Exploring Asian North American English: A YouTube Corpus-based Approach

Evidence for an association between phonetic features and Asian North American (ANA) identities is both limited and inconsistent. Moreover, it is unclear whether purported patterns correspond to ethnic-specific (e.g. Vietnamese American) or pan-ethnic identity (e.g. Asian American/Canadian). Indeed, there are good reasons why an ANA ethnolectal variety or varieties (cf. Chicano English) may not exist given the diversity of linguistic, cultural, and socioeconomic backgrounds present in ANA communities, which can include those of East, Southeast, South and West Asian as well as Pacific Islander descent (Reyes & Lo 2009). At the same time, perceptual studies indicate that some ANA speakers can be identified as 'Asian' at rates above chance (e.g. Newman & Wu 2011). To explore potential acoustic correlates of this perception, we take a computationally-driven, corpus-based approach to investigating ANA speech patterns.

Several phonetic features have been investigated in previous work on ANA English, two of which are vocalic durational variability and F2 of the mid back vowel (/oʊ/). However, while some argue that lower durational variability and backed /oʊ/ can index pan-ethnic Asian identity (Bauman, 2016), others find either no differences across ethnic/racial groups (Newman & Wu, 2011) or ethnic-specific variation instead (Cheng et al., 2016). This inconsistency in findings may be tied to local or regional factors; thus, we conduct an exploratory cluster analysis of ANA speech using data from YouTube, which allows us to access a wider range of individuals (e.g. in terms of ethnicity, region, age, etc.). In addition, clustering allows for the discovery of emergent groups of speakers along multiple dimensions.

Self-described ethnicity and regional origin information was obtained per speaker as available, mainly from user-posted Q&A or tag videos (e.g. the Growing Up Asian American tag) supplemented by other sources (e.g. YouTube channel About pages). Thus far, data has been collected from videos of 47 ANA-identifying speakers, 3 mixed/multiracial ANA speakers, and 8 non-ANA-identifying YouTubers. For this preliminary analysis, we report on 20 speakers from California: 17 ANAs who identified as East or Southeast Asian (including mixed-Asian ethnicities) and 3 non-ANAs who described their ethnic backgrounds as 'Hispanic-Caucasian', 'Danish-German', and 'half Irish', respectively. Following manual screening for audio quality and length, video captions and audio were downloaded and processed via Python. Captions were then semi-manually corrected and time-aligned. Finally, vowel segments were identified via automatic forced alignment using the Montreal Forced Aligner (McAuliffe et al. 2017). Hand-correction of forced-aligned segments is in progress; thus, data included in preliminary analyses are based on automatic measurements.

Two vowel measures per speaker were submitted to hierarchical clustering using Ward's method: (1) a normalized /oʊ/-backing score calculated as the F2 difference between each speaker's /i/ and /oʊ/ vowels (higher score represents more retracted /oʊ/), and (2) a normalized Pairwise Variability Index (nPVI) score (Grabe & Low, 2002) that assesses relative degree of vowel duration variability (lower score being less variable). Three clusters emerged from the data: One cluster (n=4) of ANA speakers with particularly *fronted* /oʊ/ ($M_{backing}$=477 Hz; M_{nPVI}=49.56), one cluster (n=4) consisting of individuals—including all three non-ANA speakers—who produced comparably less variable vowel durations ($M_{backing}$=613 Hz; M_{nPVI}=45.85), and one cluster (n=12) containing all other ANA speakers with the most retracted /oʊ/ on average ($M_{backing}$=640 Hz; M_{nPVI}=50.53). These results are inconsistent with previous findings: (1) *non*-ANA speakers were differentiated by consistently *less* variable vowel durations, (2) ANA speakers did *not* produce notably backer /oʊ/ than non-ANA speakers, and (3) ANA speakers did *not* cluster by specific ethnicity. Continued analysis—to include more speakers, additional features and hand-corrected vowel values—will confirm whether these preliminary results can be generalized. Insight into features of ANA English in production will not only contribute to documentation of understudied language variation but also enable further understanding of ANA ethnolinguistic identification in perception and its real-world consequences.

References

Bauman, C. (2016). *Speaking of Sisterhood: A Sociolinguistic Study of an Asian American Sorority.* Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. *Laboratory Phonology 7.* Newman, M., & Wu, A. (2011). "Do You Sound Asian When You Speak English?" Racial Identification and Voice in Chinese and Korean Americans' English. *American Speech.* Reyes, A., & Lo, A. (2009). *Beyond Yellow English: Toward a Linguistic Anthropology of Asian Pacific America*. McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *INTERSPEECH.* Cheng, A., Faytak, M., & Cychosz, M. (2016). Language, race, and vowel space: Contemporary Californian English. *Proceedings of the Forty-Second Annual Meeting of the Berkeley Linguistics Society*.

Please provide a maximum 100-word version of your abstract. This summary will appear in the Annual Meeting Handbook if your submission is accepted. Please provide a text-only version; if your abstract includes special characters, please also upload a PDF below

Version 1 (specific results) WC: 100

To extend previous research on Asian North American (ANA) English, we conduct an exploratory cluster analysis on vocalic durational variability and /oʊ/-backing using YouTube speech data. Preliminary analysis of 20 Californians (17 ANAs and 3 non-ANAs) result in three emergent clusters: (1) ANAs with particularly *fronted* /oʊ/ (n=4; $M_{backing}$=477 Hz; M_{nPVI}=49.56), (2) individuals with less variable vowel durations, including all three non-ANAs (n=4; $M_{backing}$=613 Hz; M_{nPVI}=45.85), and (3) the remaining ANAs ($M_{backing}$=640 Hz; M_{nPVI}=50.53). Notably, these results are inconsistent with previous findings and indicate future directions for analysis.

Version 2 (interpreted results) WC: 99

To extend previous research on Asian North American (ANA) English, we conduct an exploratory cluster analysis on vocalic durational variability and /oʊ/-backing using YouTube speech data. Preliminary analysis of 20 Californians (17 ANA-identified and 3 non-ANA-identified) revealed three emergent clusters. Results are generally inconsistent with previous findings: (1) non-ANA, rather than ANA, speakers were differentiated by consistently less variable vowel durations, (2) ANA speakers did not produce notably backer /oʊ/ than non-ANA speakers, and (3) ANA speakers did not cluster by specific ethnicity. This indicates a need for additional research, particularly including more speakers and additional phonetic features.

(e.g. Thomas & Carter, 2006; but see Arvantini, 2012)

Bauman (2016) reported that the speech of women in a mid-Atlantic Asian-interest sorority was both more syllable-timed and contained more retracted /oʊ/ than European American speakers; these features were interpreted as indexing pan-ethnic ANA identity. On the other hand, Newman & Wu (2011) did not find consistently more syllable-timed rhythm in the speech of Chinese and Korean New Yorkers compared to speakers of other racial/ethnic backgrounds while Cheng et al. (2016) reported that Korean and Chinese Californians differed in degree of /oʊ/-backing, which suggests ethnic-specific differences.

[clustering on a larger… we therefore use YouTube to collect a wider range of data than included in previous studies]

A main cluster (C1; n=12) consisting of ANA speakers of various ethnicities, a smaller cluster (C2; n=4) of ANA speakers with particularly *fronted* /oʊ/ (M_{C2}=477 Hz vs. M_{C1}=640 Hz and M_{C3}=613 Hz F2 differences), and another smaller cluster (C3; n=4) consisting of individuals—including all three non-ANA speakers—who produced comparably more syllable-timed rhythm (M_{C3}=45.85 vs. M_{C1}=50.53 and M_{C2}=49.56 nPVI scores).

Newman, M., & Wu, A. (2011). “Do You Sound Asian When You Speak English?” Racial Identification and Voice in Chinese and Korean Americans’ English. American Speech, 86(2), 152–178. https://doi.org/10.1215/00031283-1336992

Reyes, A., & Lo, A. (Eds.). (2009). Beyond Yellow English: Toward a Linguistic Anthropology of Asian Pacific America. Oxford University Press, USA.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. INTERSPEECH. https://doi.org/10.21437/interspeech.2017-1386

Cheng, A., Faytak, M., & Cychosz, M. (2016). Language, race, and vowel space: Contemporary Californian English. In E. Clem, V. Dawson, A. Shen, & A. Horan (Eds.), Proceedings of the Forty-Second Annual Meeting of the Berkeley Linguistics Society (pp. 63–78).

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), Laboratory Phonology 7 (pp. 515–546). De Gruyter Mouton. http://www.degruyter.com/document/doi/10.1515/9783110197105.2.515/html

Hanna (1997). Title. Journal.

Reyes & Lo (2009)

Wong & Babel (2017)

Malika Charrad, Nadia Ghazzali, Veronique Boiteau, Azam Niknafs (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1-36. URL http://www.jstatsoft.org/v61/i06/.

Cheng, A., Faytak, M., & Cychosz, M. (2016). Language, race, and vowel space: Contemporary Californian English. In E. Clem, V. Dawson, A. Shen, & A. Horan (Eds.), Proceedings of the Forty-Second Annual Meeting of the Berkeley Linguistics Society (pp. 63–78). /paper/Language-%2C-race-%2C-and-vowel-space-%3A-Contemporary-Cheng-Faytak/1e00e977100070a944bf78978ba93ef390394f4c

Understanding how listeners can identify ethnic ANAs despite the lack of clear identifiable ethnolectal features is important for (1) understanding how listeners generally do sociolinguistic perception and identification and how stereotypes and salience may play a role in ethnolinguistic identification, and (2) applying to our understanding and mitigation of linguistic discrimination/profiling of Asian individuals.

These were submitted to a k-means clustering algorithm … [clustering details]
one (n=4) with particularly low GOAT-backing score (i.e., fronted GOAT, M=476.6508 Hz diff), one cluster (n=4) with particularly low nPVI scores (i.e., more syllable-timed rhythm, M=45.85125), and one main cluster (n=12) with generally backed GOAT (M=640.0774)

[Possibly general result of finding ethnic-specific or pan-ethnic clustering.]

Preliminary results suggest that ANA and non-ANA speakers were able to ve

[Using the NbClust package in R (cite), the optimal number of clusters was determined to be 3.]

Speakers were grouped into 3 clusters (as determined using the NbClust package in R [cite]):

These findings support XXX () …, which suggests that...

GOAT-backing and nPVI scores then were z-scored and submitted to hierarchical clustering using Ward's method.

the Lobanov-normalized F2

several studies have reported that Korean Americans produce a more backed GOAT vowel (Cheng et al., 2016)

https://www.linguisticsociety.org/content/lsa-2022-annual-meeting-call-abstracts

some work suggests that markers of ethnic identity may in fact be leveraged by ANA speakers (Bauman 2016)

Graveyard

Although some phonetic features have been examined in the context of ANA speech, e

(East or Southeast Asian)

Growing up Asian-American (or Mexican-American), Get to Know Me, and Facts about Me tag videos

tied in with *forever foreigner* and *honorary white* ideologies

The studies that do exist mainly focus on large co-ethnic communities (e.g. California) and East Asian ethnic groups (e.g. Chinese, Korean, Japanese).

Though there are challenges with using YouTube for speech data, it allows for the collection of a much wider sample…

the current analysis therefore examines whether and how speech rhythm and /ou/-backing pattern across various ANA and non-ANA speakers.

Eligible speakers were identified through a combination of “tag” videos—themed videos where users “tag” their friends to make a video on the same topic—and recommended videos.

Perceptual experiments find that (some) listeners can identify (some) individuals as Asian (Hanna, 1997; Newman & Wu, 2011; Wong & Babel, 2017). This suggests that there may be features indexing (pan-)ANA identity, but what these features might be are unclear.

Features that have been associated with ANA identity include /ou/-backing and syllable-timed rhythm

Why is ANA English interesting/important/worth studying?

Using videos from the same “tag” topic also helped ensure that the content is more homogenous [and also allowed us to take into account speaker identity?]

—

So far: 7 Chinese, 4 EAS, 4(0) FIL, 5(3) Korean, 3 SEA, 7 VIET, 2+(0) NON, 2(0) MIX = 34(24)

Group	COUNTUNIQUE of Channel_name
chi	11
eas	7
fil	6
kor	9
mix	3
non	8
sea	4
viet	10
Grand Total	58

To date, no clear evidence of any Asian North American (ANA) ethnolectal variety (cf. Chicano English) has been reported, and indeed, there are good reasons why such a variety (or varieties) may not exist, given the diversity of cultural, socioeconomic, and linguistic backgrounds present in ANA communities (Reyes & Lo 2009). At the same time, some work suggests that markers of ethnic identity may in fact be leveraged by ANA speakers (e.g. Bauman 2016). Moreover, perceptual studies indicate that at least some ANA speakers can be identified as 'Asian' at rates above chance (e.g. Newman & Wu 2011); however, the acoustic correlates of this perception remain murky. To extend this research to a broader range of individuals and ethnic groups, we take a computationally-driven, corpus-based approach to investigating ANA speech patterns.

Evidence for an association between phonetic features and ANA identities is limited and inconsistent, and it is unclear whether patterns correspond to ethnic-specific or pan-ethnic identity. Bauman (2016), for example, reported that the speech of women in an Asian-interest sorority was both more syllable-timed and contained more retracted /ou/ (GOAT) vowels than that of non-sorority members; these features are interpreted as indexing a pan-Asian ethnic identity associated with the sorority. In contrast, Newman & Wu (2011) did not find consistently more syllable-timed rhythm in the speech of Chinese and Korean Americans compared to individuals of other racial/ethnic backgrounds. On the other hand, several studies have reported that Korean Americans specifically produce backed GOAT vowels (cites).

NWAV version

Introducing LingTube: An open-source toolkit for linguistic analysis of YouTube data

YouTube is a vast, yet relatively untapped, source of publicly-available linguistic data (Schneider, 2016). While not without challenges, there are many advantages to using YouTube as a data source, particularly for the computational study of language variation and change: Researchers can collect large amounts of 'naturalistic' speech data representing various contexts (cf., Hall-Lew & Boyd, 2020 on self-recordings), and much of this speech is already captioned, jumpstarting the transcription process. Furthermore, YouTube has the potential to improve access to lesser-studied language varieties or communities. In this vein, this study has two main contributions. First, we introduce LingTube, an open-source suite of tools for automating the downloading and processing of captioned YouTube audio. Second, we present ongoing exploratory work applying these tools to identify potential features of Asian North American (ANA) ethnolinguistic varieties.

In language variation research, data gathered from YouTube videos has been used to study both intra-speaker variation (e.g., Lee, 2017 on style-shifting across contexts) and inter-speaker variation, including larger-scale comparative analysis across regions (Coats, 2020 on articulation rate across US regions) or languages (Kramer, 2021 on cross-linguistic patterns of dependency length minimization). Analysis can be conducted at various levels of linguistic structure, including morphosyntactic, lexical and phonetic. However, working with YouTube data comes with some drawbacks, particularly for phonetic analysis, which LingTube has been developed to address. First, downloading audio and captions are not straightforward. LingTube automates the process of scraping this data from a list of videos, an entire channel, or a set of search results. Second, auto-generated text and/or time-alignment still require hand-correction, and third, unanalyzable speech (e.g., due to background noise or music) must be identified and removed; these tasks remain somewhat time-intensive. Although manual work is still required at this stage, LingTube helps to streamline the process of cleaning, correcting and time-aligning transcripts, as well as identifying usable sections of speech. Ongoing development aims to automate the process of detecting unusable speech. Finally, LingTube also automates the process of creating TextGrids for conducting forced alignment and facilitates hand-correction of forced-aligned segment boundaries.

To demonstrate this novel tool for analysing variation using computational methods, we apply LingTube to study inter-speaker variation across ethnicity and region among ANAs. Although ANAs are not a monolith, anecdotal and perceptual evidence suggest that ANA speakers can often be identified as such by local listeners, hinting at the existence of recognizable ANA varieties (Hanna, 1997; Newman & Wu, 2011; Wong & Babel, 2017). However, this contrasts with thus-far inconclusive evidence for any 'distinctive' ANA ethnolinguistic variety or variants (Reyes & Lo, 2009; Newman & Wu, 2011). Given that few studies have done comparative analyses across regions (Wong & Hall-Lew, 2014) or ethnicity (Bauman, 2016), we are developing a corpus of speech produced by 80 YouTubers, including ANA-identifying speakers and non-ANA comparisons. Although analysis is still ongoing, findings from cluster analyses of vowel space and speech rhythm will, regardless of outcomes, provide new insight into the open question of what, if any, ANA ethnolinguistic features exist as markers of (pan-)ethnic identity.

References

Coats, S. (2020). Articulation Rate in American English in a Corpus of YouTube Videos. *Language and Speech*, *63*(4), 799–831. https://doi.org/10.1177/0023830919894720

Hall-Lew, L., & Boyd, Z. (2020). Sociophonetic perspectives on stylistic diversity in speech research. *Linguistics Vanguard*, *6*(s1). https://doi.org/10.1515/lingvan-2018-0063

Hanna, D. B. (1997). Do I Sound "Asian" to You?: Linguistic Markers of Asian American Identity. *University of Pennsylvania Working Papers in Linguistics*, *4*(2), 15.

Kramer, A. (2021). Dependency Lengths in Speech and Writing: A Cross-Linguistic Comparison via YouDePP, a Pipeline for Scraping and Parsing YouTube Captions. *Proceedings of the Society for Computation in Linguistics*, *4*(1), 359–365.

Lee, S. (2017). Style-Shifting in Vlogging: An Acoustic Analysis of "YouTube Voice." *Lifespans and Styles*, *3*(1), 28–39. https://doi.org/10.2218/ls.v3i1.2017.1826

Newman, M., & Wu, A. (2011). "Do You Sound Asian When You Speak English?" Racial Identification and Voice in Chinese and Korean Americans' English. *American Speech*, *86*(2), 152–178. https://doi.org/10.1215/00031283-1336992

Reyes, A., & Lo, A. (Eds.). (2009). *Beyond Yellow English: Toward a Linguistic Anthropology of Asian Pacific America*. Oxford University Press, USA.

Schneider, E. W. (2016). World Englishes on YouTube: Treasure trove or nightmare? *World Englishes*, 253–282. http://www.jbe.platform.com/content/books/9789027267061-veaw.g57.11sch

Wong, P., & Babel, M. (2017). Perceptual identification of talker ethnicity in Vancouver English. *Journal of Sociolinguistics*, *21*(5), 603–628. https://doi.org/10.1111/josl.12264