



Holliday, Nicole R. & Burdin, Rachel Steindel & Reed, Paul E. 2024. Can listeners differentiate varieties using prosodic information? Lect identification in a persona-matching task. *Glossa: a journal of general linguistics* 9(1). pp. 1–39. DOI: <https://doi.org/10.16995/glossa.10124>

Open Library of Humanities

Can listeners differentiate varieties using prosodic information? Lect identification in a persona-matching task

Nicole R. Holliday, Pomona College, US, Nicole.Holliday@pomona.edu

Rachel Steindel Burdin, University of New Hampshire, US, Rachel.Burdin@unh.edu

Paul E. Reed, University of Alabama, US, Pereed1@ua.edu

Sociolinguists and lay people have long been interested in the mechanisms by which dialect and speaker identification occur. The current study aims to test whether realizational differences found in production studies are salient factors in identification of different lects, as well as to examine differences between groups in identification. In this study, listeners were instructed to match clips containing H* and L+H* pitch accents with specific personae who represented speakers of five lects of interest (African American English, Jewish English, Appalachian English, Southern English, and Midwestern English). Listeners from different groups (Appalachian, African American, Jewish, and those who were not from any of these groups) were presented with both low-pass filtered and unfiltered versions of the clips. Results indicate that listeners performed at or above chance in the unfiltered condition. However, we did observe a consistent pattern such that listeners confused AAE and JE voices, as well as Appalachian and Southern voices, indicating that listeners may struggle to differentiate lects with similar historical development and/or prosodic patterns. Additionally, listeners showed sensitivity to the original lect in selecting a persona in the filtered condition, though they were less accurate overall compared to the unfiltered condition. Overall, listeners appear to be more sensitive to segmental phonological than prosodic information in lect identification. Additionally, when presented with a personae matching task, listeners from a variety of backgrounds perform with a high degree of accuracy, though examining patterns of confusion help to shed light on the variables that are most salient.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 OPEN ACCESS



1 Introduction

The mechanisms by which listeners observe and classify dialect differences has been a major question in sociolinguistics since the field's inception. Indeed, non-linguists display a persistent fascination with the ability to discern regional and ethnic characteristics of a speaker based on their production of specific words. One prominent example of this is the popular "Accent Tag" series on YouTube, which contained more than 40,000 videos recorded between 2010–2015.¹ In the "Accent Tag", users answered a series of questions about what they call specific items (e.g. "What do you call gym shoes?") as well as read a word list that contained items known to display regional variation in vowel production (e.g. *aunt*, *pecan*, *caught*) (Sellers 2014). The "Accent Tag" phenomena shows that while some elements of how listeners observe dialect variation are clearly at the level of the lexicon and/or syntax, phonetic and phonological variation are also key. While most previous studies have focused on segmental features, particularly vowels, the current study aims to test whether listeners can accurately identify and categorize speakers of different ethnic and regional varieties, as represented by specific personae, even when segmental cues are not available.

Over the past several decades, lect identification studies have focused on the role of phonological and phonetic phenomena in triggering listener judgments of speaker region and/or ethnicity. Thomas (2002) provides a thorough review of the literature in this vein, from early works in the 1960s, through studies in the early 2000s. Since 2002, such studies have expanded in both number and scope, heeding the call put forth by Thomas and others. However, the majority of such studies have focused specifically on vocalic phenomena to the exclusion of other types of variables that may trigger listener judgments. Additionally, most studies were designed to ask about one aspect of speaker identity (e.g., race, region, etc.), often using written prompts (e.g., "Where is this speaker from?"), and audio-only stimuli. Other work has examined the role of visual and other biographical information in the perception of speaker characteristics (e.g., Niedzielski 1999; McGowan 2015; McGowan & Babel 2020). However, few studies have focused on a listener's ability to match voices to speakers with particular characteristics. This represents a gap in the literature, especially considering the role that personae play in the development of indexical links to particular lects and groups. The current study begins to address these gaps by examining whether listeners are sensitive to prosodic patterns, in addition to segmental features, by presenting listeners with both low-pass filtered and unfiltered speech, and proposes a new methodology for addressing how listeners make the link between prosodic features and types of regional and ethnic personae. Specifically, this work examines listeners' ability to identify three lects which have previously shown to display difference prosodic patterns: African American English, Jewish English, and Appalachian English.

¹ <https://www.pri.org/stories/2015-07-21/accent-quiz-swept-english-speaking-world>.

The paper begins with an exploration of previous literature on the role of the social information, as well as prosodic information, in lect identification. Finally, it motivates the methodology for the current study via discussion of results of prosodic differences in production studies, presents the results, and discusses implications for sociolinguistic theory and future work. Overall, results indicate that listeners are generally not especially sensitive to prosodic differences alone, and that they may struggle to differentiate lects when they share overlapping features, supporting results of previous studies (e.g. Clopper & Pisoni 2004b; Clopper et al. 2006).

1.1 Lect identification and listener knowledge

The categorization of lects, particularly different dialects, has been a major area of sociolinguistic research for decades. From Preston's pioneering work on perceptual dialectology (1986; 1996, *inter alia*) to modern studies that isolate sociophonetic variables, linguists and lay people continue to work to understand how judgments about speakers and places are made. Excellent reviews of the literature on region and identification can be found in both Thomas & Reaser (2004) and Clopper & Pisoni (2006).

In general, speakers perform well in lect identification tasks. Clopper & Pisoni (2004a) found that naïve listeners performed above chance when categorizing speakers from six U.S. dialect regions, and that miscategorizations made by the listeners were primarily between dialects with higher degrees of phonetic similarity as characterized by their vowel systems. Clopper & Pisoni also found that those with more experience with diverse dialects did better at identifying different dialects than those with less experience. Other work has explored how listeners' personal experience with the lects in questions, and/or their identity may affect their performance on dialect classification tasks. Thomas & Reaser (2004) found that Black listeners were more accurate than white listeners at identifying Black speakers from a small, rural community. Burdin (2020) likewise found that (a) Jewish listeners, and (b) non-Jewish listeners who reported familiarity or interaction with Yiddish speakers were more likely to identify particular intonational contours as coming from Jewish speakers compared to non-Jewish listeners who didn't report familiarity with such speakers.

Additionally, other work has shown that listeners integrate social information, including visual information, in their perception of particular types of speech. Niedzielski (1999) found that the perception of vowels as having the Northern Cities Vowel shift varied depending on whether listeners were told that the speaker was from Canada or Detroit. McGowan & Babel (2020) in a study of Spanish as spoken by Spanish and Quechua speakers found both (a) a shift in categorical boundaries, and (b) variation in qualitative feedback about speech depending on whether listeners were told that the speaker was a native speaker of Quechua or of Spanish. McGowan (2016) found differences in intelligibility of speech depending on what sort of face (white or Asian) was paired with Chinese-accented and non-Chinese accented English. Even when

listeners are not given a richer social context, they bring expectations about how speech is linked to multiple aspects of identity to perception tasks. Plichta & Preston (2005), for example, found that listeners rated /ɑɪ/ monophthongization not only as more “Southern”, but also frequently as more “male”. Burdin (2020) found that listeners not only rated particular types of speech as “Jewish” but also associated them with a particular type of Jewish speaker—one that is older, and a speaker of Yiddish.

These studies all show that listeners do not simply make a judgment about a listener’s region/ethnicity/race in a vacuum, but rather alongside other perceivable social information (Thomas & Reaser 2004), as well as their own identity and experiences. Indeed, modern sociolinguists have theorized that the social meanings of specific linguistic variables are invariably tied to both ongoing social change and fluid identities. Eckert (2008) defines the indexical field, arguing that “the meanings of variables are not precise or fixed but rather constitute a field of potential meanings – an indexical field, or constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable. The field is fluid, and each new activation has the potential to change the field by building on ideological connections” (454). As a result, in any perception task, linguists must be aware that for any given linguistic or social variables we believe that we’re testing, there is no way to totally isolate variables from one another.

One strategy for addressing this issue is to present listeners with tasks that are more similar to situations they would encounter in everyday life, where they are given with more robust social information about the speaker, including visual cues. While this does not give us the power to conclusively isolate variables and their social indices, it is a useful methodology if we seek to understand how listeners utilize social information to make linguistic judgments. In the current study, we present listeners with full phrases as well as descriptions of personae with accompanying biographical information in order to develop a new methodology for lect identification. We also explore to what extent listeners’ identities influence their ability to identify both lects of speakers similar to, and different from, themselves.

1.2 Production Differences in Prosody: What Are Listeners Hearing?

While the evidence that listeners generally perform well in tasks that ask them to discern regional, ethnic, and racial information is robust, these tasks assume the existence of production-level differences that are not always as well understood as the attending perceptual findings. Despite the research that has explored how listeners employ segmental phonological information in dialect classification, very few papers have examined how listeners may rely on prosodic information to do the same. A number of studies have tested judgements of specific vowel quality differences found in production, but few have looked at production/perception correspondences between non-segmental phonetic and phonological information. This section will discuss the studies that have attempted to investigate whether specific documented production differences between groups can be robustly detected in perception studies.

With respect to non-vocalic phonetic differences between AAE and MUSE (Mainstream US English), Thomas (2015) presents a number of variables that have been reported in production studies, but notes that many of these have not been tested for their perceptual salience. Among other features, researchers have tested effects of intonation and timing (Koutstaal & Jackson 1971) and f_0 (Hawkins 1992) without conclusive results. Walton & Orlikoff (1994) found that listeners could differentiate white and Black male speakers with approximately 60% accuracy on the basis of sustained /a/ vowels. In a post-hoc analysis, the authors found that jitter (F0 perturbation), shimmer (amplitude perturbation) and harmonics-to-noise ratio, differed between the clips of white and Black speakers. Additionally, listeners were more accurate at differentiating speakers with larger differences in these cues, though the authors did not claim that any of these features were necessarily individually perceptually salient. Foreman (2000) tested whether intonational differences could cue differences in listener judgments of a speaker's race, and found that even with speech that had been low-pass filtered at 900 Hz, Black and white American listeners did differentiate the voices based on intonational cues, a finding which was also supported by Koutstaal & Jackson's (1971) earlier work. Other studies that have focused on the production aspects that may trigger social perceptions have also frequently done so post-hoc. In a matched-guise listening study, Purnell, Idsardi & Baugh (1999) found that listeners could accurately differentiate between a white, Black, and Chicano guise and that these guises did contain differences with respect to voice quality features such as jitter, shimmer, and HNR (also supporting the work of Walton & Orlikoff 1994), but could not conclusively claim that the voice quality differences themselves triggered such judgments. Thomas, Lass, & Carpenter (2010) review this work and conclude that "listeners are capable of accessing a variety of cues under experimental conditions" (267). While this claim may seem self-evident, work that continues to try to better articulate the link between production and perceptual cues, especially in the domain of prosody, is vitally important for understanding mechanisms for accent and speaker identification.

The literature on production differences in prosody between regional varieties of American English is somewhat more robust, though many questions still remain about the perceptual salience of the production differences. Clopper & Smiljanic (2011), which also provides a portion of the stimuli used in the current study, finds differences in the use of pitch accents, boundary tones, and pauses as a function of speaker region and gender. Similarly, Arvaniti & Garding (2007) find both realizational and meaning differences in the use of H*/L+H* pitch accents between Minnesota and California speakers, though they do not test whether these differences are salient to listeners.

While there are a number of studies on perception of AAE and broader regional varieties, the perception of varieties such as Jewish English and Appalachian English is less well-examined. In particular, the perception of Jewish speakers is understudied, but as noted above, Burdin (2020) found that contours with a greater degree of macro-rhythm (more rising/falling patterns) were more likely to be perceived as Jewish than intonational contours with a lesser degree

of macro-rhythm. Meanwhile, perception studies of Appalachian speakers are practically non-existent. In one of the only studies to address these questions in Appalachian English, Cramer (2018), using perceptual dialectological methods, showed that Kentuckians recognized that Appalachian speakers are distinct from others. However, a distinct research lacuna exists as to whether, or how, listeners can perceive Appalachian varieties.

Burdin et al. (2018; 2022) found that the L+H* pitch accent differs in both rate of use and realization between varieties of American English, especially between African American English (AAE), Appalachian English (ApEng), and Jewish English (JE); likewise, there are differences in the production of H* contours as well. Each of these varieties also uses L+H* at a greater rate and with more extreme realizations than the Midlands (Midwestern) speakers studied in Clopper & Smiljanic (2011). In addition to these prosodic differences in L+H* and H*, there are likely other prosodic differences, including speech rate and rhythm, and other differences in intonational patterns present in the clips. Therefore, the evidence of prosodic production differences between varieties is robust, but such differences have rarely been studied in perception.

1.3 Method for Investigating the Role of Prosody in Lect Identification

In addition to whether listeners can perform accurately on lect identification tasks, and whether there is a direct relationship between production and perception differences, researchers have also tried to better understand listener judgments by controlling the stimuli and conducting a series of manipulations in order to focus on specific linguistic variables of interest. This type of design may be especially useful for suprasegmental features which are difficult to isolate, and which may have sociolinguistic indexes that have not yet been thoroughly explored in the research. For scholars interested in intonational variables, filtering of the speech signal has been a primary method for isolating potential effects of suprasegmental variables from those of segmental ones.

In an early perception study that employed a phonetic manipulation method, Lass et al. (1980) presented participants with recordings of ten African American and ten European American speakers in three conditions: unfiltered, low-pass filtered at 225 Hz, and high-pass filtered at 225 Hz, and found that listeners performed above chance at identification in all three conditions. Building on the work of Thomas & Reaser (2004), Thomas et al. (2010) conducted two experiments involving prosodic manipulation designed to uncover which cues Black and white listeners use in ethnicity judgments. In Experiment A, listeners heard read stimuli sentences in three conditions: unmodified, monotonized, and all vowels converted to schwa. Overall, they found that listeners performed with greatest accuracy on the unmodified stimuli and only slightly less well with the monotonized stimuli. However, listener accuracy dropped considerably in the schwa-vowels condition, indicating that vowel quality may play a larger role

than F0 information. In Experiment B, listeners heard stimuli where either the vowel quality or the F0 track had been swapped via resynthesis. That is, listeners heard original stimuli as well as Black speakers whose intonation and/or vowels had been resynthesized to match white speakers, and vice versa. In all cases, listeners performed with over 80% accuracy, indicating that both prosody and vowel quality can act as cues for listener racial identification. Using a similar methodology, Leeman et al. (2018) tested the contribution of segmental versus prosodic information in varieties of Swiss German. Listeners were presented with two varieties of Swiss German (Bern and Valaise) that differed on both segmental and prosodic information. Each variety was presented unmodified, with the rhythm of the other, e.g., Bern with Valaise rhythm, and with both the rhythm and intonation of the other, e.g., Bern with Valaise rhythm and intonation. Listeners were highly accurate in identification, even with the manipulations. However, accuracy did drop slightly for some listeners, showing that prosodic information does play a role in lect identification. Manipulation studies such as these serve to provide greater detail about the ways in which different listeners may use different cues in their judgments, as well as elucidate the role of prosodic cues in particular. Additionally, some studies have manipulated the f0 contour to test the salience of variation in the production of pitch accents in lect identification, including Thomas et al. (2010) and Holliday & Villarreal (2020).

In sum, previous work has found that:

- (1) Listeners bring social knowledge to lect perception tasks, including accessing particular types of personae even if (and maybe especially when) information about the speaker is not given to them; and related to that,
- (2) Performance on lect perception tasks may vary depending on listener characteristics; and
- (3) Based on previous work, both on the production and perception side, prosodic features likely play a role in lect identification, particularly for identifying African American English, Jewish English, and Appalachian English.

Given the findings discussed above, our research questions are as follows:

1. When presented with stimuli from five different lects, can listeners match personae with those lects?
2. Does this differ depending on whether segmental cues are absent or present, potentially encouraging participants to rely more on prosodic differences?
3. Does performance on this perception task vary by listener demographic?
4. How do these results bear on how listeners may engage in lect identification in everyday speech contexts?

2 Method

2.1 Stimuli

For the three lects of interest (AAE, ApEng, JE), three clips from a female speaker found to be representative of each lect were selected. These speakers reflected the general acoustic and vocalic productions typical of the areas studied, based on earlier experimental work (Burdin et al. 2022). For example, the Appalachian speaker exhibited aspects of the Southern Vowel Shift and monophthongization of the diphthong /ay/. However, we want to be mindful of the pitfalls of dialect essentialism, wherein the only “authentic” speakers of a variety are those that exhibit all the “expected” or documented features. The stimuli for the current study were those that were analyzed in a previous paper (Burdin et al. 2022), as well as those analyzed in Clopper & Smiljanic (2011). In each of the previous studies, the authors conducted a reading task with speakers from each group. African American speakers read the “Rainbow Passage” (Fairbanks 1960). Appalachian speakers read “Arthur the Rat” (Hall 1942) and Jewish speakers read “Comma Gets a Cure” (McCullough & Somerville 2000). For the Black, Jewish, and Appalachian speakers, the authors shared a racial, regional, or ethnic community with the speakers they collected this production data from. The data also consists of 2 Midwestern (specifically, the Midlands dialect, both white, and presumed non-Jewish) speakers from Clopper & Smiljanic (2011), reading the passage “Goldilocks”.² All speakers used for the stimuli creation were female and identified as speakers of each of the lects of interest. While the passages were not the same, the methodology for collecting the passages was quite similar across the works (Burdin 2016; Holliday 2016; Reed 2016). For each, passages were read as part of larger interview settings, using high quality recorders and lavalier microphones in quiet environments. Further, the listening participants only heard 1–3 word utterances from each and were told they were to hear the same reading passage from several speakers.

As described above, earlier analysis of the data found that both the H* and L + H* of JE, AAE, and ApE differ from each other, and for L + H* specifically, from Southern and Midland English. **Table 1** gives the values for the three varieties of interest for H*, in Hz and ERB. While these differences have been found to be significant in earlier work, it is unclear to what extent listeners may be able to use these prosodic cues to accurately identify the lects in question, especially in the absence of segmental material, and given only a limited number of examples from one speaker each.

Clips from each speaker were selected on the basis that they met the prosodic criteria of interest, so each clip consisted only one Intonational Phrase (IP) with a duration of <2 seconds, containing the pattern H* L + H* L-L% as labelled in an earlier study (Burdin et al. 2022), using the

² Since the Midwestern speakers were taken from Clopper and Smiljanic (2011)’s data, we rely on their criteria for Midland speakers. In this case, talkers were from central Indiana or Missouri and were selected to represent a variety approximating “General American”.

MAE-ToBI conventions (Beckman & Elam 1997); see this previous work for further discussion of interrater reliability, which generally found the annotators were consistent in labeling pitch accents in these varieties despite these phonetic differences. Finally, five clips total were taken from two Midwestern speakers. **Figure 1** shows a spectrogram of an example sentence from a JE speaker.³

	AAE	JE	ApE
Peak height	192 Hz (5.66 ERB)	200 Hz (5.83 ERB)	202 Hz (5.88)
Slope	1.135 Hz/ms (0.045 ERB/ms)	2.56 Hz/ms (0.10 ERB/ms)	2.97 Hz/ms (0.11 ERB/ms)
Rise span	17 Hz (0.66 ERB)	32 Hz (1.21 ERB)	23 Hz (0.89 ERB)
Peak alignment	6.01 ms	4.95 ms	6.01 ms

Table 1: Values of H* for the three varieties, from Burdin et al. 2022.

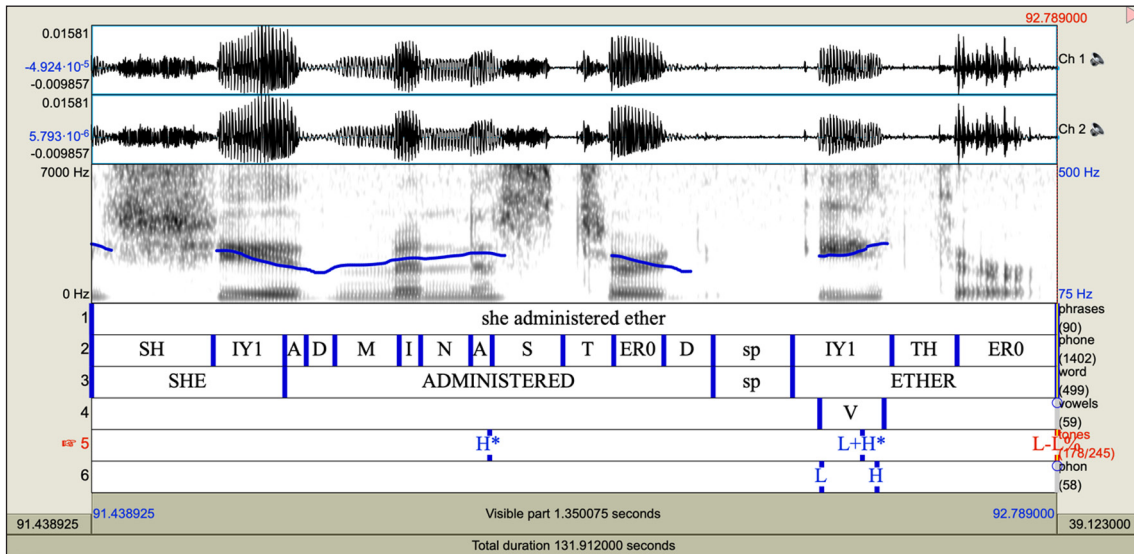


Figure 1: Example of phrase with H* L + H* L-L% pattern, labelled for phrase, phoneme, word, vowel interval length, MAE-ToBI labels, and low and high F0 points within the L + H* pitch accent interval.

³ Originally these clips were intended to be manipulated to five different guises, with the f0 of the L and H of the L + H* matching the approximate average L and H of the L + H* from each of the five lects (Southern, Appalachian English, Midlands, Jewish English, and African American English). To keep the study a manageable length, we presented original speakers only from the three varieties of interest (AAE, JE, and ApE), plus the Midlands to represent a “General American” guise. This meant that Southern, while a “correct” answer from a prosodic point of view, was never a correct answer from an original speaker point of view. However, the JE and ApE guises were inadvertently manipulated to the same level, meaning that there were only four guises. After finding that these new manipulations were not significant (and it was unclear how to interpret the results) we removed them from the statistical model. As such, our interest shifted to the differences in performance in low pass filtered vs. unfiltered speech.

Listeners heard a total of 70 clips in each block. For the three lects of interest (AAE, ApEng, JE) listeners heard 15 clips each (3 clips \times 5 guises). For the Midwestern lect, there were 25 clips. **Figure 2** demonstrates the organization of the clips for each block.

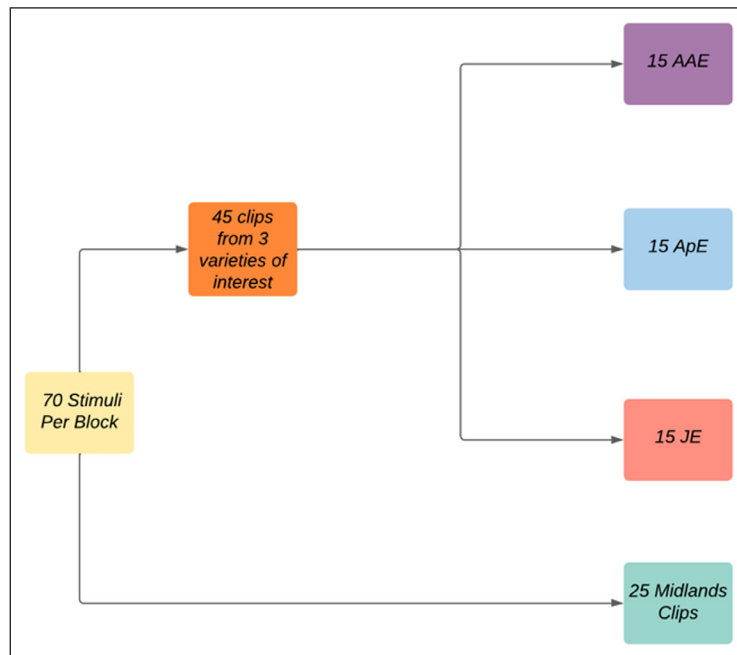


Figure 2: Visualization of stimuli presented in each block (2 blocks total).

To test the contributions of the prosodic differences in the presence and absence of segmental information, participants heard two blocks of stimuli. In the first experimental block, listeners were presented with a series of clips that were low-pass filtered at 400 Hz (following Knoll et al. 2009) to remove segmental phonological information. In the second block, listeners heard the original, unfiltered versions of the same clips.

2.2 Participants

In order to test the role of listener background, four different listener groups were recruited: Jewish listeners, African American listeners, Appalachian listeners, and Non-target listeners who did not belong to any of the previous three groups. Listeners participated in an online survey created with Qualtrics (www.qualtrics.com) and conducted via the research platform Prolific (www.prolific.co).

76 listeners from four different listener groups participated in the study. Listeners were recruited on the basis of their identification with the communities of interest, using Prolific's recruitment filters. African-American listeners ($n = 19$) self-identified as Black/African-American, and Jewish listeners ($n = 20$) identified as Jewish in their Prolific profiles. For the Non-target

($n = 17$) and Appalachian ($n = 20$) listeners, we utilized geographically bounded criteria filtered by state. Appalachian listeners were those who indicated residence in any of the states contained in the Central and Southern areas as defined by the Appalachian Regional Commission. (www.arc.gov). Since Prolific only allows filtering down to the state level, we then further filtered the participants to only include listeners who indicated they were from towns or cities within the ARC parameters. The Non-target listeners were recruited from states that were not listed as containing parts of Appalachia, and were from across the United States. None of the listeners had overlapping regional/racial/religious characteristics that would cause them to be representative of more than one group. This recruitment method allowed us to obtain a diverse sample of listeners that demographically paralleled the speakers in the stimuli, as well as to control the method in which the data was presented and the time frame during which it was collected.

2.3 Listening Task

Upon consenting to the study, the listeners were informed that they would hear clips of five women reading the same story⁴ and that they would be tasked with matching each clip to one of the five speakers. They were also instructed that in the first task, the voices may sound “quiet and perhaps a bit odd (almost like voices from another room or voices outside a door)”. They were instructed to do the survey in a quiet room and to use headphones. **Figure 1** shows the photos and biographies given for each speaker, which were presented to participants at the onset of the survey. For each clip, participants selected the name of the persona that they believed matched the voice, and they were not given a time limit. The photos were taken from free stock photo sites (Unsplash.com and freepik.com) and the invented biographies were composites of the speakers in each of the data sets. Names were chosen from various lists of popular women’s names. Prior to running the experiment, we conducted two norming tasks asking naïve listeners if the names, descriptions, and photos seemed congruent in order to confirm that the associations between photos, names, and descriptions were clear and also that the names fit the descriptions and photos. The initial norming task volunteers found overall that 4 of the 5 photos were congruent with the names and biographies provided. The incongruent photo was replaced with a different photo, and a subsequent norming task found all 5 photos, names, and descriptions to be congruent. **Figure 3** shows the five personae, with the photos and biographical information that were presented to the listeners.⁵

Following the listening task, listeners were asked a series of demographic questions about their age, race/ethnicity, religion, hometown, other places lived, current place of residence, and experience studying languages other than English. These questions were designed to test for

⁴ The fact that the stories differed by speaker was not evident to listeners due to short duration of the clips.

⁵ A reviewer pointed out that listener groups might differ in how they interpret the different facial expressions of each photo. We reiterate that our norming task did not show any incongruities. However, we recognize that such a difference might exist, and we hope to address this in future iterations of this work.

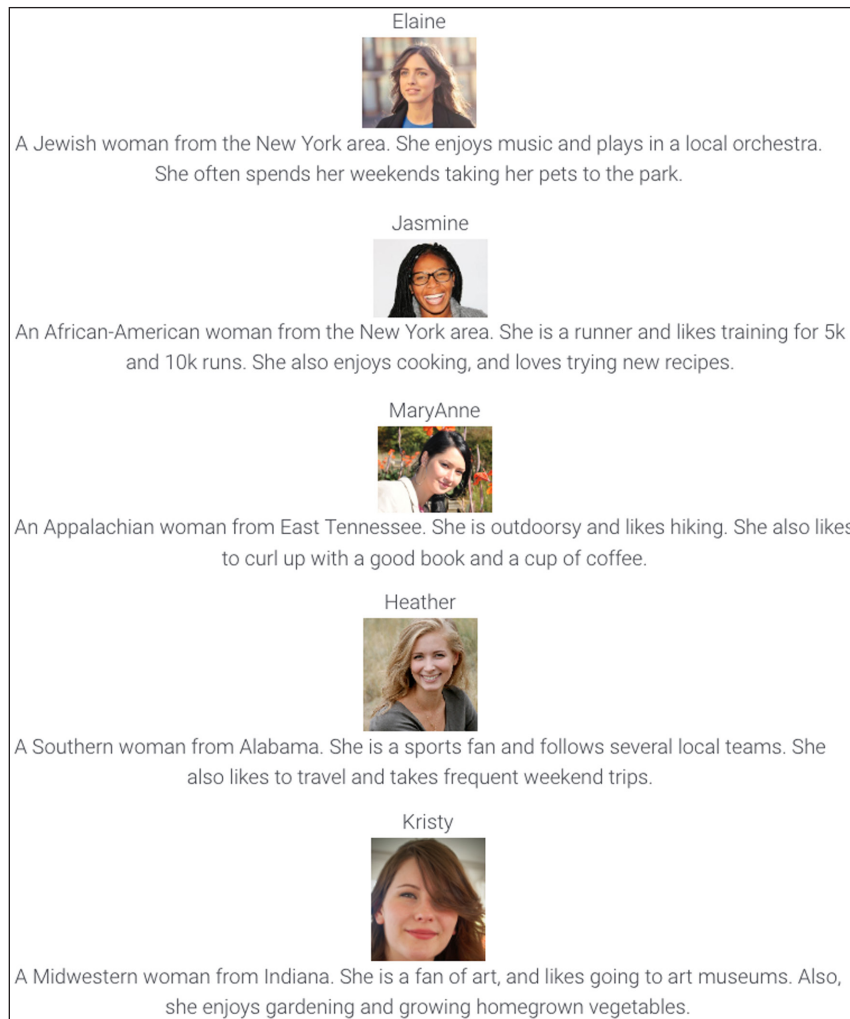


Figure 3: Personae presented to listeners at study onset.

possible effects of in-group bias and/or listener familiarity with the different lects, since earlier studies, including Clopper & Pisoni (2004a) and Thomas et al. (2010) found differences as a result of listener experience with their lects of interest.

3 Analysis and Results by Listener Group

In order to analyze the results of the perception task, we ran a log-linear model for contingency tables in R (R Core Team 2022) evaluating the performance of each listener group (Jewish English, African American, Appalachian, and Non-target) in order to understand what differences in listener judgments may exist among the groups. Rather than predicting the likelihood of a specific response given a specific stimulus, these types of models predict the frequency of

responses in a given cell, and thus are useful for analyzing results of this task. If listeners are completely accurate in their judgements, the results would look like **Table 2**, with listeners selecting the correct response 100% of the time. **Table 3** shows an example of the listeners' performing at chance in identifying the lect. We can compare the diagonal of the matrix of the actual results to chance (i.e., the values in the bolded cells in **Table 2**) to see if listeners are identifying the lects at rates above chance.

	Original Lect			
Response	ApEng	JE	AAE	Midwestern
ApEng	100	0	0	0
JE	0	100	0	0
AAE	0	0	100	0
Midwestern	0	0	0	100
Southern	0	0	0	0

Table 2: Confusion matrix, perfect performance.

	Original Lect			
Response	ApEng	JE	AAE	Midwestern
ApEng	20	20	20	20
JE	20	20	20	20
AAE	20	20	20	20
Midwestern	20	20	20	20
Southern	20	20	20	20

Table 3: Confusion matrix, performance at chance.

These tables show response and actual lect; the final model also included listener group as an additional dimension to the tables. The full table of results by listener group and original lect, with an additional break down by filtering condition, is shown in **Table 4**. The full data set, as well as code for the following statistical models are available at <https://osf.io/v49bk/>.

Listener Group	Response	Original											
		African American		Appalachian		Jewish		Midwestern					
		Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered	Filtered				
African American	African American	130 (49%)	34 (13%)	48 (16%)	78 (27%)	19 (7%)	122 (43%)	54 (11%)	78 (16%)				
African American	Appalachian	33 (12%)	61 (23%)	70 (25%)	57 (20%)	43 (15%)	21 (7%)	80 (17%)	83 (17%)				
African American	Jewish	23 (9%)	61 (23%)	45 (16%)	48 (17%)	144 (51%)	54 (19%)	71 (15%)	97 (20%)				
African American	Midwestern	69 (26%)	60 (23%)	40 (14%)	56 (20%)	41 (14%)	54 (19%)	155 (33%)	126 (27%)				
African American	Southern	11 (4%)	50 (19%)	82 (29%)	46 (16%)	38 (13%)	34 (12%)	115 (24%)	91 (19%)				
Appalachian	African American	95 (34%)	48 (17%)	23 (8%)	151 (50%)	117 (39%)	184 (61%)	27 (5%)	73 (15%)				
Appalachian	Appalachian	33 (12%)	47 (17%)	90 (30%)	37 (12%)	3 (1%)	25 (8%)	113 (23%)	85 (17%)				
Appalachian	Jewish	63 (23%)	51 (18%)	33 (11%)	38 (13%)	131 (44%)	42 (14%)	79 (16%)	87 (17%)				
Appalachian	Midwestern	75 (27%)	74 (26%)	27 (9%)	57 (19%)	40 (13%)	29 (10%)	195 (39%)	145 (29%)				
Appalachian	Southern	14 (5%)	60 (21%)	127 (42%)	17 (6%)	9 (3%)	20 (7%)	86 (17%)	110 (22%)				
Jewish	African American	80 (29%)	30 (11%)	41 (14%)	116 (39%)	80 (26%)	135 (45%)	21 (4%)	59 (12%)				
Jewish	Appalachian	59 (21%)	71 (25%)	86 (29%)	59 (20%)	28 (9%)	40 (13%)	81 (16%)	88 (18%)				
Jewish	Jewish	59 (21%)	56 (20%)	16 (5%)	37 (12%)	164 (55%)	41 (14%)	84 (17%)	92 (18%)				

(Contd.)

		Original											
		African American			Appalachian			Jewish		Midwestern			
Listener Group	Response	Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered	Filtered	Unfiltered	Filtered		
Jewish	Midwestern	67 (24%)	66 (24%)	29 (10%)	66 (22%)	27 (9%)	51 (17%)	247 (49%)	119 (24%)				
Jewish	Southern	15 (5%)	57 (20%)	128 (43%)	22 (7%)	1 (0%)	33 (11%)	67 (13%)	142 (28%)				
Non-target	African American	93 (39%)	26 (11%)	36 (14%)	102 (40%)	100 (39%)	111 (44%)	27 (6%)	54 (13%)				
Non-target	Appalachian	47 (20%)	33 (14%)	41 (16%)	39 (15%)	35 (14%)	29 (11%)	96 (23%)	76 (18%)				
Non-target	Jewish	40 (17%)	58 (24%)	27 (11%)	44 (17%)	87 (34%)	42 (16%)	96 (23%)	86 (20%)				
Non-target	Midwestern	50 (21%)	48 (20%)	38 (15%)	44 (17%)	28 (11%)	53 (21%)	134 (32%)	110 (26%)				
Non-target	Southern	8 (3%)	73 (31%)	113 (44%)	26 (10%)	5 (2%)	20 (8%)	72 (17%)	99 (23%)				

Table 4: Full table of response frequency by listener group, original stimuli, and filtering condition.

As a reminder of our research questions, based on previous work, we expected that listeners may be influenced in their choice of persona based on prosodic differences between the lects, with this being mediated by condition (low-pass filtered or not). Additionally, we expected differences by listener group, with listeners being more accurate with stimuli from their in-group than their out group. For the model, then, the dependent variable, again, is the *frequency* of the response choice (e.g., the number in one of the cells above). The independent variables are the lect of the original stimulus, the response (that is, an individual's response, rather than the frequency count of a cell), filtering condition, listener group, the diagonals of response and original as well as response, original, and group. Additionally, we included interactions between response and filtering condition; then group with original stimulus lect, response, filtering condition, and the diagonal of the response and the original; and finally a three way interaction between response, filtering condition, and group. The reference level for the response factor was Midwestern; for stimuli, Midwestern; for group, Non-target listeners; for filtering condition, the unfiltered condition.

Model fit and significance of variables were assessed using a chi-square test; the analysis of deviance table is shown below in **Table 5**. There were significant effects of the diagonal of response, original, and listener group ($p < 0.001$); diagonal of response and original ($p < 0.001$); an interaction between the diagonal of response and original with group ($p < 0.001$); as well as significant effects of original lect ($p < 0.001$), group ($p < 0.001$), an interaction of response and filtered ($p < 0.001$), and response, filtered, and group ($p < 0.001$). The diagonal terms show that, when looking at the diagonals of the contingency tables above (representing accurate matches of original lect to response), we see significant differences by group, the original lect, and filtering conditions. The full regression table is shown in appendix A, with the reference levels as described above. Due to the complexities of examining all of the variables at once, as well as, e.g., testing all pairwise comparisons for significance, in the following sections, we walk through the results for each group by filtering condition to get a general idea of the patterns driving these results.

	Degrees of Freedom	Deviance	Residual Degrees of Freedom	Residual Deviance	
NULL			159	3913.5	
Diag(response, original, group)	4	125.31	155	3788.2	$p < 0.001$
original	3	638.91	152	3149.3	$p < 0.001$
response	4	165.55	148	2983.7	$p = 1$
filtered	1	0.00	147	2983.7	1.0

(Contd.)

	Degrees of Freedom	Deviance	Residual Degrees of Freedom	Residual Deviance	
Diag(response, original)	4	389.26	143	2594.5	p < 0.001
Group	3	28.41	140	2566.1	p < 0.001
response*filtered	4	104.22	136	2461.9	p < 0.001
original*group	9	4.80	127	2457.1	0.8511
response*group	12	51.87	115	2405.2	p < 0.001
filtered*group	3	0.25	112	2404.9	0.9691
Diag(response, original)*group	8	40.31	104	2364.6	p < 0.001
response*filtered*group	12	42.64	92	2322.0	p < 0.001

Table 5: Analysis of deviance table for final log-linear model.

3.1 African American English Listeners

3.1.1. African American English Listeners-Unfiltered Condition

Table 6 shows the raw counts for African American listeners in the unfiltered condition.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	130 (49%)	48 (16%)	19 (7%)	54 (11%)
Appalachian	33 (12%)	70 (25%)	43 (15%)	80 (17%)
Jewish	23 (9%)	45 (16%)	144 (51%)	71 (15%)
Midwestern	69 (26%)	40 (14%)	41 (14%)	155 (33%)
Southern	11 (4%)	82 (29%)	38 (13%)	115 (24%)

Table 6: Results for Identification Task for African American Listeners (unfiltered).

To better visualize these results, we created stacked bar graphs showing the responses for the African American listeners by lect. **Figure 4** shows how the African American listeners responded to the clips for each lect. Bolded boxes indicate the original lect that the listeners heard (accurate responses).

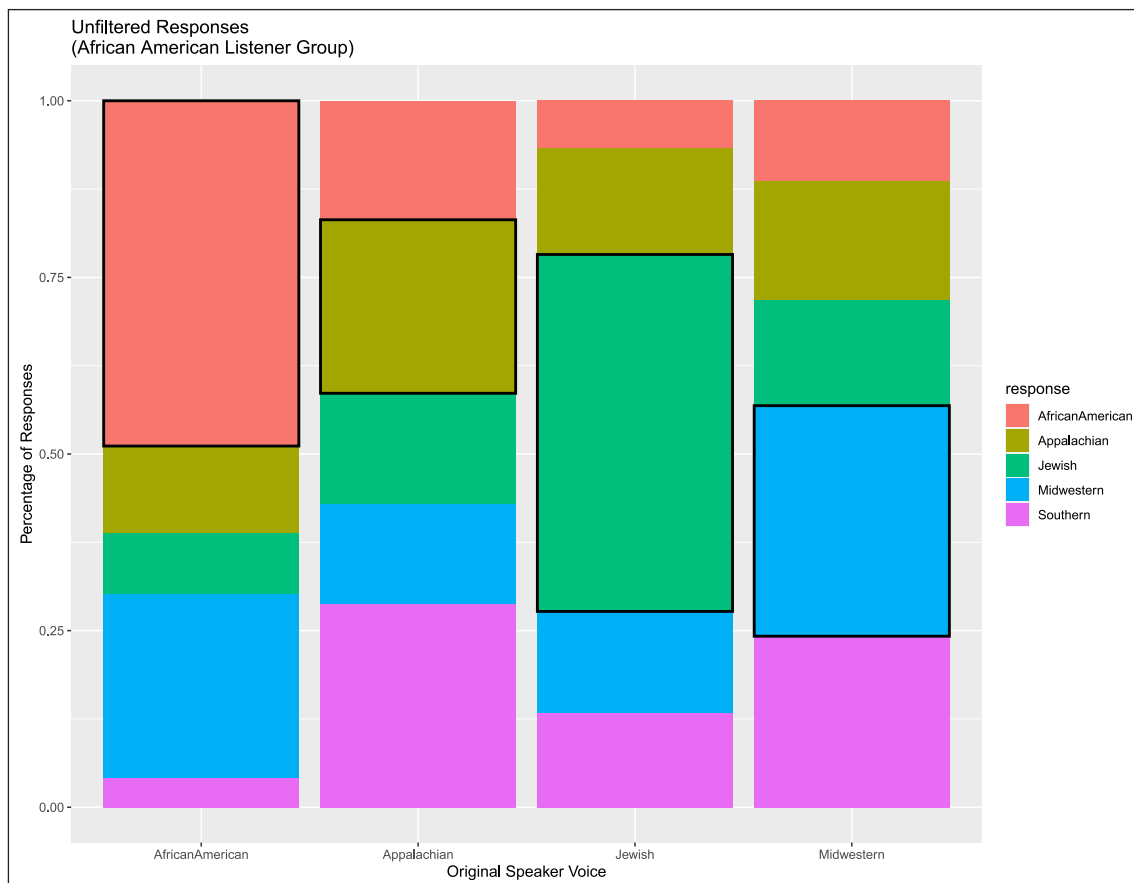


Figure 4: Percentage of African American listener responses by original lect (unfiltered); highlighted box indicates correct identification.

This figure demonstrates that 49% of the time, African American listeners were able to correctly identify the original AAE clip; 50% of the time, the Jewish English clip was accurately identified, and 32% of the time, the Midwestern clip was accurately identified. Appalachian English was misidentified as Southern (29%) more frequently than it was correctly identified (25%).

3.1.2. African American English Listeners-Filtered Condition

Differences emerge in the low-pass filter condition, with results presented in **Table 7** and **Figure 5**.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	34 (13%)	78 (27%)	122 (43%)	78 (16%)
Appalachian	61 (23%)	57 (20%)	21 (7%)	83 (17%)
Jewish	61 (23%)	48 (17%)	54 (19%)	97 (20%)
Midwestern	60 (23%)	56 (20%)	54 (19%)	126 (27%)
Southern	50 (19%)	46 (16%)	34 (12%)	91 (19%)

Table 7: Crosstab Results for Identification Task for African American listeners (filtered).

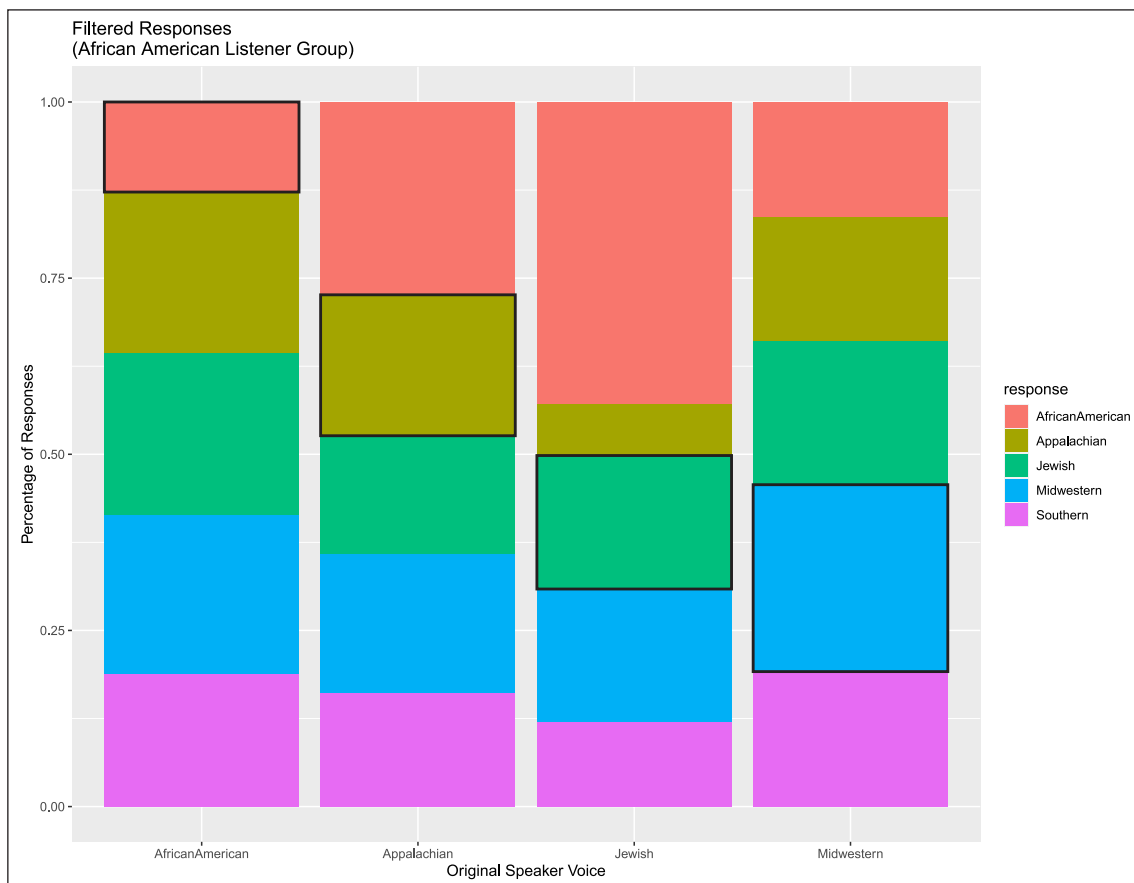


Figure 5: Percentage of African American listener responses for each lect (filtered); highlighted box indicates correct identification.

Again, the African American listeners responded primarily to the original, actual dialect of the stimuli, even if those responses were not accurate: only the Midwestern (26%) and Appalachian English (20%) were identified at or above chance, assuming chance is at 20%, given the five personae choices. As a reminder, for here and following, it is unknown if this was *significantly* above chance for this particular combination of listener group, filtering condition, and original stimulus lect; see note above about pairwise comparisons and significance testing,

Jewish English was misidentified as AAE (43%), as was Appalachian English (27%). AAE was fairly evenly misidentified as Appalachian English, Jewish English, and Midwestern (all at 23%).

3.2.1 Appalachian Listeners-Unfiltered Condition

In the unfiltered condition, Appalachian listeners again were influenced by the actual lect of the original stimulus. **Table 8** shows the cross tabs for Appalachian listeners, and **Figure 6** shows the distribution of responses for the clips.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	95 (34%)	23 (8%)	117 (39%)	27 (5%)
Appalachian	33 (12%)	90 (30%)	3 (1%)	113 (23%)
Jewish	63 (23%)	33 (11%)	131 (44%)	79 (16%)
Midwestern	75 (27%)	27 (9%)	40 (13%)	195 (39%)
Southern	14 (5%)	127 (42%)	9 (3%)	86 (17%)

Table 8: Results for Identification Task for Appalachian English listeners (unfiltered).

Appalachian listeners were 30% accurate in identifying their own lect, and also identified three other lects above chance (AAE = 34%, JE = 44%, and Midwestern, 39%). When listeners were not accurate, ApEng was most frequently confused for Southern with listeners labelling the ApEng clip as Southern 42% of the time. JE and AAE were confused, with JE labeled as AAE 39% of the time, and AAE as JE 23% of the time; AAE was also confused for Midwestern (27%).

3.2.2 Appalachian Listeners-Filtered Condition

In the filtered condition, we observe a similar result for Appalachian listeners as for the African American listeners between unfiltered and filtered conditions, as shown in the results in **Table 9** and **Figure 7**.

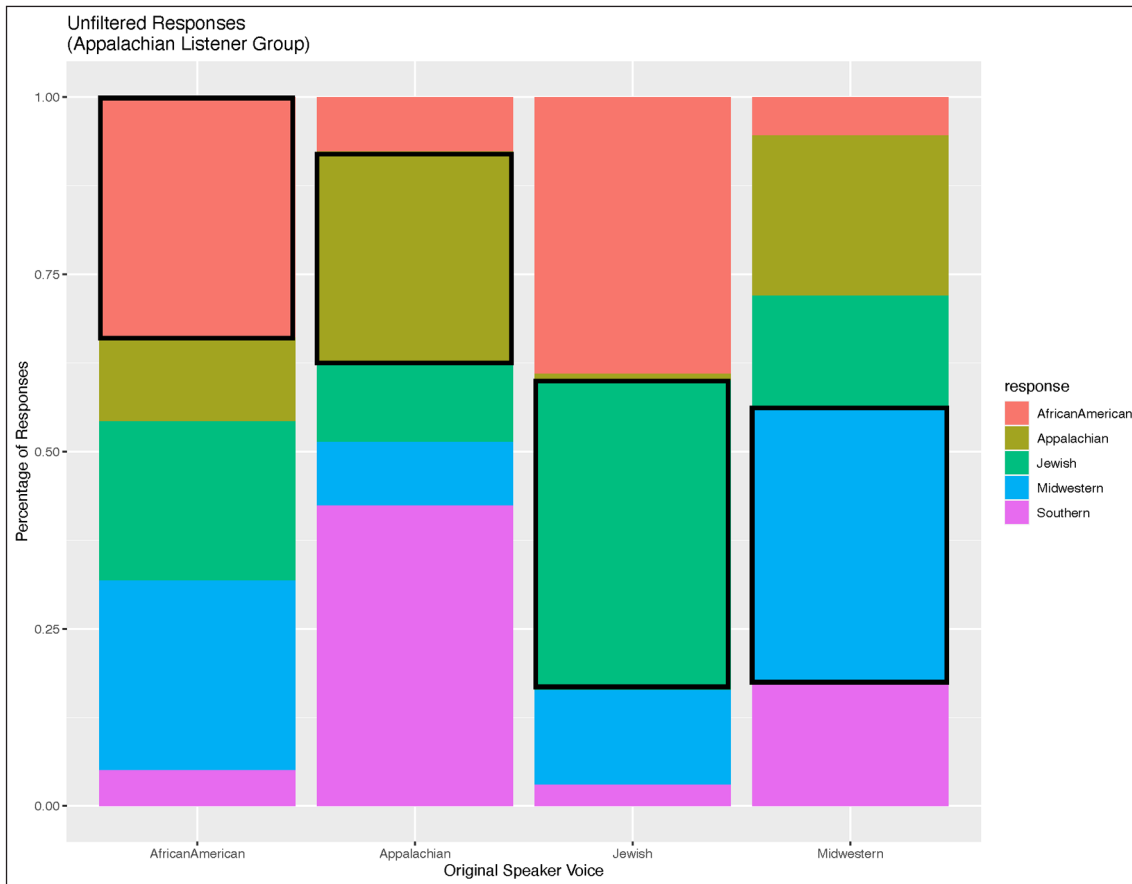


Figure 6: Percentage of Appalachian listener responses for each lect (unfiltered); highlighted box indicates correct identification.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	48 (17%)	151 (50%)	184 (61%)	73 (15%)
Appalachian	47 (17%)	37 (12%)	25 (8%)	85 (17%)
Jewish	51 (18%)	38 (13%)	42 (14%)	87 (17%)
Midwestern	74 (26%)	57 (19%)	29 (10%)	145 (29%)
Southern	60 (21%)	17 (6%)	20 (7%)	110 (22%)

Table 9: Results for Identification Task for Appalachian English listeners (filtered).

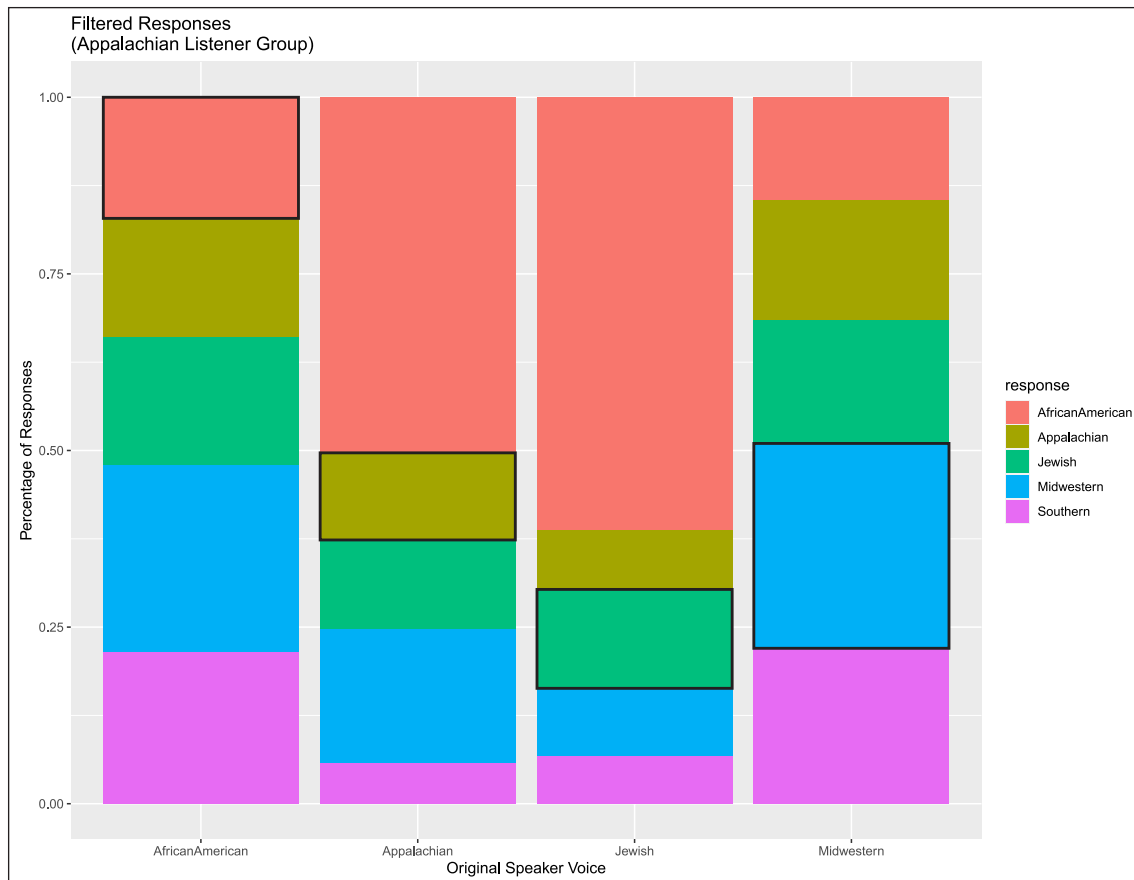


Figure 7: Percentage of Appalachian listener responses for each lect (filtered); highlighted box indicates correct identification.

Even in the presence of low-pass filtering, Appalachian listeners were influenced by the actual lect of the original stimulus. However, the filtering condition did influence which lects were confused for one another, with Appalachian listeners now somewhat more likely to classify both their own lect as AAE (50%) and JE as AAE (61%). AAE was also confused for Midwestern (26%). Midwestern was the only dialect identified correctly at a level above chance, 29% of the time.

3.3.1 Jewish Listeners-Unfiltered Condition

The Jewish listener group performed similarly to the other two groups. The results for the unfiltered condition are shown in **Table 10** and **Figure 8**. Jewish listeners were most accurate on identifying the Midwestern (49%) and Jewish (54%) stimuli, and also were above chance for AAE (29%) and ApEng (29%). However, they frequently perceived Appalachian English as Southern English (42%), and confused African American English as Midwestern (23%), and JE as AAE (26%).

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	80 (29%)	41 (14%)	80 (26%)	21 (4%)
Appalachian	59 (21%)	86 (29%)	28 (9%)	81 (16%)
Jewish	59 (21%)	16 (5%)	164 (55%)	84 (17%)
Midwestern	67 (24%)	29 (10%)	27 (9%)	247 (49%)
Southern	15 (5%)	128 (43%)	1 (0%)	67 (13%)

Table 10: Results for Identification Task for Jewish English Listeners (unfiltered).

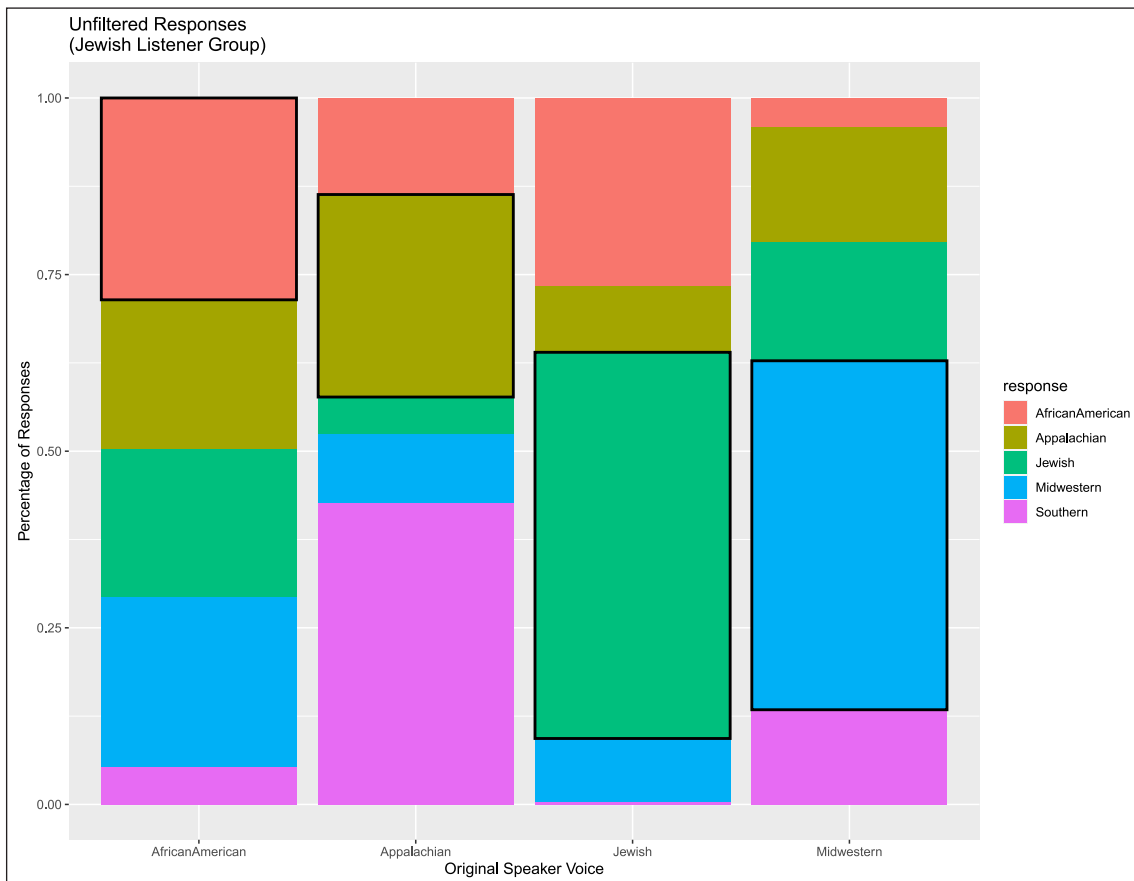


Figure 8: Percentage of Jewish listener responses for each lect (unfiltered).

3.3.2 Jewish Listeners-Filtered Condition

For the filtered condition, the results for the Jewish listener group are shown in **Table 11** and **Figure 9**.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	30 (11%)	116 (39%)	135 (45%)	59 (12%)
Appalachian	71 (25%)	59 (20%)	40 (13%)	88 (18%)
Jewish	56 (20%)	37 (12%)	41 (14%)	92 (18%)
Midwestern	66 (24%)	66 (22%)	51 (17%)	119 (24%)
Southern	57 (20%)	22 (7%)	33 (11%)	142 (28%)

Table 11: Results for Identification Task for Jewish Listeners (filtered).

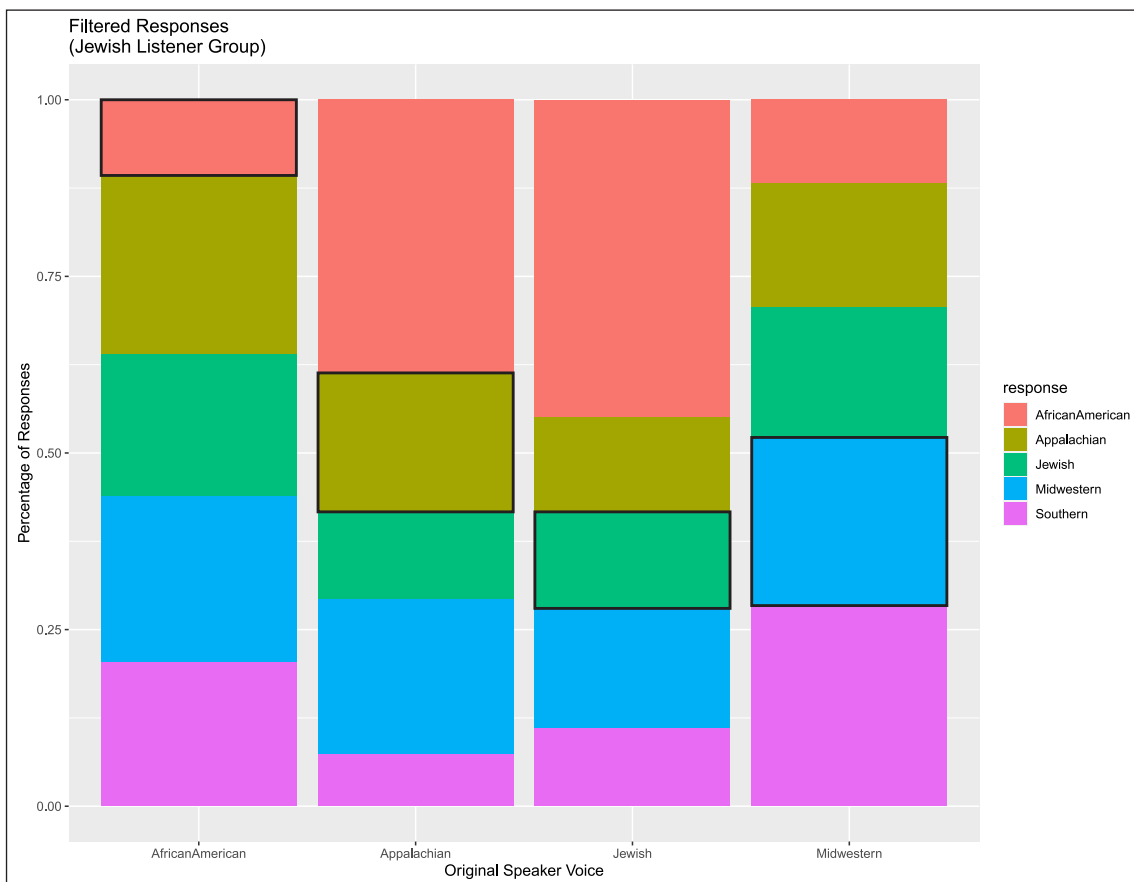


Figure 9: Percentage of Jewish listener responses for each lect (filtered); highlighted box indicates correct identification.

Again, similar to the listener groups above, even when the stimulus is low-pass filtered, the Jewish listeners were influenced by the original lect. Midwestern was the only lect identified above chance (24%). However, the filtering condition did introduce a greater degree of lect confusion. Jewish listeners were now more likely to classify both their own lect (45%) and ApEng (389%) as AAE. Additionally, in the filtered condition, Jewish listeners were also more likely to classify AAE as JE (20%), ApEng (25%), Midwestern (23%) and Southern (21%); and Midwestern as Southern (28%).

3.4.1 Non-target Listeners-Unfiltered Condition

The Non-target listener group had similar results to the previous three groups. The results for the unfiltered condition are shown in **Table 12** and **Figure 10**.

	Original			
Response	African American	Appalachian	Jewish	Midwestern
African American	93 (39%)	36 (14%)	100 (39%)	27 (6%)
Appalachian	47 (20%)	41 (16%)	35 (14%)	96 (23%)
Jewish	40 (17%)	27 (11%)	87 (34%)	96 (23%)
Midwestern	50 (21%)	38 (15%)	28 (11%)	134 (32%)
Southern	8 (3%)	113 (44%)	5 (2%)	72 (17%)

Table 12: Results for Identification Task for Non-target Listeners (unfiltered).

The Non-target listeners were influenced by the original dialect of the stimuli, in a manner similar to the other groups. Overall, Non-target listeners were most adept at identifying AAE (39%), as well as Midwestern (31%) and JE (34%); however, they were below chance for the ApEng (16%). Like the other groups, they most frequently confused Southern for ApEng (44%) and JE for AAE (39%).

3.4.2 Non-target Listeners-Filtered Condition

The results for the Non-target listeners in the filtered condition are shown in **Table 13** and **Figure 11**. As with the previous groups, even when low-pass filtered, the Non-target listeners were influenced by the dialect of the original stimuli.

Only the Midwestern speaker was identified above chance (29%). Interestingly, the filtering condition again influenced which lects were most frequently confused for one another, with Non-target listeners now more likely to label JE (44%) and ApEng (40%) as AAE, and AAE as Southern (31%).

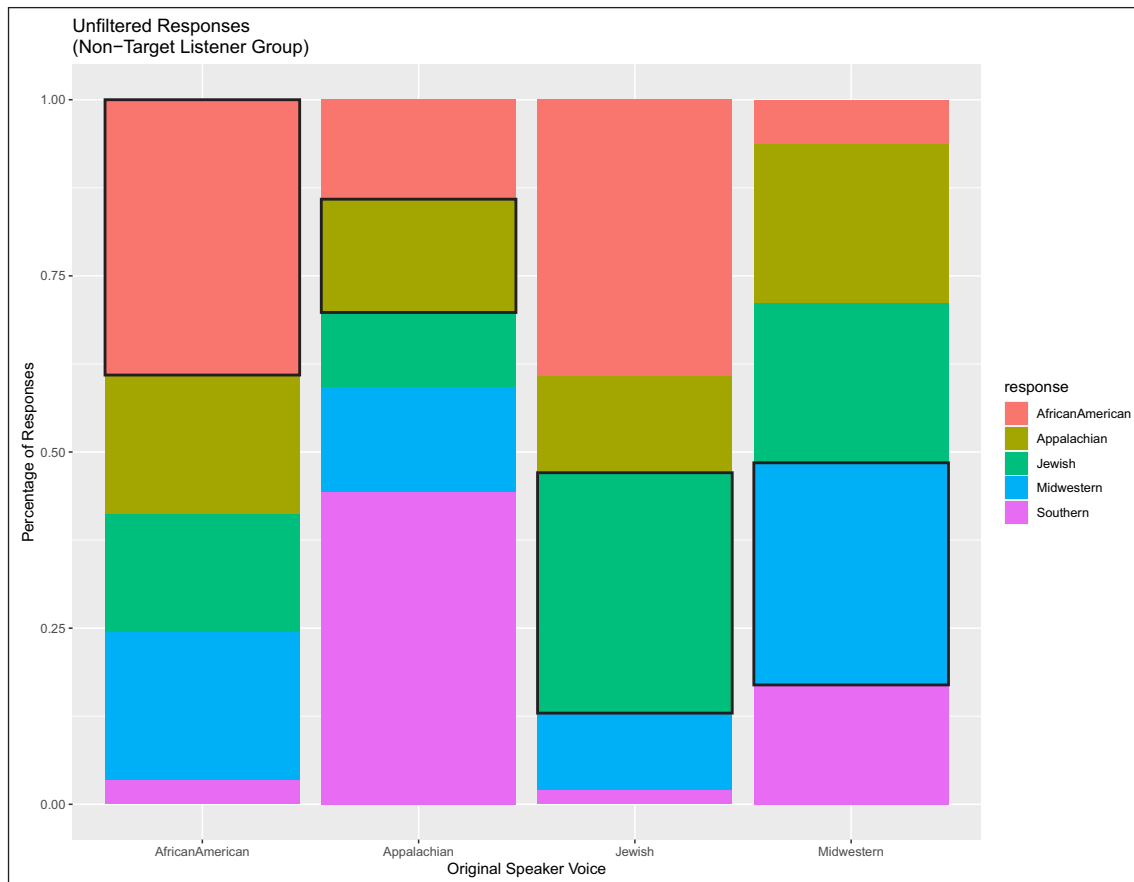


Figure 10: Percentage of non-target listener responses for each lect (unfiltered); highlighted box indicates correct identification.

	Original				
Response	African American	Appalachian	Jewish	Midwestern	Southern
African American	26 (11%)	102 (40%)	111 (44%)	54 (13%)	26 (11%)
Appalachian	33 (14%)	39 (15%)	29 (11%)	76 (18%)	33 (14%)
Jewish	58 (24%)	44 (17%)	42 (16%)	86 (20%)	58 (24%)
Midwestern	48 (20%)	44 (17%)	53 (21%)	110 (26%)	48 (20%)
Southern	73 (31%)	26 (10%)	20 (8%)	99 (23%)	73 (31%)

Table 13: Results for Identification Task for Non-target Listeners (filtered).

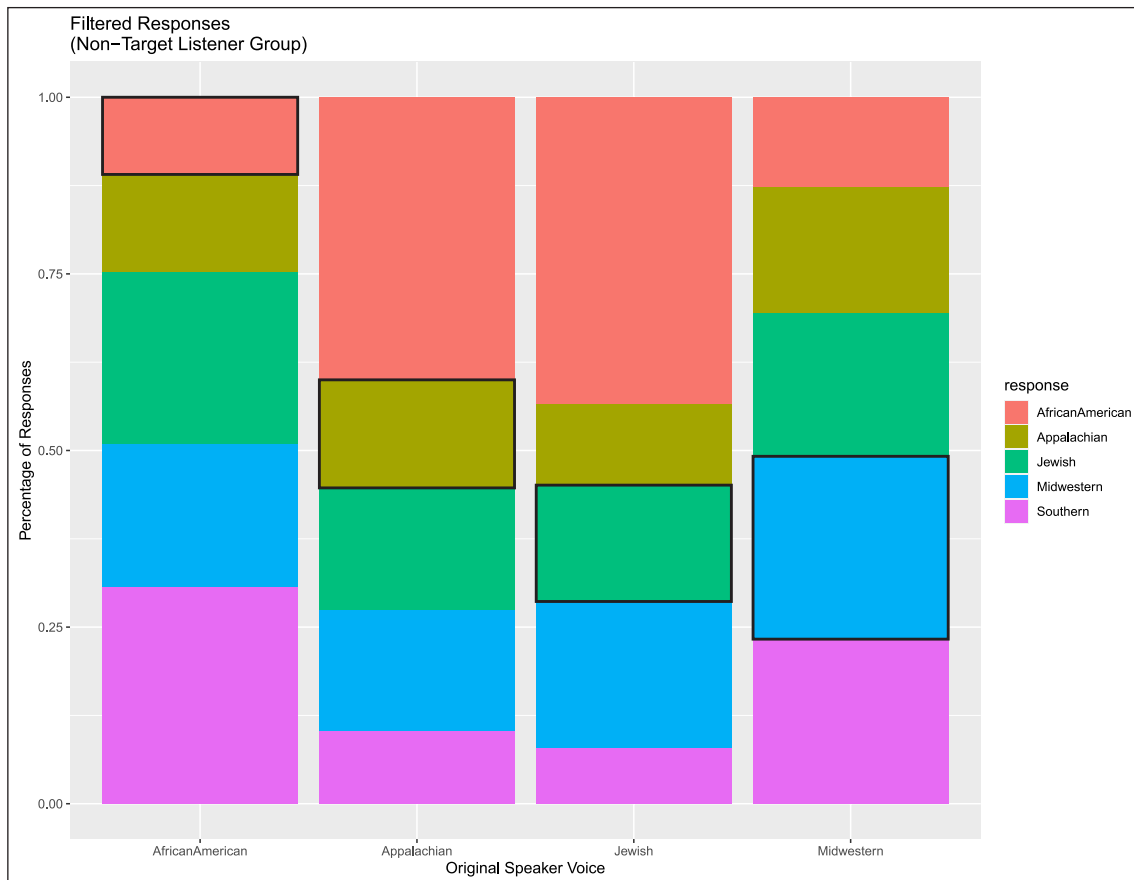


Figure 11: Percentage of Non-target listener responses for each lect (filtered); highlighted box indicates correct identification.

4 Discussion

The results of this perception study in which AAE, ApEng, JE, and Non-target listeners were presented with stimuli in a persona matching task show consistent results across groups. In general, the groups are above chance at identifying the original lect of the clip in the unfiltered condition, with the exception of Appalachian English speakers by the Non-target group. **Table 14** summarizes the hit rates across the groups in the unfiltered condition.

Tables 15 and **16** show lect confusion by listener group, for unfiltered and filtered conditions respectively, with a reporting cutoff of 20% for lect confusion, given that listeners had 5 lect options to choose from for each clip. Note that these are arranged in groups to align down the columns, rather than in ranked order for each group.

Listener Group	AAE hit	ApEngl hit	JE Hit	Midwestern hit
African American	48%	24%	50%	32%
Appalachian	33%	30%	44%	39%
Jewish	29%	29%	55%	49%
Non-target	39%	16%	34%	31%

Table 14: Hit (accuracy) rates for each listener group, own lect.

Listener Group	Lect Confusion 1	Lect Confusion 2	Lect Confusion 3	Lect confusion 4 +
African American	ApEng as Southern (28%)	—	AAE as Midwestern (26%)	Midwestern as Southern (24%)
Appalachian	ApEng as Southern (42%)	JE as AAE (39%)	AAE as Midwestern (26%)	AAE as JE (23%); Midwestern as ApEng (23%)
Jewish	ApEng as Southern (43%)	JE as AAE (27%)	AAE as Midwestern (24%)	AAE as ApEng (21%); AAE as JE (21%)
Non-target	ApEng as Southern (44%)	JE as AAE (39%)	AAE as Midwestern (21%)	Midwestern as ApEngl (23%), Midwestern as Jewish (23%)

Table 15: Lect confusion by listener group (unfiltered).

Listener Group	Lect Confusion 1	Lect Confusion 2	Lect Confusion 3 +
African American	JE as AAE (42%)	ApEngl as AAE (28%)	AAE as Appalachian, JE, Midwestern (all 23%)
Appalachian	JE as AAE (61%)	ApEng as AAE (50%)	AAE as Midwestern (26%)
Jewish	JE as AAE (45%)	ApEng as AAE (39%)	Midwestern as Southern (28%)
Midwestern	JE as AAE (44%)	ApEng as AAE (40%)	AAE as Southern (31%)

Table 16: Lect confusion by listener group (filtered).

Differences emerge between filtering conditions (and somewhat between groups) when examining which groups are most likely to be confused for one another. In the unfiltered condition, all four listener groups were likely to confuse ApEng for Southern speakers, JE speakers for AAE speakers, and finally, AAE for Midwestern. In the filtered conditions, all four groups were still likely to confuse JE for AAE. Interestingly, listeners consistently did *not* confuse ApEng for Southern in this condition, but rather, ApEng for AAE. Additionally, we generally observe a greater degree of lect confusion in the filtered condition as opposed to the unfiltered condition. This is predicated, given previous studies that have demonstrated lower rates of listener accuracy and greater confusion when speech is subjected to filtering (Purnell, Idsardi & Baugh 1999; Thomas & Reaser 2004, *inter alia*). The differences in rates of confusion for different lects and different groups, however, do demonstrate that listeners are likely relying at least in part on prosodic information in their judgments, though prosodic information alone is not sufficient for identification. The patterns of confusion that we observed will be discussed in turn.

ApEng is often confused for Southern English in the unfiltered condition. This is perhaps unsurprising, given the similarities between the two lects and their geographic proximity. Preston (1996) describes a number of earlier studies whose results indicated that American listeners generally draw perceptual dialectology boundaries around the South as a single group, rarely differentiating between the area that the Appalachian Regional Commission defines as Appalachia, and the broader U.S. South. For these reasons, we may attribute the finding that listeners struggle to differentiate Southern and Appalachian varieties to the fact that they may simply lack familiarity with Appalachia as a region distinct from the South.

The confusion between ApEng and AAE for all groups in the filtered condition may also not be surprising, given the history of both lects. While aspects of the evolution of AAE are an ongoing subject of debate for sociolinguists, authors such as Van Herk (2015) describe the ways in which AAE developed alongside Southern and Appalachian white varieties. Van Herk (2015) also discusses a number of phonological and morphosyntactic features that are shared between AAE and Southern and Appalachian varieties, though he does not focus on prosodic variables. In a different study, Hazen & Fluharty (2004) specifically discuss how ApEng shares several phonological and morphosyntactic features with AAE. For this reason, when presented with a short audio clip, especially in a filtered condition, listeners also struggle to differentiate ApEng from AAE. This may be especially true for Appalachian listeners who are aware of their own regional variety and possibly more familiar with the ways in which it bears similarities to AAE. It is also worth noting here that regional variation within AAE may influence this pattern overall, but the speakers employed in this study were white speakers from Appalachia and AAE speakers from outside Appalachia. Future work should examine whether such a pattern holds for listeners who are exposed to both white and Black ApEng stimuli.

At first glance, the pattern of confusion on the part of all listener groups between JE and AAE may not be intuitive. Further complicating this finding is the fact that while all listener groups had a tendency to guess between AAE and JE for the JE clips, only the Jewish listeners in the unfiltered condition also confused JE as AAE. In previous work, Burdin et al. (2018; 2022) described the prosodic similarities and differences between AAE and JE, finding some commonalities in the realization of the L+H* pitch accent between the two groups, especially higher peaks and similar delay intervals, compared to Midwestern English. However, Jewish English's H* (and L+H*) look more like Appalachian English than African American English; see **Table 1** for H* data. Since the L+H* pitch accent was also the subject of the manipulation in the current study, these existing similarities in production may be responsible at least in part for these confusions, but it is unclear why this would be the case for JE and AAE but not JE and ApEng, and in the filtered condition. The difference of the Jewish listeners here may be explained by the fact JE, like ApEng may be less enregistered as a lect for listeners than either Southern varieties or AAE. Burdin (2020) finds that many non-Jewish listeners are less accurate at identifying Jewish English prosody, suggesting that the variety may be less enregistered than others. As a result, non-Jewish listeners may not feel confident in their ability to differentiate JE from other, more salient varieties that share prosodic features; Jewish listeners may be more willing to make these choices, and then, overcorrect and misidentify other similar lects as Jewish English. An additional factor here is that both the AAE and JE personas were presented as being from New York City; the data from both were also collected from the NYC region. Although there are differences between AAE and JE in NYC (Becker 2014; Labov 1966), shared regional features may have also influenced confusion here, particularly in the unfiltered condition.

Finally, we observe a bias in the filtered condition such that only the Midwesterners were correctly identified above chance for all four groups of listeners. It is important to note that all of the stimuli were gathered in structured reading tasks and are therefore largely devoid of regionally or ethnically stigmatized morphosyntactic features. Perceptual dialectology research (Preston 1986; 1996, *inter alia*) has consistently found a bias towards Midwestern varieties as the most prescriptively “standard” or “correct” for lay listeners. The lack of regionally identifiable features as well as the formality of the stimuli likely contributed towards the selection of non-stigmatized varieties across conditions, thus predicting the bias towards Midwestern. The confusion of AAE for Midwestern in the unfiltered condition also supports evidence of this bias; AAE is frequently heavily stereotyped as maximally distant from the prescriptive standard specifically due to morphosyntactic differences, though this is not necessarily the case for Jewish English or Appalachian English. In the absence of stigmatized morphosyntactic features of AAE, listeners may demonstrate more confusion between it and other varieties that are perceived as “more standard”.

Finally, the current study has demonstrated the utility of a persona task for lect identification, though we acknowledge that this design functions best with a limited number of speakers. We argue that providing listeners with normed photos and descriptions of speakers allows them to engage in lect identification in a manner that more closely resembles the judgments they make in everyday life. Our results show that listeners can still perform above chance in a lect identification task when provided with this additional information; also, it allows us to better control assumptions about social information that listeners make in such a task. This study acts as a first step for future research in lect identification that utilizes a variety of methods that leverage how listeners make sociolinguistic judgments in their everyday lives. Future work should also examine the role of specific segmental phenomenon associated with different lects alongside suprasegmental phenomena, in order to better understand the features that listeners attune to in lect identification, as well as how co-occurring features may influence such judgments. Additionally, future work should compare conditions with photos and audio stimuli, audio stimuli alone, and audio stimuli with photos and the social information.⁶ While it is difficult to isolate features of ethnic, religious, and regional lects simultaneously, we do know that listeners do not individually judge each aspect of a speaker's personae; rather, they combine socio-indexical information with what they know about speaker characteristics in order to arrive at a lect judgment.

⁶ We thank an anonymous reviewer for this suggestion.

Appendix A: Regression Table

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.1621	0.0929	44.8	< 2e-16 ***
Diag(response, original, Group) Midwestern	-0.2948	0.139	-2.1	0.034 *
Diag(response, original, Group) AfricanAmerican	0.4676	0.1639	2.9	0.004 **
Diag(response, original, Group) Appalachian	0.5646	0.182	3.1	0.002 **
Diag(response, original, Group) Jewish	0.5895	0.1593	3.7	< 2e-04 ***
originalAfricanAmerican	-0.4396	0.0701	-6.3	< 4e-10 ***
originalAppalachian	-0.3296	0.066	-5	< 6e-07 ***
originalJewish	-0.4325	0.0683	-6.3	< 2e-10 ***
responseAfricanAmerican	0.2767	0.1045	2.6	0.008 **
responseAppalachian	0.1713	0.1083	1.6	0.114
responseJewish	0.1862	0.1061	1.8	0.079.
responseSouthern	0.0234	0.1059	0.2	0.825
filteredTRUE	0.0198	0.089	0.2	0.824
Diag(response, original) Midwestern	0.9267	0.0928	10	< 2e-16 ***
Diag(response, original) AfricanAmerican	0.017	0.1207	0.1	0.888
Diag(response, original) Appalachian	-0.2142	0.1378	-1.6	0.12
Diag(response, original)Jewish	0.2917	0.1192	2.4	0.014 *
GroupAfricanAmerican	0.3898	0.1267	3.1	0.002 **

(Contd.)

	Estimate	Std. Error	z value	Pr(> z)
GroupAppalachian	0.1662	0.126	1.3	0.187
GroupJewish	0.217	0.1254	1.7	0.083.
responseAfricanAmerican:- filteredTRUE	0.1152	0.1235	0.9	0.351
responseAppalachian:- filteredTRUE	-0.2327	0.1347	-1.7	0.084.
responseJewish:filteredTRUE	-0.1032	0.1276	-0.8	0.419
responseSouthern:filteredTRUE	0.0764	0.1325	0.6	0.564
originalAfricanAmeric- an:GroupAfricanAmerican	-0.1618	0.0981	-1.7	0.099.
originalAppalachian:- GroupAfricanAmerican	-0.1302	0.0924	-1.4	0.159
originalJewish:GroupAfrica- nAmerican	-0.1961	0.0963	-2	0.042 *
originalAfricanAmerican:- GroupAppalachian	0.1461	0.0968	1.5	0.131
originalAppalachian:GroupAp- palachian	-0.0282	0.091	-0.3	0.757
originalJewish:GroupAppalachian	-0.0042	0.094	0	0.964
originalAfricanAmerican:- GroupJewish	0.1755	0.0958	1.8	0.067.
originalAppalachian:GroupJewish	0.0123	0.0922	0.1	0.894
originalJewish:GroupJewish	-0.0517	0.0954	-0.5	0.588
responseAfricanAmerican:- GroupAfricanAmerican	-0.4198	0.1452	-2.9	0.004 **
responseAppalachian:GroupAfric- anAmerican	-0.3691	0.1501	-2.5	0.014 *
responseJewish:GroupAfrica- nAmerican	-0.3253	0.1458	-2.2	0.026 *

(Contd.)

	Estimate	Std. Error	z value	Pr(> z)
responseSouthern:GroupAfricanAmerican	-0.068	0.1427	-0.5	0.634
responseAfricanAmerican:-GroupAppalachian	-0.1111	0.1431	-0.8	0.437
responseAppalachian:GroupAppalachian	-0.2424	0.1488	-1.6	0.103
responseJewish:GroupAppalachian	-0.0856	0.1439	-0.6	0.552
responseSouthern:GroupAppalachian	-0.0164	0.1429	-0.1	0.909
responseAfricanAmerican:-GroupJewish	-0.3229	0.1449	-2.2	0.026 *
responseAppalachian:GroupJewish	-0.2149	0.1467	-1.5	0.143
responseJewish:GroupJewish	-0.1698	0.143	-1.2	0.235
responseSouthern:GroupJewish	-0.186	0.1439	-1.3	0.196
filteredTRUE:GroupAfricanAmerican	-0.0498	0.1207	-0.4	0.68
filteredTRUE:GroupAppalachian	-0.1196	0.119	-1	0.315
filteredTRUE:GroupJewish	-0.2229	0.118	-1.9	0.059.
Diag(response, original)Midwestern:GroupAfricanAmerican	-0.5186	0.134	-3.9	< 1e-04 ***
Diag(response, original)AfricanAmerican:GroupAfricanAmerican	NA	NA	NA	NA
Diag(response, original)Appalachian:GroupAfricanAmerican	0.4797	0.1822	2.6	0.008 **
Diag(response, original)Jewish:-GroupAfricanAmerican	0.5607	0.1605	3.5	< 5e-04 ***

(Contd.)

	Estimate	Std. Error	z value	Pr(> z)
Diag(response, original)Midwest- ern:GroupAppalachian	-0.0706	0.1319	-0.5	0.593
Diag(response, original)Africa- nAmerican:GroupAppalachian	-0.2626	0.1636	-1.6	0.108
Diag(response, original) Appalachian:GroupAppalachian	NA	NA	NA	NA
Diag(response, original)Jewish:- GroupAppalachian	0.3315	0.1614	2.1	0.040 *
Diag(response, original)Midwest- ern:GroupJewish	NA	NA	NA	NA
Diag(response, original)Africa- nAmerican:GroupJewish	-0.3141	0.1706	-1.8	0.066.
Diag(response, original) Appalachian:GroupJewish	0.4717	0.1781	2.6	0.008 **
Diag(response, original)Jewish:- GroupJewish	NA	NA	NA	NA
responseAfricanAmerican:- filteredTRUE:GroupAfricanAmer- ican	0.1323	0.1705	0.8	0.438
responseAppalachian:- filteredTRUE:GroupAfricanAmer- ican	0.2448	0.1836	1.3	0.182
responseJewish:filteredTRUE:- GroupAfricanAmerican	0.0484	0.1741	0.3	0.781
responseSouthern:filteredTRUE:- GroupAfricanAmerican	-0.1536	0.1811	-0.8	0.396
responseAfricanAmerican:- filteredTRUE:GroupAppalachian	0.5387	0.1658	3.2	0.001 **
responseAppalachian:- filteredTRUE:GroupAppalachian	0.1239	0.1836	0.7	0.5
responseJewish:filteredTRUE:- GroupAppalachian	-0.1361	0.1743	-0.8	0.435

(Contd.)

	Estimate	Std. Error	z value	Pr(> z)
responseSouthern:filteredTRUE:- GroupAppalachian	-0.1078	0.1813	-0.6	0.552
responseAfricanAmerican:- filteredTRUE:GroupJewish	0.5142	0.1694	3	0.002 **
responseAppalachian:- filteredTRUE:GroupJewish	0.4514	0.1788	2.5	0.012 *
responseJewish:filteredTRUE:- GroupJewish	-0.0509	0.1726	-0.3	0.768
responseSouthern:filteredTRUE:- GroupJewish	0.3121	0.1796	1.7	0.082.

Data availability

Data and R code is available at at <https://osf.io/v49bk/>.

Ethics and consent

This research was overseen by the Internal Review Boards of Pomona College (02/02/2017NH-FL) the University of New Hampshire (Legacy-UNH-6997) and the University of Alabama (UA IRB – 18-08-1394).

Acknowledgements

The authors would like to thank Emelia Benson Meyer for assistance with stimuli creation, Neal Fultz for statistical assistance, and Dan Villareal for comments on an earlier version of this manuscript. Partial funding for this project was provided by an internal grant from the University of Pennsylvania to the first author.

Competing interests

The authors have no competing interests to declare.

References

- Arvaniti, Amalia & Garding, Gina. 2007. Dialectal variation in the rising accents of American English. In Cole, Jennifer & Hualde, Jose Ignacio (eds.), *Phonetics and Phonology 4(3): Laboratory Phonology 9*, 547–575. New York: Mouton de Gruyter.
- Becker, Kara. 2014. (r) we there yet? The change to rhoticity in New York City English. *Language Variation and Change* 26(2). 141–168. DOI: <https://doi.org/10.1017/S0954394514000064>

- Beckman, Mary E. & Elam, Gayle Ayers. 1997. *Guidelines for ToBI labelling, version 3.0*.
- Burdin, Rachel Steindel. 2016. *Variation in form and function in Jewish English intonation*. Doctoral Thesis. The Ohio State University.
- Burdin, Rachel Steindel. 2020. The perception of macro-rhythm in Jewish English intonation. *American Speech* 95(3). 263–296. DOI: <https://doi.org/10.1215/00031283-7706542>
- Burdin, Rachel Steindel, Holliday, Nicole R. & Reed, Paul E. 2018. Rising above the standard: Variation in L+H* use across 5 varieties of American English. In Klessa, Katarzyna & Bachan, Jolanta & Wagner, Agnieszka & Karpiński, Maciej & Śledziński, Daniel (Eds.), *9th International Conference on Speech Prosody*, 354–358. DOI: <https://doi.org/10.21437/SpeechProsody.2018-72>
- Burdin, Rachel Steindel & Holliday, Nicole R. & Reed, Paul E. 2022. American English pitch accents in variation: Pushing the boundaries of mainstream American English-ToBI conventions. *Journal of Phonetics* 94. 101163. DOI: <https://doi.org/10.1016/j.wocn.2022.101163>
- Clopper, Cynthia G. & Pisoni, David B. 2004a. Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change* 16(1). 31–48. DOI: <https://doi.org/10.1017/S0954394504161036>
- Clopper, Cynthia G. & Pisoni, David B. 2004b. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics* 32(1). 111–140. DOI: [https://doi.org/10.1016/S0095-4470\(03\)00009-3](https://doi.org/10.1016/S0095-4470(03)00009-3)
- Clopper, Cynthia G. & Levi, Susannah V. & Pisoni, David B. 2006. Perceptual similarity of regional dialects of American English. *The Journal of the Acoustical Society of America* 119(1). 566–574. DOI: <https://doi.org/10.1121/1.2141171>
- Clopper, Cynthia G. & Smiljanic, Rajka. 2011. Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39(2). 237–245. DOI: <https://doi.org/10.1016/j.wocn.2011.02.006>
- Cramer, Jennifer. 2018. Perceptions of Appalachian English in Kentucky. *Journal of Appalachian Studies* 24(1). 45–71. DOI: <https://doi.org/10.5406/jappastud.24.1.0045>
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. DOI: <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Fairbanks, G. 1960. The Rainbow Passage. In *Voice and articulation drillbook*, 2nd ed., 124–139. Harper and Row.
- Foreman, Christina Gayle. 2000. Identification of African American English from prosodic cues. In *Texas Linguistic Forum*, Vol. 43. 57–66. Austin: University of Texas.
- Hall, J.S. 1942. 3. The Consonants. In *The Phonetics of Great Smoky Mountain Speech*, 85–106. Columbia University Press. DOI: <https://doi.org/10.7312/hall93950-005>
- Hawkins, Francine D. 1992. *Speaker ethnic identification: The roles of speech sample, fundamental frequency, speaker and listener variations*. Doctoral Thesis. University of Maryland, College Park.
- Hazen, Kirk & Fluharty, Ellen. 2004. Defining Appalachian English. In Margaret Bender (ed.), *Linguistic diversity in the South: changing codes, practices, and ideology*, 50–65. Southern Anthropological Society.

- Holliday, Nicole R. 2016. *Intonational variation, linguistic style and the Black/Biracial experience*. Doctoral Thesis. New York University.
- Holliday, Nicole R. & Villarreal, Dan. 2020. Intonational variation and incrementality in listener judgments of ethnicity. *Laboratory Phonology* 11(1). DOI: <https://doi.org/10.5334/labphon.229>
- Koutstaal, Cornelis W. & Jackson, Faith L. 1971. Race identification on the basis of biased speech samples. *Ohio Journal of Speech and Hearing*, 6, 48–51.
- Knoll, Monja A. & Uther, Maria & Costall, Alan. 2009. Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Communication* 51(3). 210–216. DOI: <https://doi.org/10.1016/j.specom.2008.08.001>
- Labov, William. 1966. *The social stratification of English in New York City*. Washington D. C.: Center for Applied Linguistics.
- Lass, Norman J. & Almerino, Celest A. & Jordan, Laurie F. & Walsh, Jayne M. 1980. The effect of filtered speech on speaker race and sex identifications. *Journal of Phonetics* 8(1). 101–112. DOI: [https://doi.org/10.1016/S0095-4470\(19\)31445-7](https://doi.org/10.1016/S0095-4470(19)31445-7)
- Leeman, Adrian & Kolly, Maria-Jose & Nolan, Francis & Li, Yang. 2018. The role of segments and prosody in the identification of a speaker's dialect. *Journal of Phonetics* 68. 69–84. DOI: <https://doi.org/10.1016/j.wocn.2018.02.001>
- Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1). 62–85. DOI: <https://doi.org/10.1177/0261927X99018001005>
- McCullough, Jill & Somerville, Barbara. 2000. *Comma Gets a Cure*.
- McGowan, Kevin B. 2015. Social expectation improves speech perception in noise. *Language and Speech*, 58(4), 502–521. DOI: <https://doi.org/10.1177/0023830914565191>
- McGowan, Kevin B. & Babel, Anna M. 2020. Perceiving isn't believing: listeners' expectation and awareness of phonetically-cued social information. *Language in Society*, 49(2). 1–26. DOI: <https://doi.org/10.1017/S0047404519000782>
- Plichta, Bartłomiej & Preston, Dennis R. 2005. The /ay/s have It the perception of /ay/ as a north-south stereotype in United States English. *Acta Linguistica Hafniensia* 37(1). 107–130. DOI: <https://doi.org/10.1080/03740463.2005.10416086>
- Preston, Dennis R. 1986. Five visions of America. *Language in Society* 15(2). 221–240. DOI: <https://doi.org/10.1017/S0047404500000191>
- Preston, Dennis R. 1996. Where the worst English is spoken. *Focus on the USA*, 16, 297–360. DOI: <https://doi.org/10.1075/veaw.g16.16pre>
- Purnell, Thomas & Idsardi, William & Baugh, John. 1999. Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(1). 10–30. DOI: <https://doi.org/10.1177/0261927X99018001002>
- Reed, Paul E. 2016. *The sociophonetics of Appalachian English: Identity, intonation, and monophthongization*. Doctoral Thesis. University of South Carolina.

- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Sellers, Frances Stead. 2014, December 3. How do you speak D.C.? ‘Accent tags’ help to define the District’s dialect. *Washington Post*. Retrieved from <https://www.washingtonpost.com/news/arts-and-entertainment/wp/2014/07/09/how-do-you-speak-d-c-accent-tags-help-to-define-the-districts-dialect/>
- Thomas, Erik R. 2002. Sociophonetic applications of speech perception experiments. *American Speech* 77(2). 115–147. DOI: <https://doi.org/10.1215/00031283-77-2-115>
- Thomas, Erik R. 2015. Prosodic Features of African American English. In Sonja L. Lanehart (ed.), *The Oxford handbook of African American English*, 420–435. Oxford University Press.
- Thomas, Erik R. & Reaser, Jeffrey. 2004. Delimiting perceptual cues used for the ethnic labeling of African American and European American voices. *Journal of Sociolinguistics* 8(1). 54–87. DOI: <https://doi.org/10.1111/j.1467-9841.2004.00251.x>
- Thomas, Erik R. & Lass, Norman E. & Carpenter, Jeannine. 2010. Chapter 12: Identification of African American Speech. In Dennis R Preston & Nancy Niedzielski (eds.), *A Reader in Sociophonetics*, 265–288. De Gruyter Mouton. DOI: <https://doi.org/10.1515/9781934078068.2.265>
- Van Herk, Gerard. 2015. The English Origins Hypothesis. In Jennifer Bloomquist & Lisa J. Green & Sonja L. Lanehart (eds.), *The Oxford handbook of African American Language*, 0. Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199795390.013.3>
- Walton, Julie H. & Orlikoff, Robert F. 1994. Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research* 37(4). 738–745. DOI: <https://doi.org/10.1044/jshr.3704.738>

