SOUNDING LIKE A STEREOTYPE:
THE INFLUENCE OF EMOTIONAL PROSODY ON RACE PERCEPTION

Rachel Elizabeth Weissler, Julie E. Boland, Ph.D.

University of Michigan
racheliw@umich.edu

ABSTRACT

This study investigates how perception of race and perception of emotion influence one another. Participants listened to isolated words from an African American English (AAE) speaker and a Mainstream American English (MAE) speaker in happy, neutral, and angry prosodies, and indicated perceived race and emotion of the speaker. Though an effect of race on prosody was not found, results did show that MAE was rated whitest in the happy condition, whereas AAE was rated blackest in neutral and angry conditions. This could be because culture-specific processes affect how emotions are processed in speech (Pell & Skorup 2008). Varied perceptions of emotional prosody could lead not only to misunderstandings but also perpetuate dialect discrimination, making emotional prosody a crucial point of further inquiry for sociophoneticians.

Keywords: prosody, emotional prosody processing, AAE, perception

1. INTRODUCTION

This paper investigates whether perception of race influences perception of emotion and vice versa when hearing isolated words. This research is informed by the emotional prosody processing literature [1, 3, 7, 9, 12, 13, 14], the sociophonetic literature on perception [2, 10, 15] and research on the Angry Black Woman trope [4, 5, 6].

1.1. Emotional Prosody

Prosody, is a suprasegmental feature that can convey prominence and emphasis in speech [3]. Emotional prosody has been defined as emotion-specific prosodic patterns that elicit listener perceptions of various emotional states. The field of emotional prosody is relatively small and heterogeneous, focusing mostly on Indo-European languages. Happiness and sadness are the typical, binary emotions of focus. Explorations of emotional prosody have used both behavioural [7, 12, 13], and neuroscience (e.g., fMRI, EEG) methodologies [3, 9].

1.1.1. Emotional Prosody Processing

The emotional prosody processing literature is even more narrow; it consists mostly of studies sampling from English-speaking populations (i.e. Standard/Mainstream American English), reifying standard language ideologies that dismiss dialects and non-standard varieties. In a meta-assessment of the emotional prosody field, Bak found that very few studies have explored emotional prosody processing outside speakers' L1 or native dialect [1]. Even so, the literature does point toward gender differences in emotional prosody processing, and gender stereotyping is often involved in emotion recognition from prosody [3]. It seems likely, then, that there could be differences in emotional prosody processing linked to other social identifiers.

The most closely related prior research was done by Pell & Skorup [14]. They investigated the implicit processing of emotional prosody in a foreign accent and found that when listening to pseudo-Arabic words that contained happy, sad, and neutral prosody, Mainstream American English (MAE) listeners were able to detect the emotion being indicated by the prosody [14]. However, it's possible that listeners need more exposure to emotional prosody in Arabic to make emotional prosody judgements on Arabic-accented speech. While Pell & Skorup argued that culture-specific processes affect how emotions are processed in speech, they did not consider how prejudice and stereotyping might also play a significant role in emotion processing. These culture-specific processes are defined there as, "display rules" [9]. These display rules are culture-specific social norms that regulate how emotions are expressed in socially appropriate ways. For example, in a comparison of Western cultures versus East-Asian cultures, Western cultures tend to be more individualistic while East-Asian cultures tend to be more collective [9]. In behavior, these cultural norms can be borne out in adopting actions such as indirectness and avoiding eye contact in East-Asian cultures, whereas Western cultures encourage overt expressions of emotion and interaction, such as eye contact. It has thus been surmised that display rules play a central role in emotion perception [9].

Kim & Sumner [7] posited that emotional prosody maps directly to social concepts and representations, thereby influencing the spoken word recognition process. Their work investigated whether emotional information, when carried by a semantically emotionless word, could influence the mental lexicon during online processing. Before running an affective priming paradigm, Kim and Sumner had naïve listeners rate their auditory stimuli to ensure that the stimuli were categorized by the particular prosody type they were seeking. They worked with angry, happy, and neutral prosody types. They only used a Standard American English woman's speech for the stimuli. In their study, they show that angry prosody, when preceded by a non-emotional word uttered in angry prosody, results in affective priming. Affective priming did not show the same effect for happy and neutral prosody, which both had the same effect of association strength. One reason postulated by the authors was that neutral sounding-speech possibly sounded 'happy enough', especially coming from a woman's voice which is already stereotyped as agreeable and friendly [7].

1.2. Stereotypes Influencing Perceptions of Identity

Women's language is described as being more agreeable-sounding [8]. The question remains: If all women's voices are stereotyped as happier and more agreeable, how much nuance exists in listeners' perceptions of emotion? Additionally, how might raciolinguistic stereotypes play a role in discerning emotional prosody?

Examining the influence of stereotypes on perception is not new to the field of psycholinguistics. Previous research has shown that identical speech samples are perceived differently based on different social guises, such as varied faces or varied demographic information as evidenced by Rubin [15] and Niedzielski [10]. Rubin [15] shows that visual priming of ethnicity (using ethnically Asian and European faces) can affect listeners' evaluations of identical speech samples, where an Asian face prime results in higher ratings of accentedness and lower listener comprehension. Additionally, Niedzielski [10] shows that priming with varied social information (in this case, which city a speaker is from) can affect listener's evaluations of identical speech samples, where, among Detroit participants, the speaker said to be from Canada was perceived as using diphthongs in speech, while the speaker said to be from Detroit was not perceived to be using diphthongs. These are just two studies within a large body of literature that reminds us that social information and stereotyping have a significant influence on speech perception and must be considered when doing perception experiments.

1.3. The Angry Black Woman Trope

As listener ideologies about certain speaker groups inevitably inform their sociophonetic perceptions, it is important to keep in mind social variables when looking at speech perception, as ideologies about certain speaking groups can inform those perceptions. Many African American English suprasegmental features are stigmatized, such as the falsetto register being an indicator of indignation in AAE by Nielsen [11].

The Angry Black Woman (ABW) trope is deeply entrenched in American society and is still pervasive today, [2][4][5]. With its origins in slave times, the ABW trope caricatured enslaved woman who were forced to suppress emotion. When these women did express anger, they were labelled as aggressive and overbearing, [4]. Marcyliena Morgan has claimed that the reason why the ABW trope is so sparsely studied is because researchers have accepted the stereotype to be true. Black women's anger is not seen as legitimate and is often misconstrued as an issue of control according to Perry [5]. Even today, black women often end up in situations in which they are discriminated against, due to this pervasive concept of black women as degenerate according to Jones & Norwood [6].

Perceptions of black women's speech can not only create disadvantages for black women in the U.S., but even more dire, can have life-threatening consequences. In "The Killing of an 'Angry Black Woman': Sandra Bland and the Politics of Respectability," by Gillon [4], a description of the murder of Sandra Bland is put forth with a critical eye toward the legality of the arrest. At the time of her arrest, Bland was pulled over but did not get out of her car or put out her cigarette when the officer asked her to, saying things like, "I'm in my car, why do I have to put out my cigarette," and, "You seem very irritated" to the officer. Her reluctance to put the cigarette was an action read as violent to the officer, and, "This idea of Bland being a threatening individual already existed because of the significations that said black women are inherently belligerent It is an expectation that black women will be harder to control" [4]. Thus, we see here the murder of a perceived ABW, resulting from the linguistic discourse between her and the officer. This provides anecdotal evidence that the effect of emotional prosody can be life threatening.

2. METHODS, DATA, AND PREDICTIONS

In order to investigate the role of stereotypes in emotional prosody processing, this study looks at how emotional words were rated for emotion and race depending on language variety. The stimuli were isolated words, based on the claim that a lexical item as short as "Hello" can indicate to listeners whether or not the utterance was in MAE or African American English (AAE) according to Baugh [2]. In addition, isolated words in L1 and a foreign language have been successfully used to evoke perceptions of emotional valence [7, 14].

2.1. Methods and Data

The auditory stimuli come from two women native to San Diego, California; one black AAE speaker and one white MAE speaker, both in their mid-twenties. Each speaker recorded the 24 words in Table 1 in the three emotional prosodies, used by Kim & Sumner [7]. This resulted in a 3 x 2 (i.e., emotional prosodies by speakers) design, with six conditions (HappyWhite, AngryWhite, NeutralWhite, HappyBlack, AngryBlack, and NeutralBlack), resulting in 144 stimuli. Both the choice of emotional prosodies and the lexical stimuli were taken from Kim and Sumner [7].

Table 1: List of semantically non-emotional words pretested by Kim & Sumner [7].

academy	adequate	after	bounce	collide	compact
compose	conference	copy	galaxy	garbage	listen
ministry	multiply	pending	pineapple	planet	question
recall	salary	scale	specialist	stem	transmission

One hundred U.S. based participants were recruited using Amazon Mechanical Turk to complete an Auditory Rating task, using the 144 stimuli. 87% of these participants identified the racial demographics of their K-12 schools as "Mostly White." Participants were randomly assigned to one of six groups. Each group listened to the 24 words – four in each of the six conditions – pseudo-randomized using a Latin square design. Participants listened to each audio file individually and were asked to identify the speaker's race and mood for each word, before hearing another word. They were given a binary choice for the race ("Black" or "White") and three options for emotion ("Happy," "Angry," and "Neutral"). Participants were also given a language history survey and asked

what they thought the study was about (only responses to the latter question will be referenced to in this paper). Subject means were submitted to an Analysis of Variance (ANOVA) because we were looking for significant differences across three independent groups.
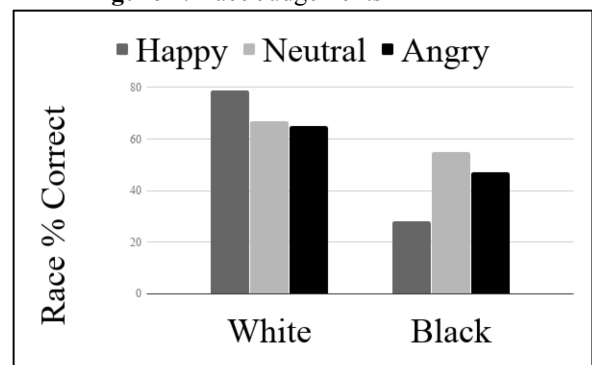
2.2. Predictions

If stereotypes do, in fact, influence how emotional prosody is perceived, race judgements should interact with emotion judgements. Specifically, (1) Happy guises will be perceived as more white, (2) Angry guises will be perceived as more black, and (3) the NeutralBlack guise will be perceived as more angry that the NeutralWhite.

3. RESULTS

A repeated measures ANOVA on race judgements found an interaction between RACE and EMOTION, $F(2, 198) = 43.43$, $p < 0.01$, despite a higher "White" base-rate of correct responses. This interaction is consistent with the first two predictions regarding happy and angry prosody: As seen in Figure 1, the White voice was rated whitest in the happy prosody condition, while the Black voice was rated blackest in the neutral and angry prosody conditions.

Figure 1: Race Judgements

In Figure 1, we can see a clear opposing pattern in terms of which voices are being perceived as which races, across three emotions. It is also clear that there is low race-rating accuracy in all emotional prosody conditions, especially for the Black voice.

Tables 2 and 3 present the proportions of Happy and Angry judgements, respectively, for all six conditions. There is no support for the third prediction, that the NeutralBlack guise would be read as angrier than the NeutralWhite guise. Across all trials, the Neutral guise was very rarely rated as Angry.

Table 2: Percent Happy Responses

	Happy	Neutral	Angry
White	93.5	6	21
Black	68	7	7

Table 3: Percent Angry Responses

	Happy	Neutral	Angry
White	2	14	42
Black	7	16	23

The observed patterns were replicated in an in-person study using the same stimuli, but with undergraduates as the population, rather than the Amazon Mechanical Turk population with ages ranging from 18-65. In both versions of the experiment, participants were not very accurate at identifying the race of the speaker, with an overall bias to identify the speaker as white.

4. DISCUSSION

The results from this study provide evidence that emotional prosody influenced race judgements, with the Happy guise overwhelmingly perceived as white, However, there was a clear low base rate for identifying the accuracy for identifying the Black voice as black, leading to a surprisingly low accuracy in the race identification component. In turn, this makes it difficult to conclude anything from the absence of a race effect on emotion identification

The tendency to identify both voices as white raises the possibility of an expertise effect – perhaps listeners are more accustomed (e.g., from media influence) to hearing a wider range of emotions from white female voices than from black female voices. This may have implications for mental representations of emotional black voices. Additionally, it is possible that culture-specific processes could affect how emotions are processed in speech, as Pell and Skorup mention [14].

It is possible that participants were reticent to make binary racial choices, as some of their responses on the post-survey questionnaire provide support for predictions that were not borne out in the data. When asked about the purpose of the study, one respondent said, "When the voice was higher and positive with good enunciation, I tended to lean towards white and when it was low, a bit aggressive and/or laxed in enunciation, I leaned towards black." Another respondent said that they perceived the study to be about, "Determining if people think black speakers are more likely to be angry compared to white

speakers even if they both are speaking in an objectively neutral tone." These responses suggest that lay people have a sense that race differences can influence emotion perception and vice versa.

The in-person replication allowed us to evaluate the possibility that Amazon Mechanical Turk workers were providing low-quality judgements. Turkers tend to complete surveys as quickly as possible, and it is possible that race and emotion judgements were not made carefully. However, the fact that the low base rate and white response bias were replicated in our laboratory suggests that accurate race identification requires more than a single word.

5. CONCLUSION

The goal of this study was to examine the relationship between the perception of emotional prosody perception and race. This experiment suggests that cultural stereotypes in U.S., such as the ABW trope, have influences on the intersection of race perception and emotional prosody perception. Our results are consistent with Bak's theory that identity plays a role in emotion recognition from prosody, not only through gender stereotyping, but evidently also through racial and raciolinguistic stereotyping [1].

Our low base rates of race accuracy for the Black voice are someone analogous to Pell & Skorup's postulation that listeners need more exposure to prosodic information in non-mainstream languages or varieties in order to make an emotional prosody judgement [14].

The current results have implications for related explorations in psycholinguistics, sociophonetics, and online processing in general. Speech perception researchers should keep in mind that listeners have knowledge of multiple languages and grammars apart from their L1, such as MAE speakers having knowledge of AAE. Furthermore, listeners can employ their knowledge of those grammars, as well as their stereotypes about a that speaker, to form expectations about the speaker. As discussed previously, this can have effects that not only lead to disfluency and miscommunication, but also have the potential to put the lives of black women and black people at risk. This work has implications for dialect discrimination research, which, if taken seriously in the public sector, has the potential to change the way people are treated in their day-to-day lives based on how they speak.

7. REFERENCES

[1] Bak, H. 2016. Emotional Prosody Processing for Non-Native English Speakers. Switzerland: Springer International Publishing.

[2] Baugh, J. 2000. Linguistic Profiling. New York: Routledge Publishing.

[3] Dahan,D. 2015. Prosody and Language Comprehension. New York: Wiley Periodicals. Inc.

[4] Gillon, V. 2015. The Killing of an 'Angry Black Woman.': Sandra Bland and the Politics of Respectability. Illinois: Augustana Digital Commons.

[5] Harris-Perry, M. 2011. Sister Citizen: Shame, Stereotypes, and Black Women in America. New Haven: Yale University Press.

[6] Jones, T., Norwood, K. 2017. Aggressive Encounters and White Fragility: Deconstructing the Trope of the Angry Black Woman. Iowa: Iowa Law Review.

[7] Kim, S., Sumner M. 2017. Beyond lexical meaning: The Effect of Emotional Prosody on Spoken Word Recognition. United States: Journal of the Acoustical Society of America.

[8] Lakoff, R. T. 1975. Language and woman's place. New York: Harper & Row Publishing .

[9] Liu, P., Rigoulot S., Pell M. 2014. Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia,* 67, 1-13.

[10] Niedzielski, N. 1999. The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*. United Kingdom: SAGE Publications.

[11] Nielsen, R. 2010. "I ain't Never Been Charged with Nothing!": The Use of Falsetto Speech as a Linguistic Straregy of Indignation. Pennsylvania: Penn Libraries.

[12] Pell, M.D. 2005a. Nonverbal emotion priming: evidence from the'facial affect decision task'. New York: Springer Publications.

[13] Pell, M.D. 2005b. Prosody-face Interactions in Emotional Processing as Revealed by the Facial affect decision task. Heidelberg, Germany: Springer + Business Media, Inc.

[14] Pell, M., Skorup, V. 2008. Implicit Processing of Emotional Prosody in a Foreign Versus Native Language. Amsterdam, Netherlands: Elsevier.

[15] Rubin, D.L. 1992. Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking teaching assistants. New York: Springer Publications.