the centroids of the total vowel spaces for both naturally spoken vowels and the cloned vowels.

This research employs new technology employing a modified Matlab® CentroidPolygon.mlx computer file which calculates vowel centroids and compares the centroids of the naturally spoken speech to the centroids of the cloned speech. The advantage of using centroids for comparison is that centroids consider the entire vowel portion of a w ord, centroids figure in the frequency wavelengths of each formant, centroids do not trim or discard data, and the centroid is the mathematical geometric center of each vowel or vowel space (Anderson et al., 2003).

Methods and materials

This research first used sentences that had been produced for a previous perception-production study that compared English and Greek vowel spaces (Botinis, et al., 2022). Native New Mexico English speakers each recorded carrier sentences containing the target words: *bit, beat, bet, bat*, *boot, butt, bought*,

and *bot*. Speech recordings were produced by both speakers using a Røde N microphone in a GretchKen™ Industries acoustic sound booth.

These eight carrier phrases were designed to produce the monothongs /iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/, which represent the corner vowels and edges of the English speakers' vowel spaces. The sentences were spoken clearly, and the key words were spoken with brief silence before and after each word. The vowel portions of the spoken target words were extracted with PRAAT, formants were computed, and the data was stored in a spreadsheet.

The Speechify ™ cloning program was trained on the 16 New Mexico English carrier sentences for both speakers. This program was then used to produce clones of the same sentences on which it was trained producing clone *bit*, clone *beat*, clone *bet*, clone *bat*, clone *boot*, clone *butt*, clone *bought*, and clone *bot*. for both male and female speakers. The vowel portions of the cloned words were extracted with PRAAT, clone formants were computed, and the formant clone data was stored in a spreadsheet.

The Matlab® CentroidPolygon.mlx program was used to calculate the centroids of the sixteen naturally spoken target vowels and the sixteen cloned speech vowels. The vowel spaces for both naturally spoken vowels and cloned vowels showed that the clone vowel spaces were modified and reshaped. This information is presented in Table 1 in the results section.

For the second part of this experiment, different sound recordings were necessary to test the viability of the computer algorithm. In this exercise, four participants each read a list of 82 monosyllabic words representing the vowels and diphthongs of American English with a variety of consonantal boundaries, producing 328 speech tokens. Speech recordings were produced by three female speakers and one male speaker using a Røde N microphone in a GretchKen™ Industries acoustic sound booth.

The words containing the /iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/ monothongs, which represent the corner vowels and edges of the English speakers' vowel spaces were identified in the lists and used in the study. The vowel portions of these target words were extracted with PRAAT, formants were computed, and the data was stored in a spreadsheet.

Speechify ™ was then trained on the 328 New Mexico English speech tokens, and this program was used to produce clones of the same words. Cloned vowel corner and edge vowels were again identified for the clone vowels space. The vowel portions of the cloned words were extracted with PRAAT, cloned formants were computed, and the clone data was stored in a spreadsheet.

The CentroidPolygon.mlx program was used to calculate the centroids of the individual vowels and the centroids of the vowel spaces for both natural speech and cloned speech. Chart 1 shows the locations of the natural spoken vowel

space centroids and the cloned vowels space centroids, and how much the centroids have moved.

Results

The spoken English natural vowels and the cloned English vowels were plotted on the same F1-F2 scale for comparison. It was expected that cloned vowel locations would be located close to the natural speech locations. This is based on the ideal that cloned words are perceived the same as original speech.

In contradiction, it was found that cloned vowel centroids were not located anywhere near to the corresponding centroids of natural speech. The change was especially evident in the cases of the New Mexico English vowels, /iː, æ, uː, ʌ, ɔ/ with F1 and F2 values. The change in vowel locations was also irregular and not predicted by the data.

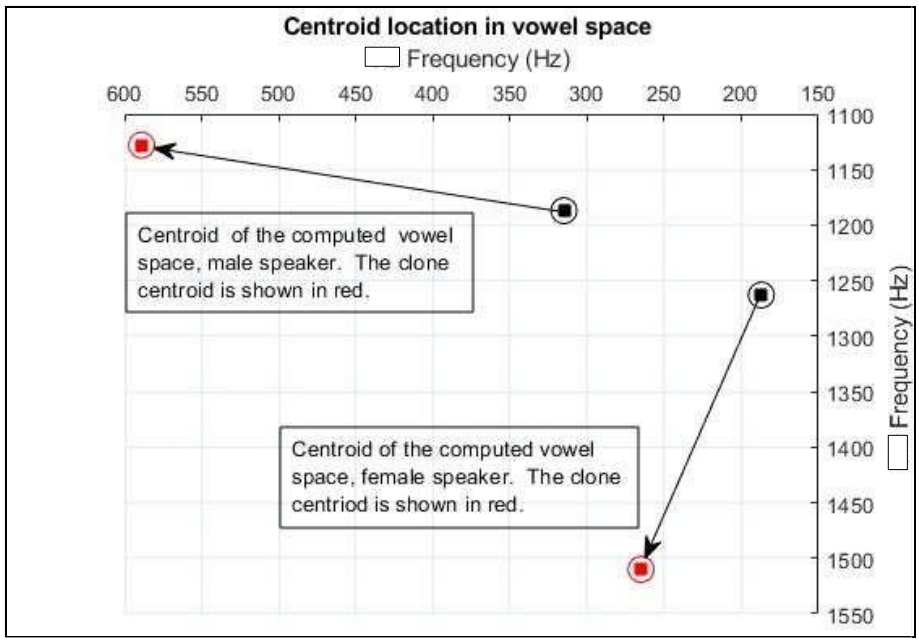
The centroids of both the natural speech and the cloned vowel spaces were computed, to show a net vowel space movement.

Table 1. The centroids values of naturally spoken and cloned English vowel spaces define different vowel spaces.

	F1 (Hz)			F2 (Hz)	
	Mean	SD		Mean	SD
Female speaker	187.	137.84		1263.3	245.72
Female cloned speech	265.1	89.73		1510.8	278.21
Male speaker	315.8	151.52		1187.2	246.50
Male cloned speech	589.2	230.80		1128.1	237.04

Significant formant changes were noted when the natural speech of each speakers' 82-word set was compared to their corresponding cloned speech word set, but the cloned vowels did not exhibit a pattern in the formant frequency changes. Therefore, centroids were computed for the total vowel space for both natural speech and the clones. Centroids represent the center of the total vowel space and show the differences between natural and cloned vowels in terms of frequency changes. The centroids for the male speaker's natural speech and clone speech, and the female speakers' natural speech and clone speech show, in summary form, the vowel shape differences.

Chart 1. A plot showing the centroid locations of the vowel spaces defined by the corner and edge vowels (/iː, ɪ, e, æ, uː, ʌ, ɔ, ɑː/) and the centroid locations of the cloned vowel spaces.



Discussion and conclusions

This research shows the vowel space centroids for the naturally spoken vowels are significantly different from the centroids of cloned vowels, and this change would be forensically useful to identify cloned speech. We understand that this is a limited study, but the mathematical algorithms that were developed are well-suited for additional study in speech analysis.

References

Anderson, T., Botinis, A., Goertz, G., Kontostavlaki, A. 2022. Vowel characterization by centroids. ExLing 2022 Paris: Proceedings 13th International Conference of Experimental Linguistics, October 13-16. Paris, France.

Botinis, A,, Goertz, G., Kontostavlaki, A., Anderson, T. 2023. Vowel discrimination of American English. ExLing 2023 Athens: Proceedings 14th International Conference of Experimental Linguistics, October 13-16. Athens, Greece.

The MathWorks Inc. 2023. MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com.