MOCK INFANTILE SPEECH: A SOCIOLINGUISTIC PERSPECTIVE

by

BAILEY BIGOTT

(Under the Direction of Jon Forrest)

ABSTRACT

The growing popularity of the internet has led to the creation of new online social spaces and subcultures, often with their own linguistic personas and associated speech styles. The thesis examines one such proposed speech style called "Mock Infantile Speech" (MIS) through a sociolinguistic lens, attempting to connect individual phonetic and phonological features to aspects of a larger "uwu" persona. Raised F0 values, retraction of sibilants, liquid modification, and interdental stopping in the mock and standard guises are compared across speakers who successfully use Mock Infantile Speech. The analysis demonstrates that a difference between guises is significant and subsequently contextualizes Mock Infantile Speech in its relationships to language, gender, the internet, and child speech.

INDEX WORDS: Mock infantile speech, Internet, Linguistics, Sociolinguistics, Indexicality, Linguistic persona, Gender, Sexuality

MOCK INFANTILE SPEECH: A SOCIOLINGUISTIC PERSPECTIVE

By

BAILEY BIGOTT

B.A., The University of Georgia, 2020

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2021

MOCK INFANTILE SPEECH: A SOCIOLINGUISTIC PERSPECTIVE

by

BAILEY BIGOTT

Major Professor:	Jon Forrest	
Committee:	Keith Langston	
	Margaret Renwick	

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2021

DEDICATION

Dedicated to my family, who continue to support me despite not really knowing what linguistics is. I hope this gives some answers.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Within the 21st century, the internet has expanded from a primarily text-based medium to a place where people can converge internationally to create new audio and video content accessible to broader audiences. The most recent generations have grown up alongside the internet, already well-versed in the rapid exchange of information it allows. This increased access results in new communities and subcultures that would have been impossible to form across global and linguistic boundaries even twenty years ago. Within these online spaces – developed through apps, forums, social media, and video streaming websites – come their own associated linguistic features. Members develop their own norms and ideas, and once a coherent culture forms, those same features can become subject to stereotyping, commentary, and parody. While these communities may superficially appear to be confined to the internet, there is always the potential for impact – positive and/or negative - on everyday life outside of those spaces. With this in mind, it is vital to recognize when such communities form, and how members mark association with them. Where the internet allows for relative anonymity, the use of linguistic features to signal belonging to a group and to assist in construction of that group's identity becomes a significant part of the online discourse.

In this study, I attempt to define a new such subculture and associated speech style, which I have designated as Mock Infantile Speech (MIS). I posit that this new identity, found primarily on social media sites such as YouTube and TikTok, has been adopted by adult women imitating features of child speech within internet gaming communities. While it is difficult to ascertain the explicit intention of this vocal pattern without further sociolinguistic research and

1

conversations with speakers, it superficially serves two purposes: to identify the speaker as a member of the community (often called the "uwu girls"), and to use targeted linguistic features often associated with submission, femininity, and a "childlike" pattern of behavior. Additionally, this vocal pattern has become prolific enough that it has become subject to parody by other app users, signifying its high level of visibility and recognizability within popular culture. While not exclusive to TikTok and YouTube, the apps' large breadth and young user base invites participants of this speech style to advertise their own audio and video content. Therefore, the abundance of clear voice recordings, where voice and image are often the sole methods of conveying information, make them an ideal subject for academic study. Through my research I investigate further the features characterizing this speech pattern as well as posit some social motivations for their production. Since this speech style is relatively new, I take a mostly descriptive approach, focusing on the phonetic and phonological features of the speech style through a sociolinguistic lens. I identify salient features of Mock Infantile Speech and attempt to connect them to the broader roles and traits that they index.

I first review previous literature on social media sites, indexicality, and the relationship between language and gender. Next, I provide an overview of "uwu" subculture, its popular characteristics, and competing perspectives towards its members and associated speech style. I then posit what features I expect to find and why, as well as outline the methodological approach for impressionistic coding and statistical analysis of audio samples from speakers. Following this is a presentation of the results and discussion of their implications and place within the larger sociolinguistic landscape.

BACKGROUND

First, I define the app TikTok and discuss more generally the relationship between language and the internet. Next, I review how speech relates to social characteristics and how people can recombine features to make new meanings and identities. Then I discuss the relationship between gender and language and associated linguistic personas commonly associated with women. Finally, I consider child speech and propose the term "Mock Infantile Speech," to describe a new speech style performing a type of femininity through the adoption of childlike phonetic and phonological features.

TikTok and Digital Communities

TikTok emerged in 2016 when the Chinese company ByteDance bought the now-defunct app Musical.ly. The app focuses on sharing short videos, typically 15 seconds to a minute in length, among 800 million users (Iqbal 2020). While content creators may caption and comment on videos, written language is largely absent from the platform. Without permanent forums, users upload videos in response to one another, creating a decentralized environment in which memetic trends and discussions occur simultaneously and develop at high speeds. The userbase of TikTok tends to skew towards younger audiences, with 60% of US users aged between 16 and 24 (Iqbal 2020). Because TikTok's content is almost exclusively visual and auditory (content creators may lip-sync to previously-uploaded audio tracks or create their own), its nature lends itself to metadiscourse. Zuo & Wang (2019) emphasize this decentralized network as well as identify individual motivations for using TikTok, including self-expression and a desire to participate/comment on social trends.

There is almost no existing literature on TikTok, likely due to its relative youth. Most studies focus on its alleged censorship and cybersecurity breaches on behalf of the parent company (Ryan, Fergus, et al. 2020). Other research has covered its use as a platform for communicating public health information within China (Zhu, Xu, et al. 2019) and a social media site through which cyberbullying can influence the outcome of elections (Wekesa 2019). While these discussions elucidate the global impact TikTok can have, they nonetheless do not introduce a linguistic analysis of TikTok despite its wealth of auditory data.

Other studies, however, have delved into the idea of language on a digital platform and the impact of the internet on language at large. In terms of written speech on the internet, people generally tend to write as if they were talking (Davis & Brewer 1997). Any person frequenting internet discussion boards and social media has observed the online preservation of linguistic features, including regional colloquialisms, syntax patterns, and even mispronunciations. Crystal (2001) also identifies online social communities, giving the example of "hackers" as one relevant subculture. Like any community of practice that would traditionally be grounded in physical interaction, these social communities utilize a shared vocabulary to communicate belonging to the group and exclude others identified as not part of the group (Wenger & Lave 1991). Thus, internet-based communities of practice do exist based on some larger identity that people can be included into or excluded from. However, Crystal restricts his discussion to textual exchanges, likely due to the early publishing of *Language and the Internet*, before the internet became as widely accessible as it is today. As established above, TikTok is unique in that encourages an audio and video component to internet communication, where text is possible but not the primary method of contributing content. This shifts the definition of an online community of practice slightly differently, incorporating not only and not primarily lexical and syntactical features, but

also phonological features. Evidence for this newer kind of online community of practice can be found in the "YouTube voice". "YouTube voice" has not received much in the way of academic examination, but colloquially describes a kind of hyperarticulation affectation adopted by popular YouTube creators (Beck 2015). Therefore, YouTube voice presents an example of a speech style being established and performed almost exclusively on the internet and for the purpose of consumption by an online audience.

Phonetic Features and Social Meaning

A large body of work in sociolinguistics connects phonetic features, such as those of the "YouTube voice," to social meaning and social categories. A successful description of a new linguistic persona (online or otherwise), must do the same thing, establishing how the associated phonetic features can impact the overall persona being created. These meanings can include physical characteristics such as age, sex, and attractiveness, as well as social features including economic class and level of education, among others (Kreiman & Gerratt 2005). While phonetic features are employed by speakers, their interpretations are also subject to the demographics of the listener, whose own experiential and cultural background contributes to how they may understand the speaker's voice pattern (see Silverstein 2003). D'Onofrio (2015) notes how personas with social meaning can influence perception of linguistic features. Voice patterns are strongly tied to the speaker's identity and the listener's understanding of the speaker's identity beyond just the application of demographic information, but also as signaling membership to a certain group to which additional cultural assumptions and stereotypes may be applied. In these identity groups, demographics may overlap such that a superseding conceptualization is created: "Identities encompass macro-level demographic categories; local, ethnologically-specific cultural positions; and temporary and interactionally specific stances and participant roles"

5

(Bucholtz & Hall 2005: 592). Through a "persona style," linguistic features tend to index for both social groups and lower-level social meanings (Eckert 2008). Physical presentation through dress, makeup, etc. signaling membership to these groups, known as "embodied style," also contributes to identity performance (Eckert 2008).

<u>Language and Gender</u>

One way in which identity ties to voice is through gender. The language and gender subfield of linguistics is dedicated to investigating the way gender identity and perceptions about gender across demographic categories manifest through linguistic features. Lakoff (1973) characterizes a social structure in which "men's language" is perceived as the neutral, unmarked speech style, while "women's language," characterized by features such as tag questions, hedging, uptalk, etc., is marked as powerless and insecure, thus conferring upon women those same qualities. While Lakoff's original work has since been disputed, studies have generally shown how women's speech tends to be devalued relative to men (Bradley 1981).

Indeed, the denigration of (perceived) women's language encapsulates a larger societal norm. While a broad discussion of sexism lies outside the purposes of this study, a few key points are particularly relevant. Women regularly find their identities and roles reduced to that of "girls" while to call an adult man a "boy" is considered offensive (Richardson 1981). Pornographic material often emphasizes the youth of a female participant contextually, visually, and within the title of the video (Bridges, Wosnitzer, Sharrer, Sun, and Liberman 2010). Even though the participating woman is an adult, she is often made to adopt the mannerisms and appearance of someone much younger, thus equating to some extent sexuality with youth. Goffman (1971) draws attention to the fact that women's placement in advertisements shares commonalities with children, including being portrayed lying down on surfaces, being spoon-fed

by an adult (often a man), and bodily postures indicating shyness, submissiveness, etc. These equations receive a kind of "idealized" status in advertisements regardless of their frequency in reality. In this way, Western societal expectations and pressures for women force an image of childlike submissiveness within a larger patriarchal structure.

However, while Lakoff's work opened up a general discussion about the relationship between language and gender, other research has complicated "women's language" and "men's language" beyond the bilateral distribution of power described above. Hall (1995) gives credence to the idea that women may use devalued features stereotypically characteristic of women's speech in order to obtain power and craft new identities for themselves. One example is the phone fantasy hotline, where interviewed women enjoy being able to express creative uses of "women's language" in order to possess financial independence outside of the traditionally patriarchal workplace. These women convey images of a submissive, sensual woman exclusively through vocal performance regardless of their affinity for that image in their actual lives. While the popularity of the phone fantasy hotline has diminished with widespread internet access, modern equivalents may be drawn to restricted content distributed by women online, where a (typically male) audience must pay for the privilege of seeing or hearing a woman fit some sort of prescribed image, whether sexual or otherwise. In this way, some women have claimed financial, social, and sexual power even by using so-called "powerless" images and linguistic features.

Phonetic Features Associated with Gender

More specific speech styles include so-called "breathy voice" and "creaky voice"/vocal fry, which face conflicting perceptions by women and by the greater public at large. While both women and men do "speech performance," as established above, it is often women's speech

styles that receive public and academic attention, especially when "women's speech" deviates from some perceived norm (i.e. where men's speech is considered the neutral or prestigious standard) (Eckert & McConnell-Ginet 1992). Listeners judge creaky voice negatively for both women and men, though women are judged to a stronger extent. Women with creaky voices are judged to be less trustworthy, less competent, and less likely to be hired than women without, leading to the popular conclusion that elimination of creaky voice is necessary for women to be successful in the business market (Anderson, Klofstad, Mayew & Venkatachalam 2014). Simultaneously, production of creaky voice is correlated with lowering one's F0, meaning that some women may employ creaky voice in order to give off the impression that they are more dominant or educated as a result of a lower vocal pitch (Davidson 2020). Other ways speakers use creaky voice to indicate personality or identity include high schoolers who attempt to seem more laid-back than their peers (Pratt 2018) and a woman who connected vocal fry with her Chicano heritage and "tough" persona (Mendoza-Denton 2011). While some women naturally have "creakier" voices than others, they nonetheless can modulate the amount of "creak" permitted in order to suggest positive attributes and membership to a popular group. Multiple dimensions of perceived power dynamics overlap so that what may largely be shunned in the public eye has a subtle advantage in other contexts and communities of practice.

"Breathy voice" covers a different set of associations but can also suggest the identity of the speaker: voices judged to be more "breathy" than others are rated as more feminine (Borsel, Janssens, & Bodt 2008). In adoption of a breathy voice on the same fantasy phone lines discussed above, the speaker communicates a seemingly "aroused" state while establishing the listener as the dominant participant within the exchange (Hall 1995). In another study, breathy voice was associated with intimacy, friendliness, sadness, timidity, and feeling content and

relaxed (Gobl & Chasaide 2010). Therefore, from our understanding of the social impact of creaky voice and breathy voice, we can conclude that it is possible to indicate different emotional states and demographic information through specific methods of phonation.

More generally, studies conflict regarding the perception of vocal pitches. Across languages, men tend to have an average F0 value of approximately 120 Hz, while women's voices are closer to 210 Hz (Traunmüller & Anders 1995). American English values tend to be consistent with these findings, with one study finding a mean of 225 Hz for women and 116 for men (Yin et al. 2003). As mentioned above, men and women both rate lower-pitched voices as more confident, strong, and trustworthy than higher-pitched voices (Klofstad, Anderson, & Peters 2012). Additionally, women lower voice pitch parameters when speaking with men they perceived as more desirable (Pisanski, Oleszkiewicz, et al. 2018). At the same time, men tend to prefer women with higher-pitched voices, which they believe to indicate fertility and positive physical characteristics such as facial femininity and attractiveness (Fraccaro, Hones, et al. 2011). Since pitch and its associations are related to gender, we might expect that linguistic personas indexing for femininity will interact heavily with pitch.

On a larger scale, we may conclude that listeners believe that perceived voice characteristics and speech styles indicate traits about the speaker, and speakers may modulate their speech style in order to promote or discourage those associations. Phonetic features such as those employed in breathy voice, creaky voice, valley girl voice, etc. index for social meanings and complex social stereotypes like personas. Speakers may "perform gender" and construct gendered personas in different ways depending on context, stance, and relationship to larger societal conventions (see Butler 2011). While speech styles associated with women tend to be devalued at large, individual speakers may navigate gender with language in nuanced and

9

complicated ways. In constructing a linguistic persona, speakers take these effects into

consideration, and people using Mock Infantile Speech may be expected to do the same.

<u>Child Speech</u>

If we suggest that Mock Infantile Speech attempts to approximate linguistic features

linked with children, it is necessary to address what constitutes "child speech." It is true that

children also have their own speech styles, though these features are more associated with

language development and acquisition – and errors in that acquisition – than intentional

productions for some kind of social effect. Many of these errors can be traced to age and rate of

acquisition; in English, plosives, nasals, and glides /b, n, m, p, h, w, d, g, k, t, ŋ, j/ are acquired

first (2;0-3;11), /v, dʒ, s, tʃ, l, ʃ, and z/ are typically acquired next (4;0-4;11), and / ɹ, ð, θ, ʒ/ are

acquired last (5;0-6;11) (Crowe & McLeod 2020). In other words, while mispronunciations may

be present for many years at varying rates during language development, a child is more likely to

pronounce /θ/ wrong over /b/. In many circumstances, children substitute sounds for one another

when they share a place or manner of articulation: "the substitutions reflect the use of easier and

earlier sounds for those that are acquired later, and they are usually similar in that they possess

some, if not all, of the distinctive features of the correct phoneme" (Van Riper & Emerick 1990).

We are more likely to see errors with sounds that are similar, such as substituting /l/ and

/ɹ/, then replacing /w/ with something like /p/. Common patterns of errors include turning

fricatives and affricates into stops ([tɹi] 'three') and turning liquids into glides ([wik] 'leak')

(Van Riper & Emerick 1990). Reduplication, vowel epenthesis, reduction of complicated clusters

and sounds, deletion of final consonants and unstressed syllables, assimilation, etc.: all are

hallmarks of errors produced while the capacities for speech – physical and cognitive – are still

being developed (Van Riper & Emerick 1990). It might follow, then, that a person attempting to

adopt a childlike demeanor may do so by adopting some of these features in order to replicate how a child actually talks, errors and all. Even if not every feature is adopted consistently, the overall impression of child speech may be enough to index associated characteristics.

Mock Infantile Speech

With the creation of the internet, and a wider and newer community of practice, speakers can create new linguistic personas in order to access niche audiences from which they previously would have been isolated. We have already established that speakers may intentionally adopt linguistic features – whether segmental or suprasegmental – in order to index particular physical features or personality. I propose and attempt to describe a new kind of linguistic persona designated "Mock Infantile Speech" (MIS) that adopts features of child speech, creating an image of sensual hyperfemininity for consumption in online spaces such as TikTok and YouTube.

Mock Infantile Speech has many colloquial names on these social media sites, the most popular being the "uwu voice" in reference to the word *uwu* (/uwu/). "uwu" (or "owo") originated as a Japanese emoji conveying happiness, smugness, or a general aura of "cuteness" (the *u* symbols represent closed eyes, while *w* is an approximation of a smiling closed mouth) but has now become a sort of catchphrase indexing membership to the Mock Infantile Speech persona (Dictionary.com). Other monikers include the "loli voice" (in reference to Japanese *lolicon* media, which focuses on suggestive content of young girls (Gagné 2008)), "anime voice," and the "egirl voice" (*egirl* meaning internet ('e') girl, referencing another internet subculture focused on modern alternative dress, video games, and sexual content ranging from the subtle to the overt). These names are used by both speakers themselves and by those outside of the community; YouTube videos and TikTok accounts describing "uwu speech" are tagged

11

with these labels (GothLoliVA 2019; Anie PotatoNya 2019; vexzenie 2020). Speakers of MIS are usually depicted as attractive young women with an interest in video games as well as Korean and Japanese culture. They may be referred to as "egirls" (though there is a subtle distinction in aesthetic, both the "uwu" and egirl subcultures incorporate video games as cultural touchpoints). Other designations include "lolis" and "gamer girls."

Associated dress incorporates colorful, feminine clothing that often resembles that of popular animated characters (see Figure 1). Makeup includes excessive blush and highlighter to accentuate the eyes, and hair also may be colorful and elaborately styled. Several components of the "uwu girl" image index youth and femininity, including Japanese schoolgirl outfits, pigtails,

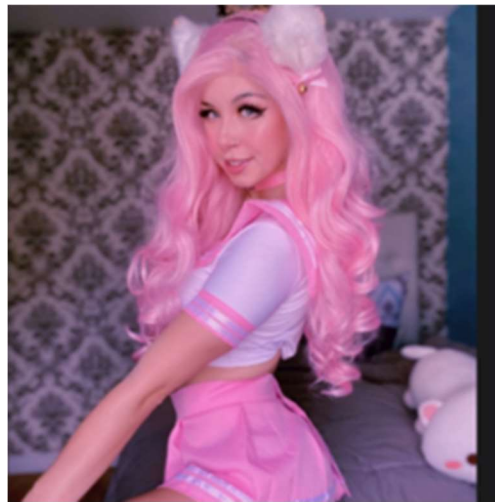Figure 1: Image of performers who use Mock Infantile Speech

Figure 1: Image of performers who use Mock Infantile Speech

and an emphasis on a "cute" environment, including small prepackaged snacks and the color pink. There is often a subtle sexual subtext to the "uwu girl" persona (hence reference to suggestive "lolicon" content), but this aspect ranges in degree of overt display. Many speakers

adopt the persona as a playful aesthetic choice, while others produce openly sexual content within and relying upon "uwu girl" subculture (including MIS). Gagne (2008) highlights how sexualization is sometimes intentional but occasionally imposed when talking about gothic Lolita subculture; the same argument may be extended here in discussion of "uwu girls". It should be noted that these characteristics are not exclusive to women, and men too can participate in Mock Infantile Speech and the associated social persona, although it is not as common.

The general origins of Mock Infantile Speech are unknown, other than that it is a relatively recent phenomenon. The oldest video sampled is from 2017; most other videos are from 2019 and 2020. However, it is likely that some inspiration for MIS comes from a (skewed and often stereotyped) Western perspective on East Asian cultures, particularly those of Korea and Japan. As mentioned above and elaborated further upon below, the MIS persona is heavily tied to an enthusiasm for East Asian culture, including anime, Japanese video games, Korean music, and cosplay (the practice of dressing up as a fictional character). Often this interest borders into orientalism, where fans of Japanese culture create a hegemonic, idealistic "Pure-Japan" through enthusiasm for Japanese cultural exports that nevertheless fails to accurately represent Japanese culture and history (Fliss 2012). Hiramoto and Pua (2019) emphasize the hypersexualization of East Asian women within Western media's James Bond franchise, a concept reinforced within the embodied style of the MIS persona when participants use Japanese schoolgirl outfits and styles in suggestive contexts. Participants often include Japanese and Korean phrases within English utterances as sexual or humorous non-sequiturs; within the space of the online Mock Infantile Speech community, these utterances help to stylize an identity merging Western cultures with racialized, sexualized ideas about East Asian cultures. This is not to say that every speaker who uses Mock Infantile Speech or the accompanying iconography is

attempting to reproduce a racialized image, or is even aware of connections to East Asian cultures at all. Instead, I propose that the MIS style has grown larger than this potential origin to encompass people using the persona who have no interest in Japanese and Korean media and music at all. Several speakers observed as part of this study appeared not to have a connection to East Asian cultures and simply produced Mock Infantile Speech on their own terms. Additionally, it should be noted that while demographic information for the speakers of Mock Infantile speech included here was impossible to verify, speakers generally tended to be either white or of East Asian descent and all native speakers of American English. Mock Infantile Speech is therefore not restricted to use by one culture or demographic group, but has a potentially problematic relationship with sexualization and stereotyping of East Asian cultures.

Financial incentives also contribute to the popularity of the style and the choice by performers to employ it. Apps like YouTube and TikTok allow payment depending on viewership and ad revenue, and the TikTok creator fund finances users who reach 10,000 followers (TikTok 2020). Particularly popular content creators may also advertise personal websites and receive sponsorships. On more private chat websites such as Discord, "uwu girls" are partly identified by the local prestige they acquire in male-dominated spaces (especially in online cooperative gaming) and the in-game or real-life currency they may accumulate as a result. It is commonly assumed that men are the primary consumers of "uwu girl" and "uwu speech" content for the purposes of sexual gratification, but this has not been reviewed empirically, and the female fans of such creators should not be disregarded.

Wider perceptions of "uwu girls" and MIS are generally mixed. Within the male consumer base, popularity of "uwu girls" is consistently high, as attested to by the large number of followers such accounts accumulate. A brief search on TikTok returns accounts using the

"uwu persona" (denoted by the visual cues including those above and hashtags like #gamergirl, #egirl, and #uwugirl) with 1.8 million; 1.2 million; and 4.9 million followers as of February 20, 2021. Some accounts within this study had follower counts of 170,000; 3487; and 2 million followers the time of initial sampling. People often comment supportively on such videos, including the statements "you're so cute" and "I'm your simp please notice me," referring to the practice of following a person's content because they are attractive (rusty.fawkes 2020). On videos where the speaker is not visible but still uses Mock Infantile Speech, it is not uncommon for people to leave comments asking the age and/or gender of the person behind the account: "When I want to simp but don't know the age". Comments like these indicate that there is often a sexual element to MIS, and commentors feel they must confirm that the speaker is a legal adult before they can continue sexualizing the voice in question. There is frequently a self-critical element to these comments, evidenced in the statements "it's to [sic] early in the morning to be back here" and "I was in public watching this [for real]" (vexzenie 2020). Both statements hint towards a sense of shame concerning the cultural stigma of openly consuming such content and the sexual subtext underlying them. Other comments referring to the "loli police" further this argument by bringing specific attention to the youthful persona of content creators and the moral and legal reprehensibility of sexualizing them.

Other positive approaches focus on the "uwu girl" aesthetic, and specifically Mock Infantile Speech, as a favorable behavioral model. Videos instructing how to adopt the "uwu voice" garner millions of views and positive feedback. Instructors take a folk-linguistic approach, including instructions such "pitch yourself up" (raise one's F0 values) and "soften your voice a little bit" (lower amplitude and make phonological contrasts less distinct). The following is from a TikTok titled "how to make a loli voice" with over 1.1 million views:

First of all, you need to close, close, close your vocal cords as possible. After that, please say "ah" [æ]. I think it will sounds like [æ], [æ], [æ]. Like that, right? To this, try to add a whisper sound. [h]. [æ]. It's gonna be like [hæ], [hæ], [hæ]. Then after this, you're gonna go high as possible, like [hæ], [hæ], [hæ], [hæ], [hæ], [hæ], [hæ], [hæ], [hæ]. *Onii-chan*! [laughs] (masaebation 2020).

This video identifies a few key practices considered central to MIS. One must "close the vocal cords," colloquially referring to the nasalization of produced phonemes. Adding a "whisper sound" is also important, meaning a lowering of amplitude, breathy voice phonation, decreasing phonological contrast, or increasing aspiration. Finally, as in the instructions above, hopeful MIS speakers are instructed to raise their pitch as high as possible. The phrase *onii-chan* is a Japanese word and accompanying honorific meaning "older brother," but within the "uwu girl" subculture has taken on an explicitly sexual connotation.

Another video on YouTube titled "uwu voice subliminal" has over 15,000 views as of April 19, 2021, and promises to subliminally alter the listener's voice to sound like Mock Infantile Speech. The accompanying video is a loop of an animated child with a song sung by someone speaking in MIS. The description of the video lists specifically some of the changes that viewers can expect, including adjustments to:

- Sound extremely adorable, soft, sleepy, and nervous during any emotion!
- Sound like a soft sleepy anime loli girl!
- Have the cutest pronunciation!
- Have the cutest soft Korean accent!
- Have a voice that reminds people of an anime girl!
- Have a voice that is able to persuade people!
- Have thin, short vocal chords (which makes ur voice higher)! (୨ ⌨ ୧ pwuppi 2020)

Comments on this video enthusiastically support and verify this change, stating "when I sneeze or cough it sounds a lot more cuter now [thank you so much]!" and "I listened once and I sound so adorable thank you." Commentors directly connect this voice to "uwu girl"

subculture, as indicated by one user commenting "Might as well say UwU."

These comments and videos above highlight a positive perspective on "uwu" subculture and Mock Infantile Speech; the subliminal YouTube video in particular underlines what specifically about the voice is considered desirable, including sounding "cute" and sounding like an animated East Asian character, with the effect being that the speaker is better able to affect and persuade their audience. From observing the popularity of creators online who are able to successfully produce the "uwu voice," we can confirm that such efforts are in some part successful.

While none of these videos outline specific phonological changes or articulatory gestures, they consistently highlight the importance of sounding "soft," indicating softness through a low amplitude, decreased phonological contrast, and raising one's pitch. The goal is implied to be that of a hyperfeminine, sensual, childlike persona, as testified to by the expressions *onii-chan* and "loli girl," as well as the explicitly stated intentions by instructors of MIS. Therefore, we might expect other speakers attempting to emulate Mock Infantile Speech to adopt other features from child speech that index youth and immaturity.

On the other hand, a large portion of online users appear to hold negative perceptions of the "uwu voice," extending that criticism to both the creators and consumers of "uwu girl" content. On a video where a woman uses Mock Infantile Speech to speak to her partner, comments such as "we use our grown up voice okay?" and "I would just leave the relationship if my girlfriend had a voice like this dang" garner hundreds to thousands of likes. Of course, these are also alongside comments like "Honestly tho [sic], if I date a girl who can do that voice I could never be mad at her" (Eliza 2020). One popular TikTok trend has been to parody gamer girls using Mock Infantile Speech, especially when such women are

perceived as adopting the voice in order to gain attention or money from men. Skits depicting

confrontations between "uwu girls" and girls speaking with their unaffected voices are

particularly popular, and almost always paint "uwu girls" as aggressive towards women they

perceive as threatening their prestige in male-dominated spaces. In this way we see how the

gamer girl persona, as well as Mock Infantile Speech in particular, is devalued outside of its

immediate consumer base. This attitude reflects an overall trend in devaluation of speech

styles associated with women as a whole, including valley girl speech (see Preston 1996 and

Villarreal 2016) and creaky voice (as discussed above).

In this study, I attempt to empirically establish what kinds of features are used to create

Mock Infantile Speech, and how they contribute to the "uwu girl" persona as a whole. Where

much creation of this new kind of linguistic persona happens online and without visual

accompaniment, it is important to discover what variables come into play and how and why

those variables index complimentary physical and behavioral characteristics.

METHODS

For my data I gathered video and audio clips from 14 speakers for a total of 1009 tokens across 23 clips. A "token" was defined as an environment where a phonological change in the target feature could occur, regardless of whether or not one did occur. 8 speakers were sampled from TikTok and 6 speakers were from YouTube. Videos were located primarily through the "search" features on both websites, which include both video titles and hashtags. Queries included "uwu voice," "anime girl," "anime voice," "loli voice," etc. Other videos were found through the TikTok For You Page feature, which algorithmically recommends and provides videos based off of recent account activity. In other words, TikTok recommended videos based off of my searches of these hashtags even when I was not looking within them at that time. These speakers appeared to be native American English speakers based on impressionistic judgements, though specific demographic information was not possible to collect online. Voice files were downloaded from their respective sources as MP3 files then converted to WAV files for analysis in Praat. Audio clips without background music or other speakers were preferred, but this was not possible for 3 speakers, though words and phonemes were still easily perceptible. Clips were chosen regardless of whether they showed specific phonological changes or not, though speakers who did not were preferred if they provided a contrasting "normal" voice. 5 speakers did not have other examples of their standard guise, so for those speakers, only clips where they use MIS are available for analysis. Though speakers would often contrast MIS with a "normal" guise within a video for contrastive or humorous purposes, it should be noted that during online speech performance, it is often difficult to verify even whether this "unaffected" guise was truly the

speaker's standard speech style. Therefore, it would be more accurate to state that the "standard" guise is one in which the speakers were clearly not indexing Mock Infantile Speech. Additionally, speakers 6 and 7 did not use any of the variables under analysis while using MIS, but were still able to successfully index for that linguistic persona (as indicated by hashtags such as #uwuvoice on videos and affirmative responses in comment sections), indicating that additional features may be at work in order to construct that specific identity. It is also for this reason that they are excluded from the analysis and graphical representations of phonological alternations below (though they were included within pitch measurements). Videos ranged in length from 5 seconds to 21 minutes, though multiple videos were often examined from speakers who provided only short clips. A large amount of data was coded from speaker 4 in particular because of the clarity of this speakers' audio, but this was accounted for in quantitative analysis through mixed-effects models with speaker as a random effect. Table 1 below outlines information about each speaker.

Data Coding

I used impressionistic coding in Praat in order to check for the presence or absence of specific phonological variables. Sociolinguistic research has a substantial grounding in impressionistic coding, and further studies have demonstrated its general accuracy and reliability (see Pope, Meyerhoff, Ladd 2007; Brown 2009; Hall-Lew & Fix 2012). While close analysis of phonetic details provides more precise readings, impressionistic coding is suitable for observing discrete, binary features such as those observed in this study, especially where spectrogram data corroborates such observations (Milroy & Gordon 2003). Annotations specifying word and phoneme were made within TextGrids. Tokens were then encoded in Excel sheets before analysis in R/RStudio. When the environment for a targeted feature was observed, I coded for

the speaker, clip (for multiple videos), word, preceding and following place of articulation,

phonological change in question, phonetic IPA transcription of the word, guise (standard or

MIS), context, and if the change was realized. I also noted in two additional columns the

underlying IPA phonemic transcription of the targeted phoneme for each token, and, if it

changed due to some phonological process, an IPA transcription of the final phonetic realization.

Table 1: Speaker details

Speaker	Gender	Platform	Number of Clips	Average Length of Clips Analyzed (seconds)
1	Female	TikTok	2	42.5
2	Female	TikTok	1	17
3	Female	TikTok	2	27
4	Female	YouTube	2	247.5
5	Female	TikTok	1	41
6	Female	TikTok	1	50
7	Female	TikTok	2	59
8	Female	YouTube	1	71
9	Female	YouTube	1	64
10	Female	YouTube	3	78
11	Female	TikTok	2	32.5
12	Female	TikTok	2	10
13	Female	TikTok	2	13.5
14	Male	YouTube	1	65

I have selected four linguistic variables that impressionistically appear to be particularly

salient features for signaling Mock Infantile Speech and the "uwu girl" persona as a whole, and

these variables were isolated for coding and analysis. This discussion's focus on them is not to

imply that other features do not also play a role, only that these features in particular were

relatively consistent and salient in their distribution.

The first variable analyzed is higher F0 formant values associated with higher pitch production. As outlined above, speakers of Mock Infantile Speech designate pitch as vital for an accurate "uwu voice." Additionally, it is well documented that Japanese women tend to speak with a higher pitch (averaging at about 230 Hz) compared to English speakers when indexing femininity (Loveday 1981; Hanley & Snidecor 1967; Tsuge et al 1987; Terasawa et al 1984). If the MIS guise is influenced by racialized understandings of Japanese and/or Korean speech styles just as the embodied style appears to be influenced by orientalist ideas about East Asian dress, we might expect English speakers of MIS to incorporate this idea about higher pitch values into the MIS style. If native English speakers' understandings of Japanese speech styles also come from Japanese media such as anime as opposed to interaction with native speakers in actualized environments, it is also worth noting that F0 values may be even higher, as average pitch of female anime characters tends to be higher than the average native Japanese speaker, at 250-500 Hz (Utsugi et al 2019). Pitch was measured by using the "get pitch" function in Praat, with measurements taken by hand for every continuous utterance. I used the standard Praat pitch setting for each speaker, 75 to 500 Hz. For longer clips, only the first 60 seconds were sampled for pitch measurements. Pitch measurements were divided into two categories based on the linguistic persona of the speaker at the time: Mock Infantile Speech and standard (unaffected speech). Where speakers switched between both guises within the same video, their F0 recordings were assigned to each category depending on the guise they were adopting at that time. 13 speakers were biologically female and identified as such, and one speaker was male; when speakers did not identify their sex or gender on their social media accounts, impressionistic judgements were made. Studies have shown that impressionistic judgements on speaker gender

based on voice are typically accurate regardless of voice modulation, including whispering and filtering (Lass et al. 1976). Once a series of pitch measurements were found for a speaker, these values were averaged in order to obtain the overall mean F0 for each speaker in each guise. These values were also averaged across speakers in order to find the overall mean pitch for each guise.

Sibilant Retraction

The first phonological change observed was the retraction of place of voiced and voiceless English alveolar fricatives /s/ and /z/ so that they become [ʃ] and [ʒ], respectively. For example, a speaker may pronounce [ʃəʊ] for the underlying representation /səʊ/ 'so', and [ʒibɹə] for /zibɹə/ 'zebra'. Like pitch, I also identify this variable as another place for potential racialized imitations of Japanese speakers; Japanese speakers tend to retract /s/ and /z/ before the vowels /i/ and /u/, but English speakers more unfamiliar with this allophonic variation may apply that change across phonetic environments (Ohata 2004). Tokens were coded regardless of place within the word. A phonological change was noted based on impressionistic coding of a contrastive difference. Marked changes were labeled "retraction" as opposed to "no retraction". Tokens were collected for every instance where speakers had the option of whether or not to retract a sound; in total, this amounted to 403 tokens.

Liquid Modification

Impressionistic coding also revealed marked changes in English liquids /ɹ l/. /ɹ/ (as in /ɹut/ 'root') is defined as a voiced alveolar approximant while /l/ is a voiced alveolar lateral approximant, as in /lʊk/ 'look'. /ɹ/ was also distinguished from the rhotic vowel /ɚ/ (as in /əsɚt/ 'assert'), where /ɚ/ is defined here as a rhoticized vowel with a lowered third formant (Ladefoged & Maddieson 1996). Additionally, a distinction was made between "light /l/" ([lɑɪ]

23

'lie') and "dark /l/" ([pʊɫ] 'pull'), where "dark l" has a secondary dorsal gesture opposed to "light l" (Oxley, Roussel & Buckingham 2007). Possible alternations included two processes: change to the voiced labiovelar glide /w/ (/laɪk/ 'like' as [wɑɪk], and /ɹaɪt/ 'right' as [wɑɪt]) and deletion of the liquid in its entirety (/wɛl/ 'well' as [wɛ]). I additionally noted derhoticization of the NURSE vowel, as in /hɚ/ 'her' as [hə]. While not the same as a direct deletion of this phoneme, derhoticization in this instance represents a general salient weakening of the rhotic vowel. These changes were encoded as "gliding" and "liquid deletion" (which also included derhoticization), respectively. If pronunciation did not change, tokens were marked as "liquid present." Like sibilant retraction, I noted each place where speakers had the opportunity to modify a liquid whether or not that modification actually occurred, resulting in 462 total tokens.

Interdental Stopping

Finally, I expected that speakers often changed the interdental fricatives voiceless /θ/ and voiced /ð/ to the stops /t/ and /d/, or occasionally deleting such consonants entirely. A speaker employing Mock Infantile Speech may, for example, pronounce /ðə/ 'the' as [tə] or [də], or /θink/ 'think' as [tink] or [dink]. This feature was observed impressionistically, but is also a common error made in child speech during language acquisition. Tokens were denoted as "present" if the fricative was produced as a stop, and "absent" if it was not. Opportunities for this variant were more infrequent, and 144 tokens were collected.

Statistical Analysis

For measuring differences in pitch between the mock guise and standard guise, I used a two-tailed t-test after establishing normal distribution of data in both guises with a Shapiro-Wilk test. For the categorical phonological variables, I used a logistic mixed-effects model with speaker encoded as a random effect to account for by-speaker variation in number of tokens.

Other variables, such as guise, were encoded as fixed effects accordingly. While I often coded

for other variables such as preceding and following place of articulation, these features were

largely not significant (unless otherwise noted). Table 2 summarizes each targeted feature

alongside examples and token counts.

Table 2: Targeted phonological features

Feature	Phonological Change	Example	Token Count
Sibilant Retraction	/s z/ → [ʃ ʒ]	/saʊ/ → [ʃaʊ]	403
Liquid Modification	/l ɹ ɫ ɚ/ → [w Ø]	/laɪk/ → [waɪk] /hɚ/ → [hə]	462
Interdental Stopping	/ð θ/ → [t d]	/ðə/ → [də]	144

RESULTS

Overall, all four variables targeted within this study proved to be significantly different in the Mock Infantile Speech guise as compared to the standard guise. This means that these features – pitch, sibilant retraction, liquid gliding and eliding, and interdental stopping – are all relevant features used in the construction of the MIS style and associated persona. While these variables are all relevant, it cannot be said that they are consistently relevant across speakers, or even within individual speakers. Therefore, I posit that while these features are important for performing Mock Infantile Speech, they are not the only variables at play. Detailed results and discussion follow below.

<u>Pitch</u>

The mean frequency for speakers using MIS was 349.2935 hertz (Hz), while speakers in the standard guise had a pitch average of 236.645 Hz. In other words, pitch was considerably higher when speakers adopted Mock Infantile Speech as opposed to when they did not. If we consider the five speakers who did not provide a standard guise counterpart, the mean F0 value was marginally higher at 351.183 Hz. The difference in mean pitch varied according to speaker, with speaker 4 having the smallest difference of 33.3 Hz while speaker 8 had the largest at 235.44 Hz. In all cases, speech within the "uwu girl" persona using MIS was at a higher pitch than speech without. Furthermore, the mean F0 value for "uwu speech" was markedly higher than mean F0 of expected modal speech for both men and women: compare 351.99 Hz to the standard 120 Hz and 210 Hz, respectively. The average pitch of uwu speech was in fact much closer to that typical of child speech, which ranges from 200 to 325 Hz for children, regardless of

gender (Kent 1976). Figure 2 below shows pitch values for each speaker data for "uwu speech" as compared to the standard guise when available.

Additionally, a two-tailed t-test demonstrated that the difference in pitch between the mock guise and the standard guise was significant ($p < .001$; df = 18.989; t-value = 4.3776). Therefore, speakers raise their F0 values during performance of the "uwu" identity compared to their unmarked speech styles, possibly serving to index femininity or a childlike presentation through a connection with the higher mean F0 of those groups.
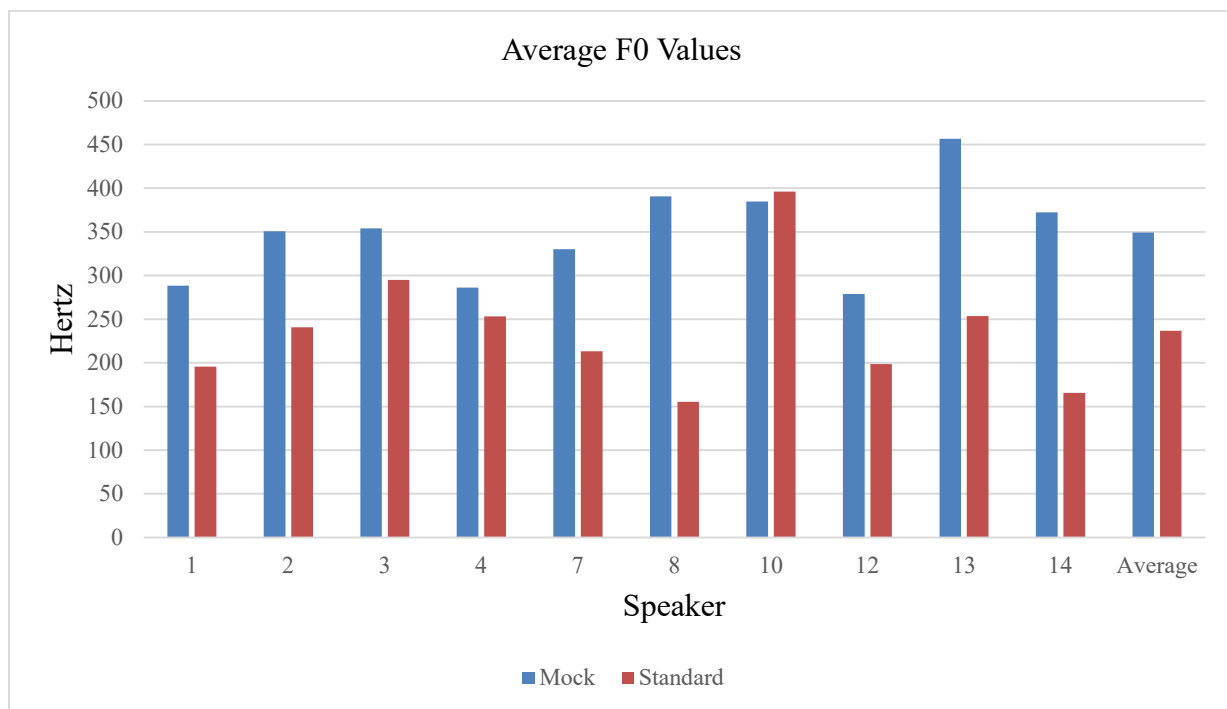
Figure 2: Average F0 values by speaker

Sibilant Retraction

Nearly all speakers demonstrated retraction of some kind, with retraction of /s/ to /ʃ/ being much more frequent than the voiced counterpart /z/, possibly due to the infrequency of /ʒ/ in English. Retraction was absent in speakers 2, 3, 6, 7, and 10; all other speakers demonstrated

at least some tokens of retraction in speech. Significant differences were observed between

guises, and the results are visible in Figure 3. Note that speakers who exhibited retraction at all

only applied it during Mock Infantile Speech. In other words, speakers marked a distinction

between guises by retracting /s/ and /z/ during MIS, while not retracting their alveolar fricatives

at all during the standard guise. This is also visible in Figure 3. I attempted to run a logistic

mixed-effects model in order to judge the statistical significance of retraction across speakers,

where speaker was coded as a random effect, while adopted guise and the phoneme in question

undergoing change (either /s/ or /z/) were all fixed effects, with realization was the dependent

variable. However, the lack of tokens of sibilant retraction in the standard guise meant
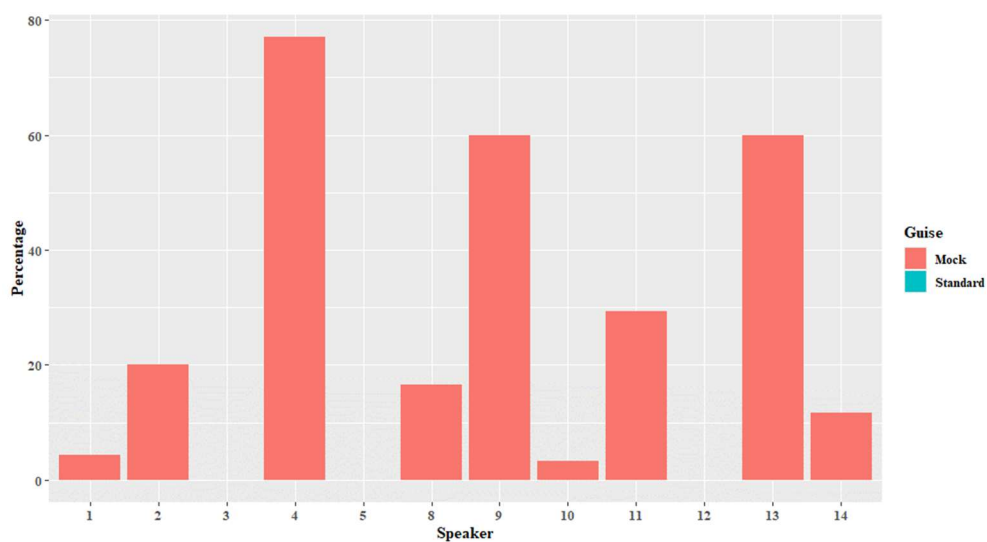
Figure 3: Percent sibilant retraction by speaker

that the model was unable to converge. Nonetheless, this finding reinforces the conclusion that

the two guises are significantly different in their rates of sibilant retraction. Speakers varied

greatly between each other and were not always consistent in realization of each phoneme. For

example, Speaker 2 retracted alveolar fricatives about 20% of the time in the Mock Infantile

Speech guise, while Speaker 9 demonstrated retraction 60% of the time (though neither speaker showed retraction in the standard guise). While more speakers on the whole retracted /s/, Speakers 8, 11, and 13 retracted /z/ more frequently than /s/, though it is unknown if this difference would hold with a larger available dataset for these speakers. Some speakers (3, 5, 6, 7, 10, and 12) did not show retraction at all, either in the mock guise or the standard guise.

Liquid Modification

For liquid modification, I created another generalized linear mixed effects model, which also highlighted the differences between gliding and deletion of liquids in the affected guise as opposed to the modal guise. Speaker was a random effect while the places of articulation and guise were fixed effects, with realization of the phonological change as the dependent variable:

(1) glmer(factor(Realization) ~ Guise + Following + Preceding + (1|Speaker)

The difference between guises was found to be significant ($p < .001$), indicating that speakers using Mock Infantile Speech generally glided and elided liquids at markedly different rates than speakers in the standard guise. The results of the model are visible in Table 3. Figure 4 shows the distribution of gliding and deletion of liquids broken down by speaker.

There were some differences found in the kind of phonological alternation related to the environment. Liquids tended to be modified in the mock guise following palatal and velar consonants and following vowels. Not all speakers demonstrated gliding or deletion of liquids: speakers 6 and 7 continued to enact no salient phonological changes at all. Additionally, speaker 14 did demonstrate deletion of the alveolar approximant /ɹ/ in the standard guise 50% of the time, but still deleted and glided other liquids in the mock guise more frequently, thereby still emphasizing this change as relevant to the construction of Mock Infantile Speech. Figure 5 compares which phonemes and allophones underwent modification.

Table 3: Liquid modification Model

Predictors	Odds Ratios	CI	p
		factor(Realization)	
(Intercept)	0.42	0.10 – 1.77	0.236
Guise [Standard]	0.01	0.00 – 0.04	**<0.001**
Preceding [bilabial]	0.97	0.37 – 2.55	0.946
Preceding [coda	0.49	0.17 – 1.40	0.183
Preceding [dental]	9.94	0.05 – 2190.41	0.404
Preceding [interdental]	0.00	0.00 – Inf	0.992
Preceding [labiodental]	0.99	0.28 – 3.49	0.987
Preceding [labiovelar]	0.44	0.06 – 3.00	0.401
Preceding [palatal]	12.84	1.43 – 115.36	**0.023**
Preceding [postalveolar]	1.57	0.08 – 29.05	0.762
Preceding [velar]	12.06	2.20 – 66.15	**0.004**
Preceding [vowel]	3.36	1.50 – 7.52	**0.003**
Following [bilabial]	0.49	0.10 – 2.37	0.377
Following [coda]	1.84	0.55 – 6.13	0.324
Following [glottal]	4.31	0.10 – 192.10	0.451
Following [labiodental]	0.91	0.11 – 7.80	0.928
Following [labiovelar]	0.84	0.04 – 18.46	0.914
Following [postalveolar]	22083448.68	0.00 – Inf	0.996
Following [rhotic vowel]	0.00	0.00 – Inf	0.998
Following [velar]	0.87	0.09 – 8.83	0.908
Following [vowel]	2.53	0.76 – 8.42	0.132

Random Effects

σ^2	3.29
$\tau_{00 \ Speaker}$	1.51
ICC	0.32
$N_{Speaker}$	12
Observations	461
Marginal R^2 / Conditional R^2	0.696 / 0.792
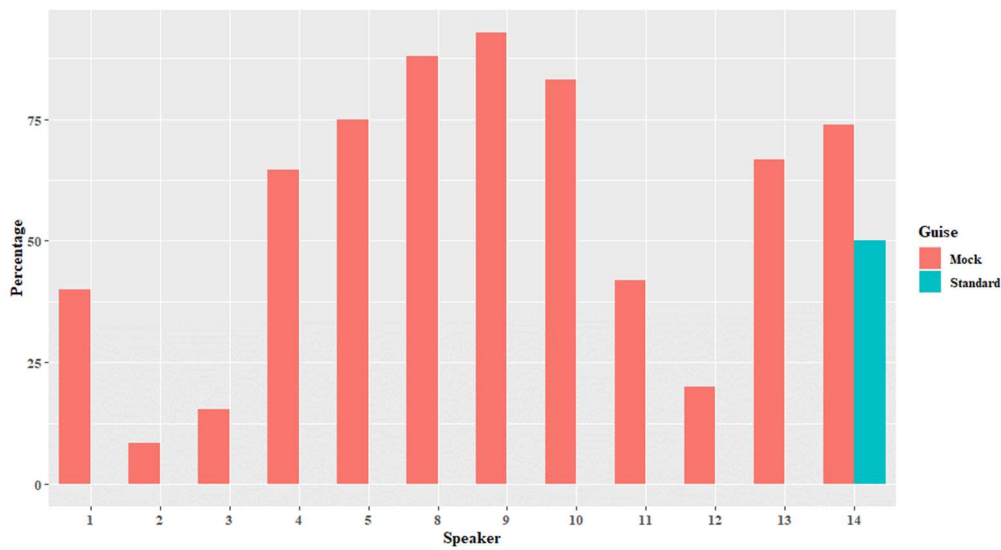
Figure 4: Percent liquid modification by speaker
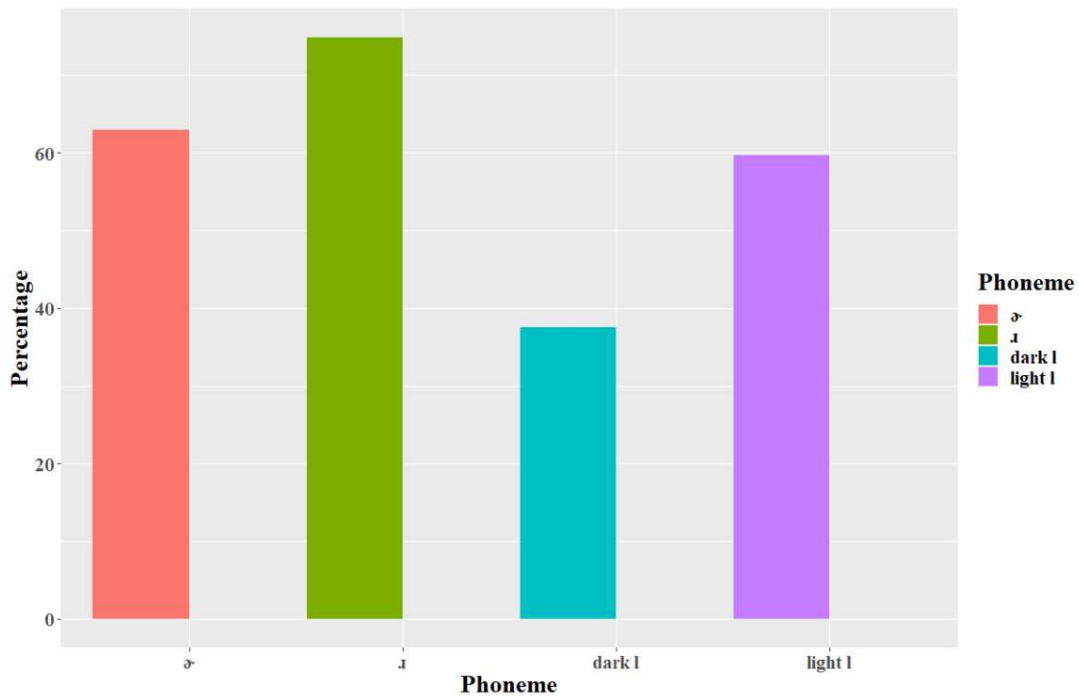
Figure 5: Percent liquid modification by phoneme

Some words were also strongly associated with gliding, particularly the word 'hello' (/hɛloʊ/), which at the surface level became [hɛwoʊ] in accordance with liquid gliding. Of the 24 tokens of 'hello' present within the collected data, 23 were changed in this way. This lexical item may have special salience for the persona, as speakers performing the "uwu persona" often tag videos with "#hewwo", "#imbaby", and "#uwu" (Lamia 2019; Lisa/Dylan 2021). More research would need to be done to conclude whether or not other words are themselves highly associated with the speech style, but "uwu" as an utterance (not just as an emoticon) also appears to be relevant to the persona.

Interdental Stopping

Finally, speakers often change the voiced and voiceless interdental fricatives /θ/ and /ð/ to the alveolar stops /t/ and /d/, respectively. While I attempted to run a logistic mixed effects

31

model to determine if the differences in reduction in the standard and mock guise were significant (using guise as a fixed effect and speaker as a random effect similar to the other models), I found that the model would once again not converge. However, interdental stopping only occurred for speakers in mock guise, so patterns are clear even without statistical modeling. In other words, speakers did not demonstrate interdental stopping in the standard guise at all, a fact reflected in the figures below. Figure 6 shows the percentage of realization for each speaker. As stated above, speakers 6 and 7 did not demonstrate any phonological changes, and thus did not alter any interdental fricatives. Speaker 9 also did not demonstrate reduction of interdental fricatives in her own Mock Infantile Speech. Notably, all other speakers demonstrated interdental stopping at least a fourth of the time. Speakers were more likely to stop the voiced fricative rather than the unvoiced, as indicated by Figure 7, though the presence of stopping for both phonemes indicates that both are associated with the "uwu persona." Interestingly, speakers did not necessarily alternate the fricatives according to voicing; for example, the voiced fricative /ð/ could become either /t/ or /d/ during realization. The same was true for the voiceless fricative /θ/, which was also stopped to /t/ and /d/.
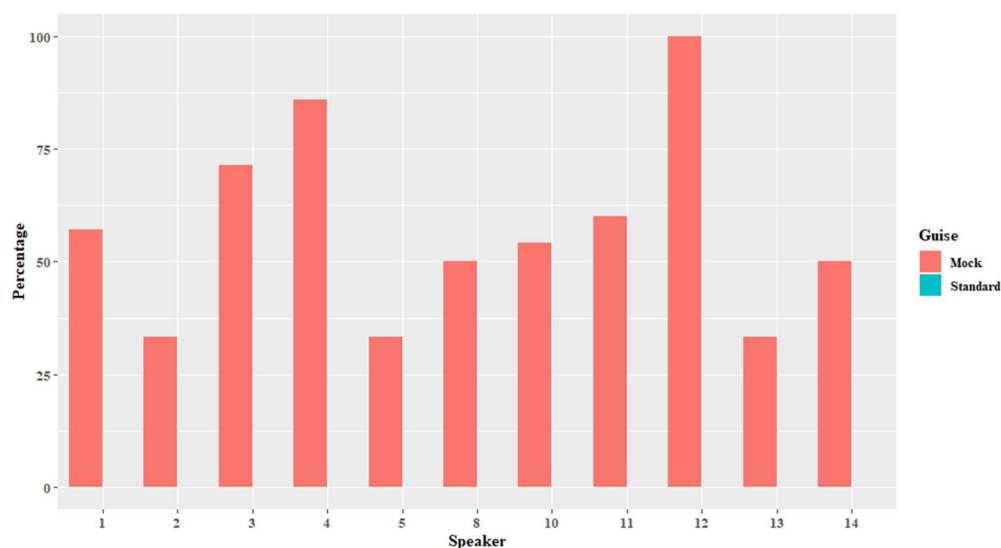
Figure 6: Percent interdental stopping by speaker
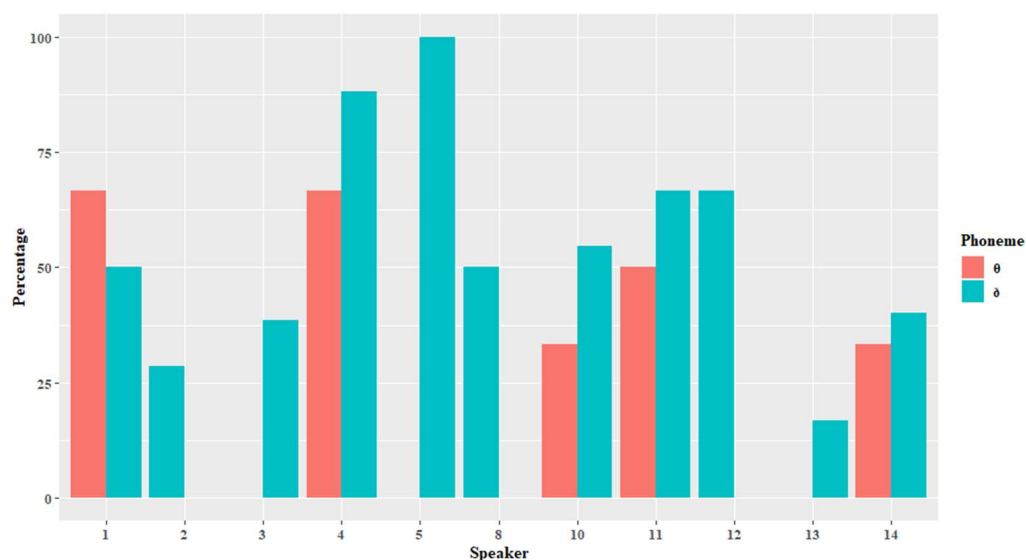
Figure 7: Percent interdental stopping by phoneme and speaker

Examined as a whole, raising F0 values, retracting alveolar fricatives, gliding and deleting liquids, and stopping of interdental fricatives all appear to be variables used in the production of Mock Infantile Speech, demonstrating that speakers employ specific features in order to index the "uwu" identity in opposition to unmarked speech styles.

DISCUSSION

Now that we have established that significant difference between the marked Mock Infantile Speech and unmarked standard guise exist, I turn to the implications and specific indexical qualities of this speech style. Additionally, I consider other phonological changes that support the construction of Mock Infantile Speech, but which occur less frequently.

First, as noted above, this generalized analysis shows that speakers tend to use both segmental and suprasegmental features in Mock Infantile Speech. All speakers who provided examples of Mock Infantile Speech and the unmarked guise demonstrated a significantly raised pitch at an average of 349.2539 Hz during "uwu" identity construction. Even when speakers did not have an unmarked counterpart, they still displayed F0 values remarkably higher than the average for both men and women. Though the vast majority of speakers sampled were women, this disparity in pitch also applied to speaker 14, who identified as male. This contingency suggests that while "uwu speech" is stereotypically associated with women, men too can successfully apply Mock Infantile Speech and its features. A high F0 range is more commonly associated with women's speech and child speech, which have higher mean F0 values at 210 Hz and 200-325 Hz, respectively. The high pitch found in Mock Infantile Speech therefore signals social associations with femininity and youth. Additionally, it may be posited that since MIS actually exceeds the standard F0 values for these categories, it calls forth a "hyperfeminine" and/or "hyperchildish" speech style, where the unrealistically high pitch aims not necessarily for accuracy, but for additional emphasis on those specific traits. If speakers also associate high pitch with orientalist understandings of Japanese speech styles, where Japanese women's high

mean F0 index femininity, and the average female anime character's pitch is even higher than that, then that may also be a contributing factor in production of high pitch in the Mock Infantile Speech guise as a way to index those cultural associations. At the same time, because not every speaker of MIS is involved in East Asian media as elaborated upon above, this additional hypothetical factor may not be generalizable to all speakers.

From pitch we consider retraction, gliding and deletion of liquids, and interdental stopping, the three phonological alternations isolated in this preliminary study. All of these changes proved to be statistically significant and likely salient in their differentiating of Mock Infantile Speech and the standard guise. Like any phonological alternation, these variables mean little in isolation, but in context liquid modification and interdental stopping target expected features of child speech in order to index youth and innocence to the listener, and sibilant retraction, while not typically associated with child speech, may index an "East Asian" component to the persona as well. As delineated above, children often struggle to pronounce phonemes they have not yet fully acquired. A child may simplify complex consonant combinations or miss target phonemes altogether. 'Right there' in child speech becomes 'wight dere' and 'hippopotamus' becomes 'hippopomus.' By replicating the same kinds of mistakes that children make in early language acquisition, adult speakers signal supposed membership to that same group. Levon (2014) relates stereotypes to "group concepts (e.g. *man*) with collections of both trait attributes (e.g. *athletic, domineering*) and roles (e.g. *father*)." Through applying that same lens here, I conclude that speakers of Mock Infantile Speech incorporating phonological alternations and mistakes from child speech do not just index themselves as "childlike," but also call into the mind of the listener related attributes (*youth*, *cuteness*, *submissiveness*) and roles (*child*, *dependent*). The listener has the choice of whether or not to respond positively, as

evidenced by the mixed perspectives towards Mock Infantile Speech described above. Even when a specific feature such as deletion of /l/ could potentially index for another identity (especially where the variation is part of another speech style such as African-American Vernacular English, see McLarty, Jones, & Hall 2019), it is the combination of these features as a whole that creates the "uwu" identity. The same can be applied to sibilant retraction as mentioned above, which is a feature often found in Japanese non-native English speakers. If speakers were universally imitating Japanese and Korean phonology, we would expect not deletion or gliding of liquids, but confusion between /l/ and /ɹ/, a common error in native Japanese and Korean speakers when acquiring English (Ohata 2004). Instead it might be said that imitating child speech is the primary component of successfully producing Mock Infantile Speech (indicated by the two alternations associated with child speech as opposed to one for Japanese phonology), but indexing an "East Asian" quality to the linguistic persona is of possible secondary nature.

It should be noted these phonological variables are not the only features of child speech, and are indeed not the only features that may be replicated. Instead, I hypothesize that they are targeted as specifically salient features intended to be perceived by the listener. They are based on the *idea* of what a child sounds like, not necessarily the reality of child speech and language acquisition. Other features present in child speech, such as vowel epenthesis and reduplication, may not signal child speech as strongly, and thus are not incorporated into Mock Infantile Speech and not found in the videos here, though more research would be necessary before making a definite conclusion. Speakers also varied in how frequently they adopted these phonological changes, even within their own speech. However, such variation can be expected in the construction of a linguistic persona; when performing a specific guise such as Mock Infantile

Speech, we can expect speakers to target the most salient features and places of articulation in order to give an overall impression of a persona without necessarily having to retract every individual phoneme for the same effect (see Eriksson 2010; Neuhauser 2008). All speakers did not display all phonological alternations either; speakers 6 and 7 did not display any of the variables noted here at all, and other speakers included some features but not all of them. This variability suggests that the phonological changes outlined here are not the only ones which contribute to MIS and the "uwu" persona.

Other variables not yet mentioned did appear in Mock Infantile Speech and also seemed to replicate features of child speech, though not at the same rate as the pitch raising, sibilant retraction, liquid gliding and deletion, and interdental stopping described above. Speakers 2 and 9 demonstrated a word-final (in all but one example) interdental lisp (fronting of the sibilant /s/ as in [ðɪθ] from /ðɪs/ 'this') in the mock guise but not the standard guise. While a lisp does not necessarily directly index child speech (more research has been done on the lisp as a stereotypical indicator of sexuality, see Mack & Munson 2012), it should be noted that Smit (1993a, 1993b) reported in a large-scale phonological study that the most common error for children acquiring /s/ was to pronounce the phoneme as [θ] or [ð]. Therefore, it would not be unlikely for speakers of Mock Infantile Speech to incorporate a lisp into their speech as a less salient quality of child speech. Speakers 1, 4, and 11 stretched the deletion of liquids during "uwu speech" to include stop consonants as well, an alternation once again not observed in the standard guise. For example, speaker 11 pronounced 'everybody' (/ɛvɹibədi/ as [ɛvwədi]). While it is not uncommon to delete consonants in casual speech, these speakers' tendencies to do so unpredictably while constructing the "uwu" persona but not otherwise potentially indicates an intentional decision to replicate another aspect of child speech. Finally, speakers 5 and 11 each

demonstrated one token of "TH-fronting," where /ð/ and /θ/, respectively, both were fronted to /v/, as in [smuvi] for /smuði/ 'smoothie' and [wɪv] for /wɪθ/ 'with'. Though this alternation was infrequent, its marked appearance in the mock guise suggests an additional avenue for speakers to index for youth and inexperience by incorporating phonological errors into their speech style. TH-fronting is not uncommon in dialects including Cockney and New York English (see Labov 2006; Ranzato 2019), but in the context of identity construction here, it can be assumed that these changes are not meant to stereotype or otherwise reference these other speech styles.

On the whole, I conclude that some linguistic variables including raising one's F0 values, retracting alveolar fricatives, gliding and deleting liquids, and changing interdental fricatives to stops, are important in the construction of Mock Infantile Speech, which speakers use in order to construct a childish, feminine identity in online spaces. Through applying specific, salient phonological features, they create a new linguistic persona to fit with the relevant subculture. The persona helps to characterize a social group (people who participate in social media exchanges and have an interest in subjects like gaming, anime, and "cute" aesthetics) and lower-level social meanings (including innocence, femininity, childishness, etc.), which are then expressed within individual exchanges where phonological features are put to use. Though speakers vary in frequency of different phonological changes, they all serve the same purpose in distinguishing Mock Infantile Speech from the standard guise. This is by no means an exhaustive review of Mock Infantile Speech and its phonetic and phonological qualities, but instead introduces MIS as a new, relevant speech style constructed in online spaces and attempts to connect its linguistic features to social meaning.

This analysis of Mock Infantile Speech places it in the larger context of women's connections to speech styles, sexuality, and internet culture. It is not news that speech styles

associated with women are typically devalued. However, women's use of MIS demonstrates the complicated relationship between gender and speech, in that some women feel empowered by the social and financial gains of "uwu speech," using it while promoting an online brand and creating YouTube videos discussing its supposed benefits, while others reject it as degrading, as evidenced by the negative comments and parodies left on videos of "uwu girls" using Mock Infantile Speech as described above. The women using Mock Infantile Speech and the larger "uwu persona" likely have varied motivations for their use of this online identity. However, there are implicit, potentially harmful associations of femininity and sexuality with immaturity and adolescence, even if speakers are not aware of it. Many content creators utilizing the "uwu persona" (as indicated by their use of MIS, hashtags on videos, and visual cues like those mentioned above) are adult women with sexually explicit content advertised on their platforms. Even if these women are not necessarily posting audio content with Mock Infantile Speech in their own voices, they often lip-sync to audio clips of other, real children for the same effect, strengthening the direct association of the "uwu" subculture with underaged individuals for consumers. Additionally, the MIS persona may contribute to the racial stereotyping of East Asian cultures and the creation of a Westernized notion of those regions. If sibilant retraction (and potentially high pitch) index East Asian cultures while coupled with patterns in child speech, there lies the risk of reinforcing orientalist assumptions of hypersexuality and submissiveness in Japanese and Korean women.

This review is not to say that women who use Mock Infantile Speech intend for these effects or are even aware of its potential harms. After all, the consumers of such content are the online users ultimately encouraging this persona; a parallel may be drawn to the adult film industry, where it is the buyers and watchers of explicit material – especially where such material

39

depicts women as underage even when they are not – who motivate its creation. Furthermore, not everyone who speaks in MIS intends to draw a connection to sexuality. However, for people who view such content, that distinction is not always clear, and as awareness and popularity of Mock Infantile Speech continues to grow, the negative consequences of a link between "uwu speech" and sexualization of underage women become clearer. The fact that companies have started to market to the "uwu aesthetic" with clothing and computer gaming supplies that have "uwu" as a phrase printed directly on them indicates that this community is expanding. With Mock Infantile Speech reaching more people, particularly through fast-paced and relatively unmonitored platforms like TikTok, its possible negative implications can reach a wider audience.

CONCLUSION

This review provides an initial description of the characteristics of this new internet-based persona, in particular focusing on the linguistic features of MIS and the qualities and roles such features index for. A high pitch, retraction of sibilants /s/ and /z/, alternation of liquids to become /w/ or deleted altogether, and stopping of interdental fricatives all contribute to the creation of this speech style and associated characteristics.

As stated above, this is by no means a complete analysis of "uwu subculture" and Mock Infantile Speech; that is left for future work. There is a need for a more comprehensive overview with more speakers and more audio clips. Speakers may also hyperarticulate vowels and/or shift the entire vowel space in an effort to reinforce associations with child speech. Children's anatomically shorter vocal tracts means that they have higher formant frequencies than the average adult, so a speaker approximating child speech may attempt to imitate those same acoustics (Vorperian & Dent 2007). A deeper investigation into the role of breathy voice may also indicate other phonation styles relevant to the production of Mock Infantile Speech. While I have also suggested that attempts to approximate linguistic features typical of Japanese and Korean speakers may also be a contributing factor to the Mock Infantile Speech guise and associated persona, more research into this relationship would be necessary before establishing a definitive correlation. Additionally, perceptual studies may further address the saliency of MIS features, the ability of speakers to distinguish the Mock and the standard guises, and how strongly individual linguistic variables index particular characteristics of speakers. Conversations with actual speakers of Mock Infantile Speech that help elucidate their methodology of

production and reasons for doing so would shed light on the larger cultural context of MIS and the individual motivations for participating in "uwu" culture. Where Mock Infantile Speech has largely yet to reach audiences outside of certain internet subcultures, perceptual studies have room to determine if other individuals can identify the persona and its associated indexical meanings.

We have largely discussed the implications Mock Infantile Speech may have for women and children existing on the internet, but these effects are worth repeating. A subtle implication that child speech is somehow desirable due to its inherent qualities of youth, immaturity, and hyperfemininity endangers the children growing up within those online communities. It is relevant to consider identities that are both growing in popularity and strengthening sociocultural ties between women, children, sexuality, language, and the internet.

REFERENCES

- Anderson, R., Klofstad, C., Mayew, W., & Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. PLoS One, 9(5), e97506. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0097506

- Anie PotatoNya. (2019, January 22). *How to make a loli voice* [Video]. YouTube. https://www.youtube.com/watch?v=VNtGN7htEeM&ab_channel=AniePotatoNya

- Anna. [@shaiwaifu]. 😅 *(also PSA grandpas house is in the middle of a remodel) it's all disaster up in here atm* [Video]. TikTok. https://www.tiktok.com/@shaiwaifu/video/6910322431680941318?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Anna. [@shaiwaifu]. *Ahhhh ty @jjpuffinstuff LOL* [Video]. TikTok. https://www.tiktok.com/@shaiwaifu/video/6908928152983342342?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- *ASMR - WEADING A STOWY ~ Reading a Bedtime Story about YOU (in OwO Voice) ~.* (2019, July 16). [Video]. YouTube. https://www.youtube.com/watch?v=qHMa1csyaj4&t=167s&ab_channel=SeafoamKitten%27sASMR

- *ASMR - MOISTURE EMERGENCY!! ~ Forcefully Moisturizing Your Crusty Hands ~.* (2019, November 29). [Video]. YouTube. https://www.youtube.com/watch?v=U4b66q26xJ4&feature=youtu.be

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

- Beck, J. (2015, December 07). Why do so many people on youtube sound the same? Retrieved February 21, 2021, from https://www.theatlantic.com/technology/archive/2015/12/the-linguistics-of-youtube-voice/418962/

- Borsel, J., & Janssens, J., & Bodt, M. (2008). Breathiness as a Feminine Voice Characteristic: A Perceptual Approach. Journal of voice: official journal of the Voice Foundation. 23. 291-4. 10.1016/j.jvoice.2007.08.002.

- Bradley, P. (1981) The folk-linguistics of women's speech: An empirical examination, Communication Monographs, 48:1, 73-90, DOI: 10.1080/03637758109376048

- Bridges, A. J., Wosnitzer, R., Scharrer, E., Sun, C., & Liberman, R. (2010). Aggression and Sexual Behavior in Best-Selling Pornography Videos: A Content Analysis Update. Violence Against Women, 16(10), 1065–1085. https://doi.org/10.1177/1077801210382866

- Brown, E. K. (2009). The relative importance of lexical frequency in syllable-and word-final/s/reduction in Cali, Colombia. In Selected Proceedings of the 11th Hispanic Linguistics Symposium (pp. 165-178). Somerville, Massachusetts: Cascadilla Proceedings Project.

- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. Discourse Studies, 7(4–5), 585–614.

- Butler, J. (2011). Gender trouble: Feminism and the subversion of identity. Routledge.

- Cato. [@anacreons]. (2020, October 9). *#pov the #gamergirl on the #discord gets an unexpected call... yes this is my real voice!1!!1! #uwu #kitten #fyp #GimmeSomeTruth* [Video]. TikTok. https://www.tiktok.com/@anacreons/video/6881836091624803589?is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Cato. [@anacreons]. (2020, October 17*). #pov the gamer girl gets saved and also makes a new friend (?) #pov #fyp #gamergirl #kitten #discordkitten #uwu #girlinred* [Video]. TikTok. https://www.tiktok.com/@anacreons/video/6884785034511797509?is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Chang, W. (2014). extrafont: Tools for using fonts. R package version 0.17. https://CRAN.R-project.org/package=extrafont

- Crenshaw, S. [@possibly.a.fae]. (2020, December 9). Reply to @soupslayer7 love you two🥴 *#lolivoice #lolivoiceimpression #loli #anime #weeb #weebtiktok #fyp #fyp ツ #foryoupage #animegirls #animegirl* [Video]. TikTok. https://www.tiktok.com/@possibly.a.fae/video/6904142210606861573?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Crenshaw, S. [@possibly.a.fae]. (2020, December 9). *Reply to @minty.man it's okie I hate me too 💖 #fyp #lolivoice #fyp ツ #loli #anime #weeb #weebtiktok #animegirl #foryoupage #animegirls #egirl #egirls* [Video]. TikTok. https://www.tiktok.com/@possibly.a.fae/video/6904140552715963654?lang=en&is_cop

y_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102
533

- Crowe, K., & McLeod, S. (2020). Children's English Consonant Acquisition in the
United States: A Review. American Journal of Speech-Language Pathology, 29(4),
2155–2169. https://doi.org/10.1044/2020_AJSLP-19-00168

- Crystal, D. (2001). *Language and the internet*. ProQuest Ebook Central
https://ebookcentral.proquest.com

- *Curious Elf inspects you* 🎄 *(⌣ ‿ ⌣✿) [uwu voice] [ASMR]*. (2020, December 15).
[Video]. YouTube.
https://www.youtube.com/watch?v=nSbwD99K1uk&t=398s&ab_channel=OopsydaisyA
SMR

- D'Onofrio, A. (2015), Persona-based information shapes linguistic perception: Valley
Girls and California vowels. J Sociolinguistics, 19: 241-256.
https://doi.org/10.1111/josl.12115

- Davidson, L. 2020a. The versatility of creaky phonation: Segmental, prosodic, and
sociolinguistic uses in the world's languages. WIREs Cognitive Science n/a(n/a). e1547.
https://doi.org/10.1002/wcs.1547.

- Davis, B. H., & Brewer, J. P. (1997). *Electronic Discourse : Linguistic Individuals in
Virtual Space*. SUNY Press.

- 🐬lisa/dylan🐬. [@_utsugi_]. (2021, Februrary 17). omg don't kinshame me
😥‼️#MiFamilia #kins #kinnie #😥 #uwu #bruh #subscribetowafflessed #kin
#kightyagami #yagamilight #pp #lol #hewwo #jacksmells #hi #shawty #yuh [Video].
TikTok.

https://www.tiktok.com/@_utsugi_/video/6930414786593213701?lang=en&is_copy_url=0&is_from_webapp=v3&sender_device=pc&sender_web_id=6923022839206102533

- Eckert, P., & McConnell-Ginet, S. (1992). Language and Gender. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139245883

- Eckert, P. (2008), Variation and the indexical field1. Journal of Sociolinguistics, 12: 453-476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

- Eliza, D. [@darlaeliza]. (2020, December 17). the correct answer is three year anniversary. #fyp [Video]. TikTok. https://www.tiktok.com/@darlaeliza/video/6907405739543842054?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Eriksson, A. (2010). The disguised voice: imitating accents or speech styles and impersonating individuals.

- Fliss, C. J. M. (2012). Communal Identity through Cultural Essentialism: The Evolution of the American Anime and Manga Fan Community and the Orientalism of its Conception of Japan.

- Fraccaro, P., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C., Feinberg, D. R., Little, A., & DeBruine, L. (2011). Experimental evidence that women speak in a higher voice pitch to men they find attractive. Journal of Evolutionary Psychology. 9. 57-67. 10.1556/JEP.9.2011.33.1.

- *fruit time ~ BUT WITH VOICES UwU*. (2019, November 16). [Video]. YouTube. https://www.youtube.com/watch?v=nFafFG78idY&feature=youtu.be

- Gagné, I. (2008), Urban Princesses: Performance and "Women's Language" in Japan's Gothic/Lolita Subculture. Journal of Linguistic Anthropology, 18: 130-150. https://doi.org/10.1111/j.1548-1395.2008.00006.x

- Gobl, C. & Ní Chasaide, A. (2010). Voice Source Variation and Its Communicative Functions. In The Handbook of Phonetic Sciences (eds W.J. Hardcastle, J. Laver and F.E. Gibbon). doi:10.1002/9781444317251.ch11

- Goffman, E. (1979). Gender advertisements (1st Harper colophon ed.). Harper & Row.

- GothLoliOVA. (2019, March 8). Loli voice tutorial uwu (i'm still new to this) [Video]. YouTube. https://www.youtube.com/watch?v=nBBkSfA0h5U&ab_channel=GothLoliOVA

- Hall, K. (1995). Lip Service on the Fantasy Lines. 10.13140/2.1.4942.2402.

- Hall-Lew, L., & Fix, S. (2012). Perceptual coding reliability of (L)-vocalization in casual speech data. Lingua, 122(7), 794-809.

- Hanley, T. & Snidecor, J. (1967). Some Acoustic Similarities among Languages. *Phonetica*, *17*(3), 141-148. https://doi.org/10.1159/000258584

- *Hewwo. . .? UwU*. (2019, October 26). [Video]. YouTube. https://www.youtube.com/watch?v=zgb-8UAXogc&feature=youtu.be

- *hewwo mr obama?* (2017, November 11). [Video]. YouTube. https://www.youtube.com/watch?v=RCB9p8o7iaM&feature=youtu.be

- Hiramoto, M., & Pua, P. (2019). Racializing heterosexuality: Non-normativity and East Asian characters in James Bond films. Language in Society, 48(4), 541-563. doi:10.1017/S0047404519000381

- Iqbal, M. (2020, October 22). TikTok Revenue and Usage Statistics (2020). Retrieved October 31, 2020, from https://www.businessofapps.com/data/tik-tok-statistics/

- Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. Journal of speech and hearing Research, 19(3), 421-447.

- Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. Proc. R. Soc. B.2792698–2704http://doi.org/10.1098/rspb.2012.0311

- Kreiman, J., Vanlancker-Sidtis, D., & Gerratt, B. R. (2005). Perception of voice quality. In D. Pisoni & R. Remez (eds.), Handbook of speech perception (pp. 338–362). Malden, MA: Blackwell

- Kuznetsova A., Brockhoff, P.B., & Christensen, R.H.B. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, *82*(13), 1-26. doi: 10.18637/jss.v082.i13 (URL: https://doi.org/10.18637/jss.v082.i13).

- Labov, W. (2006). The Social Stratification of English in New York City (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618208

- Ladefoged, P., & Maddieson, I. (1996). The Sounds of the World's Languages. Cambridge: Blackwell.

- Lakoff, R. (1973). Language and Woman's Place. Language in Society, 2(1), 45-80. Retrieved February 7, 2021, from http://www.jstor.org/stable/4166707

- Lamie, A. [@astridlamia]. (2019, November 10). As much attitude as I could muster 😂 #dojacat #hewwo #imbaby #softgirl #pastel #softaesthetic #pastelgoth [Video]. TikTok. https://www.tiktok.com/@astridlamia/video/6757785790161456390?lang=en&is_copy_url=0&is_from_webapp=v3&sender_device=pc&sender_web_id=6923022839206102533

- Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. The Journal of the Acoustical Society of America, 59(3), 675–678. https://doi.org/10.1121/1.380917

- Lave, J., & Wenger, E. (1991). Situated learning : legitimate peripheral participation. Cambridge University Press.

- Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. Language in Society, 43(5), 539-566. doi:10.1017/S0047404514000554

- Loveday, L. (1981). Pitch, Politeness and Sexual Role: An Exploratory Investigation into the Pitch Correlates of English and Japanese Politeness Formulae. Language and Speech, 24(1), 71–89. https://doi.org/10.1177/002383098102400105

- Lüdecke, D. (2018). "sjmisc: Data and Variable Transformation Functions." Journal of Open Source Software, *3*(26), 754. doi: 10.21105/joss.00754 (URL: https://doi.org/10.21105/joss.00754).

- Lüdecke, D. (2020). sjlabelled: Labelled Data Utility Functions (Version 1.1.7). doi: 10.5281/zenodo.1249215 (URL: https://doi.org/10.5281/zenodo.1249215), <URL: https://CRAN.R-project.org/package=sjlabelled>.

- Lüdecke, D. (2021). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.7, <URL: https://CRAN.R-project.org/package=sjPlot>.

- Mack, S. & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. Journal of Phonetics - J PHONETICS. 40. 10.1016/j.wocn.2011.10.002.

- maiseyolleh. [@maiseyolleh]. (2020, August 10). *this is why tiktok wont let me interact w ppl #lolimeme #loli #meme #lolivoice #lolivoiceimpression* [Video]. TikTok. https://www.tiktok.com/@maiseyolleh/video/6859540379050446085?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- maiseyolleh. [@maiseyolleh]. (2020, August 10). *I love my cat and paranormal shit but someone explain #realityripple* [Video]. TikTok. https://www.tiktok.com/@maiseyolleh/video/6859370780438826246?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Masae. [@masaebation]. (2020, July 11). *@skittles.brat さんへの返信 how to make loli voice #lolivoice #animevoice #เสียงอนิเมะ* [Video]. TikTok. https://www.tiktok.com/@masaebation/video/6848169017362926849?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- McLarty, J., Jones, T., & Hall, C. (2019). Corpus-based sociophonetic approaches to postvocalic r-lessness in African American Language. *American Speech*, *94*(1), 91-109.

- Mendoza-Denton, N. (2011). The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. Journal of Linguistic Anthropology, 21(2), 261–280.

- Milroy, L., & Gordon, M. J. (2003). Sociolinguistics : method and interpretation. Blackwell Pub.

- Muffy. [@itsurdreamgirl]. (2020, August 23). *#kawaiivoice #voiceover #uwuvoice #fyp* [Video]. TikTok. https://www.tiktok.com/@itsurdreamgirl/video/6864321931051552005?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Neuhauser, S. (2008). Voice Disguise Using a Foreign Accent: Phonetic and Linguistic Variation. International Journal of Speech Language and the Law. 15. 131-159. 10.1558/ijsll.v15i2.131.

- Ohata, K. (2004). Title: Phonological Differences between Japanese and English: Several Potentially Problematic Areas of Pronunciation for Japanese ESL/EFL Learners. Asian EFL Journal. 6.

- Olivia. [@ramen0odle]. (2020, July 18). Here's part 2 #POV for you guys lmao #fyp #discord #gamergirl [Video]. TikTok. https://www.tiktok.com/@ramen0odle/video/6850977115303644422?lang=en&is_copy_url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=69230228392061025 33

- Oxley, J., Roussel, N., & Buckingham, H. (2007). Contextual variability in American English dark-l. Clinical linguistics & phonetics, 21(7), 523–542. https://doi.org/10.1080/02699200701356485

- Pinheiro J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152, <URL: https://CRAN.R-project.org/package=nlme>.

- Pope, J., Meyerhoff, M., & Ladd, D. R. (2007). Forty years of language change on Martha's Vineyard. Language, 83(3), 615-627.

- Pratt, T. (2018). Affective sociolinguistic style: An ethnography of embodied linguistic variation in an arts high school. (PhD dissertation). Stanford University, Palo Alto.

- Preston, D. R. (1996). Where the Worst English Is Spoken. In E. W. Schneider (Ed.), *Focus on the USA* (pp. 297–360). John Benjamins Publishing Company; Groos. https://doi.org/10.1075/veaw.g16.16pre

- *Queenie Goldstein comforts you [ASMR] [Fantastic Beasts & Where to Find Them]*. (2020, January 11). [Video]. YouTube. https://www.youtube.com/watch?v=zs7jPk0uJYA&feature=youtu.be

- R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

- Ranzato, I. (2019). The Cockney persona: the London accent in characterisation and translation. Perspectives Studies in Translatology. 27. 235-251. 10.1080/0907676X.2018.1532442.

- Renei [@heyrenei]. (2021, February 9). *#pov pt.5: does Mia get to stay in the server w her?? #fyp #duetme #foryou #uwu #enemiestolovers #discord #acting #wlw* [Video]. TikTok. https://www.tiktok.com/@heyrenei/video/6927318745232739590?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Richardson, L. (1981). Feminist frontiers: Gender stereotyping in the English language. New York, NY: The McGraw-Hill Companies, Inc.

- Rusty.fawkes [@rusty.fawkes]. (2020, October 18). *you're watching as I'm trying to get us into the club lmao ❤️ my wig ohno #cringe* [Video]. TikTok. https://www.tiktok.com/@rusty.fawkes/video/6884832614587305221?lang=en&is_copy _url=0&is_from_webapp=v1&sender_device=pc&sender_web_id=69230228392061025 33

- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. Language & Communication, 23(3-4), 193–229. doi:10.1016/s0271-5309(03)00013-2

- Sky, Aurae. [@bigyikes420]. (2019, September 18). *I hate mtself so much 😂#foryourpage #fyp #foryou #frankenoc #kawaii #uwu #lolivoice #thatsrelatable #noob #twosides #* [Video]. TikTok. https://www.tiktok.com/@bigyikes420/video/6738139639825075461?is_copy_url=0&is _from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Smit, A. (1993a). Phonologic error distributions in the Iowa-Nebraska articulation norms project: consonant singletons. Journal of Speech and Hearing Research, 36, 533-547.

- Smit, A. (1993b). Phonologic error distributions in the Iowa-Nebraska articulation norms project: consonant clusters. Journal of Speech and Hearing Research, 36, 931-947.

- *SPOOKY ASMR - Hewwo? OwO ~ The Adorable Organ Harvester ~*. (2017, October 13). [Video]. YouTube. https://www.youtube.com/watch?v=7bSdXgXZQa4&feature=youtu.be

- Terasawa, R., Kakita, Y., & Hirano, M. (1984). Simultaneous measurements of mean air flow rate, fundamental frequency and voice intensity: Results from thirty normal male and thirty female subjects. Onseigengoigakir, 25, 189 - 207.

- TikTok. (2019, August 16). TikTok creator Fund: Your questions answered. Retrieved February 10, 2021, from https://newsroom.tiktok.com/en-gb/tiktok-creator-fund-your-questions-answered

- Traunmüller, H. & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. 2.

- Tsuge, S., Kakami, K., & Fukuya, M. (1987). Speakers' height, weight and voice pitch (1): On correlation of speakers' height, weight, and pitch frequencies of Japanese five vowels. Birll. Phoizef. Jap., 185, 2 - 5.

- Utsugi, A., Wang, H., & Ota, I. (2019). A Voice Quality Analysis Of Japanese Anime.

- uwu. In Dictionary.com. Retrieved February 2, 2021, from https://www.dictionary.com/e/slang/uwu/

- Van Riper, C., & Emerick, L. L. (1990). Speech correction: an introduction to speech pathology and audiology (8th ed.). Prentice Hall.

- Vex. [@vexzenie]. (2020, August 30*). #uwu #uwuvoice #weeb #anime #loli #weebs #weebo #animees #animegirl #animeboy* [Video]. TikTok. https://www.tiktok.com/@vexzenie/video/6866682994426498310?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Vex. [@vexzenie]. (2020, December 9). *#i #think #vex #has #lost #it #!!! #ThinkingAbout #HouseTour #uwuvoice #animeboy #weeb #anime #pog #animegirl* [Video]. TikTok. https://www.tiktok.com/@vexzenie/video/6904153922907163905?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Vex. [@vexzenie]. (2020, May 5). *UWU?#uwu #uwuvoice #loli #lolita #lolivoice #lolivoiceimpression #owo #owowhatsthis* [Video]. TikTok. https://www.tiktok.com/@vexzenie/video/6828885658610732294?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Vex. [@vexzenie]. (2020, Jun 16). Pleasesimpthanks *#UwU #UwUgirl #uwuvoice #owo #owowhatsthis #loli #lolita #lolivoice #lolivoiceimpression* [Video]. TikTok. https://www.tiktok.com/@vexzenie/video/6838885558719122694?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Villarreal, D. (2016). 4. "Do I Sound Like A Valley Girl To You?" Perceptual Dialectology And Language Attitudes In California. Publication of the American Dialect Society. 101. 55-75. 10.1215/00031283-3772901.

- Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: a synthesis of acoustic and anatomic data. Journal of speech, language, and hearing research : JSLHR, 50(6), 1510–1545. https://doi.org/10.1044/1092-4388(2007/104)

- Wekesa, M. (2019). Hate online: The creation of the "Other", Lodz Papers in Pragmatics , 15(2), 183-208. doi: https://doi.org/10.1515/lpp-2019-0011

- Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

- Wickham, H. (2020). tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

- Wickham, H. & Bryan, J. (2019). readxl: Read Excel Files. R package version 1.3.1. https://CRAN.R-project.org/package=readxl

- Wickham, H. & Seidel, D. (2020). scales: Scale Functions for Visualization. R package version 1.1.1. https://CRAN.R-project.org/package=scales

- Wickham, H., François, R., Henry, L. & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. R

- package version 1.0.4. https://CRAN.R-project.org/package=dplyr

- Yan, Q. & Vaseghi, S. & Rentzos, D. & Ho, C. & Turajlic, E. (2003). Analysis of acoustic correlates of British, Australian and American accents. 345 - 350. 10.1109/ASRU.2003.1318465.

- Zehra. [@prnzsssn]. (2020, August 23). *we gotta be nice to our female community but eh I'm new to that so I've never experienced this lol. The vid is just a joke ramen0odle #fyp #discord* [Video]. TikTok. https://www.tiktok.com/@prnzsssn/video/6864215883858988294?lang=en&is_copy_url =0&is_from_webapp=v1&sender_device=pc&sender_web_id=6923022839206102533

- Zhao, S. Y. Stop-like modification of the dental fricative /ð/: an acoustic analysis. J Acoust Soc Am. 2010;128(4):2009-2020. doi:10.1121/1.3478856

- Zhu, C., Xu, X., Zhang, W., Chen, J., & Evans, R. (2019). How Health Communication via TikTok Makes a Difference: A Content Analysis of TikTok Accounts Run by Chinese Provincial Health Committees. International Journal of Environmental Research and Public Health, 17(1), 192. MDPI AG. Retrieved from http://dx.doi.org/10.3390/ijerph17010192

- Zuo, H. & Wang, T. Analysis of TikTok User Behavior from the Perspective of Popular Culture. *Frontiers in Art Research* (2019) Vol. 1 Issue 3: 1-5. https://doi.org/10.25236/FAR.20190301.

- 夢 *uwu :: — uwu voice subliminal* 63. (2020, December 18). [Video]. YouTube. https://www.youtube.com/watch?v=5AnS07G4sCI&feature=youtu.be

- [Digital image of TikTok user Dessycc]. (n.d.). https://www.reddit.com/r/Dessyyc/comments/juyss4/24_hours_old/