Cross-Dialectal Perceptual Experiences of Speech-Language Pathologists in Predominantly Caucasian American School Districts

Gregory C. Robinson
University of Arkansas at Little Rock and University of Arkansas for Medical Sciences
Ida J. Stockman
Michigan State University, East Lansing

Over the years, school-based speech-language pathologists (SLPs) have become aware of the need to differentiate between speech-language differences due to a clinical disorder and those associated with a nonmainstream dialect (American Speech-Language-Hearing Association [ASHA], 1983, 1985; Individuals With Disabilities Education Improvement Act [IDEA], 2004). There are challenges to making this distinction in a practical sense, particularly when dealing with dialect patterns that are similar to developmental or disordered speech patterns in the standard dialect. The very definition of *disorder* may be culturally determined (Payne & Taylor, 2005). The practice of school-based SLPs in the United States is governed in part by IDEA, which offers the following definition of *speech-language impairment*: "a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that adversely affects a child's educational performance" (IDEA, 2004, Sec. 300.8(c) (11)). However, this definition leaves the term *communication disorder* underspecified. SLPs might refer to Van Riper's (1978) classic definition of a *speech disorder,* which specifies speech that (a) calls attention to itself, (b) interferes with communication, or (c) places an emotional burden on the speaker. It is conceivable that all of Van Riper's criteria could "affect a child's educational performance," as referenced in the IDEA definition. Van Riper's definition offers little help for SLPs, however, because the three criteria could also apply to a child who speaks a nonmainstream English dialect or an English accent influenced by another language. Alternatively, Payne and Taylor suggested that this classic Van Riperian definition

ABSTRACT: **Purpose:** This study aimed to determine if the number and type of African American English (AAE) features that are spoken in sentences influence speech-language pathologists' (SLPs') judgments of (a) how noticeable the dialect is (dialect detectability) and (b) how understandable a speaker is to others (comprehensibility).
Method: Certified SLPs with little conversational experience with AAE were recruited from predominantly Caucasian American school districts in Michigan. They listened to sentences that contained varying amounts and types of AAE phonological features. The SLPs rated the sentences on 5-point scales regarding dialect detectability and comprehensibility. The ratings for the different sentences were compared to determine how the variables contributed to the SLPs' judgments of AAE.
Results: Both dialect detectability and comprehensibility ratings were affected by the number of AAE features that were included in the sentences. The types of AAE features consistently affected the comprehensibility ratings but less consistently affected the dialect detectability ratings.
Conclusion: Multiple factors may affect SLPs' perceptions of AAE. The outcomes have both theoretical and practical implications.

KEY WORDS: African American English, speech perception, speech-language pathologists, dialect, comprehensibility, dialect detectability, dialect density

0161-1461/09/4002-0138

of a speech disorder is sufficient, with the understanding that all communication is socioculturally based. Therefore, an individual has a speech disorder when his or her speech draws attention to itself, is difficult to understand, or places an emotional burden on the speaker *within his or her home speech community.*

Diagnostic Issues Related to Normal Speaker Variation

School-based SLPs, who are not familiar with normal non-mainstream dialect development, may have a tendency to judge speakers as disordered when they are not. Such issues of misdiagnosis have been discussed in IDEA (2004). IDEA states that schools with predominantly Caucasian students and teachers are apt to place minority students into special education classes at disproportionately high rates.

Diagnostic issues may be complicated by normal intraspeaker variation. Over the last decade, studies have revealed that speakers of African American English (AAE) vary on a continuum of using the various features or patterns associated with the dialect (Craig, Thompson, Washington, & Potter, 2003; Craig & Washington, 2004; Stockman, 2004; Washington & Craig, 1994). Some speakers may use just a few features; others use many. Washington and Craig were among those who first exposed this kind of variation across AAE child speakers by calling attention to dialect density. *Dialect density* refers to the average number of dialect features that are used in a given linguistic unit. This has been a scholarly way to conceptualize what lay listeners have regarded as the *thickness* of accents or dialects. Researchers have used the notion of dialect density with the aim to quantify dialect use (for a summary, see Oetting & McDonald, 2002).

Little is still known about how dialectal production influences perceptual judgments, particularly those made by SLPs. Intuitively, it would seem that the more AAE features that are used by a speaker, the more detectable the dialect will be. At the same time, increased dialect density may reduce speech intelligibility for listeners who are unfamiliar with the dialect. The latter effect has obvious clinical relevance because measures of speech intelligibility are often viewed as key tools for speech-language assessments. It is already known that perceived intelligibility is affected by some speech articulation patterns more than others. For example, Klein and Flint (2006) revealed that final consonant deletion affected intelligibility more than the fronting of velars and stopping of fricatives and affricates did. Notably, final consonant deletion is a commonly cited feature in AAE (Bailey & Thomas, 1998; Pollock et al., 1998; Stockman, 1996, 2004, 2006). It is unknown how final consonant deletion and other AAE phonological features may influence SLPs' judgments of intelligibility and dialect detectability when listening to AAE speakers. The likelihood of misdiagnosis may increase, however, if intelligibility is affected by the use of particular AAE phonological patterns.

Such misdiagnoses may explain why African American students are overrepresented on SLPs' caseloads in U.S. schools when compared to the general population (IDEA, 2004; U.S. Department of Special Education Programs, 2004). Most SLPs (93%) identify themselves as Caucasian American (ASHA, 2005) and may not speak AAE, which is spoken by many of their African American clients. AAE speakers vary in the number and types of dialect features used, so the language of some of these speakers may be difficult to accurately assess for non-AAE-speaking SLPs.

Focus of This Study

This study broadly aimed to determine if and how the variable use of AAE phonological features influences the perceptions of dialect use and listener comprehension. It focused particularly on the judgments of school-based SLPs who have had little experience with AAE.

Dialect detectability and comprehensibility. A review of second-language acquisition research informed the choice of perceptual variables that were investigated. This review identified two different ratings as relevant to speech perceptibility: detectability and comprehensibility (Brennan & Brennan, 1981a, 1981b; Burda, Scherz, Hageman, & Edwards, 2003; Derwing & Munro, 1997; Flege & Fletcher, 1992; Johansson, 1978; Munro & Derwing, 1995a, 1995b). *Detectability* is the degree to which a language variety is noticed by a listener; *comprehensibility* is the level of difficulty experienced in recognizing the words (Munro & Derwing, 1995a). Comprehensibility, while related to intelligibility, differs from it. *Intelligibility* is a dichotomous, as opposed to a gradient, measure, which refers to whether the words are successfully recognized or not. In rating comprehensibility, participants indicate their *effort* in retrieving the message (Burda et al., 2003; Derwing & Munro, 1997; Gass & Varonis, 1984; Munro & Derwing, 1995a, 1995b). Using these concepts to study the perception of AAE was important because the relationship between AAE use by school-age children and academic achievement has been a widely debated topic in educational settings (Adger, Wolfram, & Christian, 2007; Baugh, 2000; Ramirez, 2005).

Perception of AAE phonology. Although AAE's phonology has been studied less than its morphosyntax (Bailey & Thomas, 1998), phonology may contribute most to its perceptibility and was the focus of the current study. It is known that the phonological features of nonnative accents carry greater perceptual weight than do morphosyntactic features and therefore could have the greater effect on intelligibility (Johansson, 1978). One reason is that phonological features occur more frequently than morphosyntactic ones (Craig et al.; 2003; Stockman, Guillory, Newkirk, & Seibert, in preparation; Thompson, Craig, & Washington, 2004). In fact, phonological features may occur so frequently that they create a gestalt perception of a dialect rather than awareness of specific features. In contrast, morphosyntactic features may be more categorically salient. Listener judgments about speech pronunciation are likely to be influenced by numerous factors, two of which relate to the dialect patterns that are used in a given unit of perceived speech: (a) the number of dialect features present and (b) the type of dialect features present.

Number and type of phonological features. If a speaker increases the number of dialect features that are used in a given unit of spoken language, then it is reasonable to expect that particular dialect to be detected more quickly and easily by others. For example, Ryan, Carranza, and Moffie (1977) revealed that the degree of perceived foreign accent was positively correlated with the number of segmental substitutions in a brief reading passage. This outcome offers some perceptual justification for methods designed to quantify dialect density. Oetting and McDonald (2002) have summarized several of these methods.

Dialect features may not affect perceptual judgments equally, as is implied by some of the formulae for calculating dialect density. The perceptual effect may depend on what type of dialect features are included in the sentence. Each additional feature could affect perception in differing ways. It also is possible that using a certain number of features is necessary before a dialect can be detected (i.e., threshold) or that the inclusion of a certain number of features may cause the perceptual effect to level off as features are added (i.e., a plateau).

Furthermore, some dialect features may affect perception more than others (Johansson, 1978). For example, a single linguistic cue may carry so much perceptual weight that it overrides the perception of other dialect features in the same context. Anecdotally, one may consider the /aɪ/ pronounced as /a/ in Southern U.S. dialects, or *ask* pronounced [æks] in AAE. Graff, Labov, and Harris (1983) demonstrated this phenomenon empirically. When the onset of the diphthong /aʊ/ was acoustically altered to become [æʊ] in the speech of an African American, a group of Philadelphia judges identified the speaker as Caucasian American—despite the AAE phonological features that were included in the rest of the sentence. Purnell, Idsardi, and Baugh (1999) discovered that even relatively subtle linguistic cues may have grave social consequences for speakers. When a speaker pronounced the word, "hello," in three different dialects (Standard American English [SAE], AAE, and Chicano English), variations in vowel tenseness and pitch on the first vowel were enough to trigger assumptions about the speaker's race/ethnicity. In a second experiment, the utterance was expanded to "Hello, I'm calling about the apartment you have advertised in the paper" and was presented to actual property owners. The SAE utterances elicited appointments approximately 70% of the time. The AAE and Chicano English versions elicited appointments approximately 30% of the time.

Johansson (1978) studied how various speech errors that were produced by native Swedish speakers, who were learning to speak English, affected perceptual ratings of "irritability" and intelligibility made by native English speakers. It was concluded that (a) phonological errors carried more weight than morphosyntactic ones, (b) phonemic errors were judged as more serious than subphonemic ones, and (c) subphonemic errors that differed greatly from the target were judged more severely than were those that did not.

The features that are likely to be detected or to impair comprehensibility to unfamiliar listeners may be based on their perceptual salience (i.e., the degree to which listeners can detect a feature in nonmeaningful speech). Research suggests that, cross-linguistically, speakers tend to use phonological alternations that are the least perceptually distant from an intended phonemic prototype, all while attempting to produce the least marked production (Steriade, 2004). If so, then those features that are highly salient in nonsense words may increase dialect detectability ratings and decrease perceptions of comprehensibility more than less salient features.

Research Question

This study aimed to determine if the perceptions of AAE by school-based Caucasian American SLPs were influenced by the number of features that were present in sentences and their degree of perceptual salience (i.e., the degree to which the pattern could be detected in nonsense words). It was hypothesized that (a) the number of dialect features in sentences would increase dialect detectability ratings and decrease comprehensibility ratings, and (b) the features with high perceptual salience would elicit higher dialect detectability ratings and lower comprehensibility ratings than those features with low salience.

METHOD

Participants

The participants were 16 Caucasian American SLPs who were recruited from two adjacent, rural school districts in Michigan's Lower Peninsula. They each signed a statement of informed consent per the guidelines of the Michigan State University Institutional Review Board. The school districts' student populations were at least 90% Caucasian American and less than 5% African American, as indicated by the Census from the preceding school year (Annie E. Casey Foundation, 2003). The participants' total years of experience as SLPs ranged from 3 to 33 years ($M = 11.88$, $SD = 8.77$).

To qualify for participation in the study, the SLPs indicated on a questionnaire that they (a) conversed with AAE speakers less than 5% of the time, (b) had had fewer than three clients within the last 3 years who spoke AAE, (c) did not speak AAE, and (d) had worked as an SLP for at least 2 years. Each participant also passed an air conduction hearing screening at 25db HL for 1000, 2000, and 4000 Hz. The screening was administered with a Beltone portable audiometer (Model #119) and bilateral earphones in a quiet setting.

Description of the Stimuli

The independent experimental variables in this study were (a) frequency of AAE features and (b) perceptual salience of AAE features. The stimuli consisted of six different sentence wordings that were selected to feature various combinations of these two variables (Table 1). Sentences varied only in the subject noun and object of preposition positions. The words of the sentence recordings were produced with different pronunciation patterns, as follows: (a) six sentences had no AAE phonological feature, (b) six sentences had one AAE feature, and (c) six sentences had three AAE features (see Table 2).

In total, the participants made 42 ratings: 36 ratings for the dialect detectability task and 6 for the comprehensibility task. The 18 items in Table 2 were presented twice. They were spoken by one AAE speaker during one presentation and by another AAE speaker in the following presentation (both randomized), creating a total of 36 stimulus sentences. The participants rated each of the 36 sentences individually during the dialect detectability task and rated the 36 sentences presented in six groups of six items each

Table 1. Six different stimulus sentences as categorized for perceptual salience opportunities. Underlined words indicate variable pronunciation patterns among the stimulus sentences. See Table 2 for specific pronunciation patterns.

Sentence number	*Sentence*	*Perceptual salience classification*
1	A <u>sock</u> may be <u>with</u> my <u>teeth</u> Monday.	More salient
2	A <u>cat</u> may be <u>with</u> my <u>sock</u> Monday.	
3	A <u>tooth</u> may be <u>with</u> my <u>boat</u> Monday.	
4	A <u>desk</u> may be <u>with</u> my <u>teeth</u> Monday.	Less salient
5	A <u>guest</u> may be <u>with</u> my <u>desk</u> Monday.	
6	A <u>tooth</u> may be <u>with</u> my <u>list</u> Monday.	

Table 2. Listing of stimulus sentences and targets for planned alternations phonetically transcribed.

Sentence number	*Noun phrase*	*Verb complex*	*Prep.*	*Det.*	*Obj.*	*Adverbial*	*AAE phonological features present*	*No. of features*	*Perceptual salience of features*
1	ə sɑk	me bi	wɪθ	maɪ	tiθ	mʌnde	None	0	More
2	ə kæt	me bi	wɪθ	maɪ	sɑk	mʌnde	None	0	More
3	ə tuθ	me bi	wɪθ	maɪ	bot	mʌnde	None	0	More
4	ə dɛsk	me bi	wɪθ	maɪ	tiθ	mʌnde	None	0	Less
5	ə gɛst	me bi	wɪθ	maɪ	dɛsk	mʌnde	None	0	Less
6	ə tuθ	me bi	wɪθ	maɪ	lɪst	mʌnde	None	0	Less
1	ə sɑ	me bi	wɪθ	maɪ	tiθ	mʌnde	FDk	1	More
2	ə kæ	me bi	wɪθ	maɪ	sɑk	mʌnde	FDt	1	More
3	ə tut	me bi	wɪθ	maɪ	bot	mʌnde	t/th	1	More
4	ə dɛs	me bi	wɪθ	maɪ	tiθ	mʌnde	FRsk	1	Less
5	ə gɛs	me bi	wɪθ	maɪ	dɛsk	mʌnde	FRst	1	Less
6	ə tuf	me bi	wɪθ	maɪ	lɪst	mʌnde	f/th	1	Less
1	ə sɑ	me bi	wɪt	maɪ	tit	mʌnde	FDk, t/th, t/th	3	More
2	ə kæ	me bi	wɪt	maɪ	sɑ	mʌnde	FDt, t/th, FDk	3	More
3	ə tut	me bi	wɪt	maɪ	bo	mʌnde	t/th, t/th, FDt	3	More
4	ə dɛs	me bi	wɪf	maɪ	tif	mʌnde	FRsk, f/th, f/th	3	Less
5	ə gɛs	me bi	wɪf	maɪ	dɛs	mʌnde	FRst, f/th, FRsk	3	Less
6	ə tuf	me bi	wɪf	maɪ	lɪs	mʌnde	f/th, f/th, FRst	3	Less

Note. AAE = African American English, FDk = final deletion of /k/, FDt = final deletion of /t/, t/th = substitution of [t] for /θ/, FRsk = final cluster reduction of sk, FRst = final cluster reduction of st, and f/th = substitution of [f] for /θ/. Shaded areas represent changes from Standard American English.

during the comprehensibility task organized according to the variables of frequency and salience. For example, stimuli with one low-salience AAE feature were grouped together and those with three high-salience AAE features were grouped together until all combinations of the two variables were exhausted.

Construction of the Stimuli

As stated previously, the number of dialect features (i.e., frequency) was represented by three groups of stimulus sentences, those with zero, one, and three AAE features. When one feature was present, it always occurred in the subject noun position. When three features were present, they occurred in the subject noun, the preposition *with,* and the prepositional object noun positions. Such consistency among the sentences was necessary to control for suprasegmental and coarticulatory differences that might enhance or diminish the perception of features. In the case of this study's stimuli, the subject noun and object noun were stressed by the speakers, and the preposition *with* was unstressed. Furthermore, all of the opportunities for AAE patterns preceded a word beginning with /m/, which controlled for coarticulatory effects among the sentences.

The dialect features included in the stimulus sentences were (a) final consonant deletion of /t/, (b) final consonant deletion of /k/, (c) final substitution of /t/ for /θ/, (d) final substitution of /f/ for /θ/, (e) final cluster reduction of /st/ to /s/, and (f) final cluster reduction of /sk/ to /s/. These six specific patterns represent three of the most frequently occurring consonantal patterns that were used by African American children in a study by Craig et al. (2003): (a) final consonant deletion, (b) "th" substitutions, and (c) cluster reduction. In the present study, patterns involving /t/ sounds were categorized separately from those involving /k/ sounds. Final /t/ has been cited as a commonly deleted phoneme in many American English dialects (Guy, 1980; Stockman, 2006); final /k/ deletion is not as commonly documented. The t/th and f/th substitutions were also separately categorized because they are distinct "th" substitution patterns that are variably found in AAE (Pollock et al., 1998).

These six phonological patterns were categorized into two salience groups. This was done to combine different sound patterns in individual sentences while maintaining a systematic basis for comparison. Evidence from Labov (1966) suggested that if the same phonological feature is perceived multiple times, then a cumulative perceptual effect could result. Although this was a valid point to consider, it was beyond the scope of this study. Categorization allowed different features to be clustered in single sentences, thus preventing effects from a sole feature being repeatedly experienced in single sentences.

The relative perceptual salience of the phonological patterns under investigation was determined from a preliminary experiment that was developed from suggestions by Steriade (2004; see Robinson, 2006, Appendix A, for a full description of the preliminary study). In summary, Caucasian American speech-language pathology graduate students were presented with pairs of two-syllable nonsense words that simulated the phonological environment of the variable subject noun followed by the word *may* used in this study's stimuli. When the nonsense words differed, they simulated the phonetic contrasts between AAE and SAE, which were examined in the study. For example, to determine the perceptual salience of final /t/ deletion, the contrasted words were [dat me] and [da me]. The [dat] in the first syllable of the first nonsense word simulated a subject noun with a final /t/; the [da] in the first syllable of the second nonsense word simulated the same word in which the final /t/ was deleted. The [me] in the second syllable of both words represented the word *may* immediately following the subject nouns in the study.

Participants rated each pair of nonsense words on a 4-point scale (1 = *exactly the same,* 2 = *almost the same,* 3 = *significantly different,* and 4 = *extremely different*). Some pairs of nonsense words were identical (e.g., [dat me] – [dat me]). These were considered control pairs. The ratings for the control pairs were subtracted from the ratings for the pairs containing different words (i.e., the experimental pairs) to get a perceptual salience rating for each phonological pattern to be studied. The perceptual salience averages were ranked for each pattern and resulted in the following groupings. The more salient features were (a) final consonant deletion of /k/ (FDk), M = 1.88, SD = .86; (b) final substitution of /t/ for /θ/ (t/th), M = 1.87, SD = .77; and (c) final consonant deletion of /t/ (FDt), M = 1.65, SD = .83. The less salient features were (a) final cluster reduction of /sk/ to /s/ (FRsk), M = 1.48, SD = .74; (b) final cluster reduction of /st/ to /s/ (FRst), M = 1.33, SD = .77; and (c) final substitution of /f/ for /θ/ (f/th), M = .35, SD = .63.

Description of the Sentences

The stimulus sentences consisted of eight words each (see Table 1). The sentences included both variable and fixed words. The fixed words created a consistent carrier phrase: *A ____ may be with my ______ Monday.* Variable words were placed in the two blanks; the pronunciation of *with* also varied among the different recordings (i.e., /wɪθ/, /wɪt/, and /wɪf/). The carrier phrase was used to create consistency for coarticulatory effects and to create the opportunity for the AAE phonological patterns represented in the low and high perceptual saliency groupings. Fixed words were selected that did not have codas; thus, there was no opportunity for any of the phonological patterns investigated in this study to occur.

Preparation of the Stimulus Recordings

The speakers were two 18-year-old African American females who were recruited from a university community. They exhibited the bidialectal skill needed to add or take away AAE features on cue. In a quiet room, a Logitech USB desktop microphone (Model #980186-0403) was placed approximately 6 in. from each speaker's mouth. The speech samples were digitized using Wavesurfer software (Sjolander & Beskow, 2004) at a sampling rate of 22,050 Hz.

The speakers were recorded reading the sentences three times without pronunciation instruction (*neutral readings*). The neutral readings were used to create the phonetically consistent carrier phrases into which variable words were inserted. Only the recordings of the lexically and phonologically consistent words were used from this reading (i.e., *a, may, be, my,* and *Monday*). Immediately following the neutral readings, the speakers read the sentences aloud again. This time, they were instructed to pronounce each word making sure that the final consonants were consistent with the word spellings (*SAE readings*). The SAE readings were also recorded three times. Finally, the speakers read the sentences again, making specific AAE phonological changes on the subject noun, the preposition *with,* and the object noun of the preposition (*AAE readings*). The AAE readings were also recorded three times.

In total, 54 sentence recordings were obtained from each speaker. The six different sentence wordings were repeated three times in three different ways (neutral, SAE, and AAE). All of the sentence recordings were spectrographically displayed using the Wavesurfer software (Sjolander & Beskow, 2004) that was used during the audio recording process. The first author modified the neutral readings from the two speakers by replacing the acoustic signal corresponding to the subject noun, the preposition *with,* and the object noun of the prepositional phrase with a corresponding acoustic signal in either the SAE readings or the AAE readings.

The 54 sentence recordings underwent two levels of verification to ensure that the recordings used were representative of AAE and were the most natural sounding. First, all recordings were reviewed by both investigators to determine if they were reasonable representations of the intended AAE phonological patterns. Second, six naturalness raters—3 men and 3 women (age range of 18–35) who were African American and AAE speakers—were recruited from the university community. They chose the most "natural sounding" recording from a choice of three for each stimulus sentence. See Robinson (2006) for a complete description of the naturalness rating procedures and results.

Randomization and Counterbalancing of Stimuli

The digital files for each recording selected during the naturalness rating process were embedded into slides of PowerPoint presentations. Four different randomized stimulus compilations were made. The sentences were randomized and counterbalanced to control for order effects related to speakers and tasks.

The order of the speakers varied across the compilations. Specifically, for the dialect detectability ratings, two compilations had 18 sentences read by Speaker 1 first, and two had 18 read by Speaker 2 first. For the comprehensibility ratings, two compilations had Speaker 1 first on every slide, and two had Speaker 2 first on every slide. Within each compilation, the speaker order was consistent between the two rating tasks. That is, if Speaker 1 was first in the first task, then she was also first in the second task.

The task order varied among the compilations in that two of the compilations had the participants rate comprehensibility first, and the remaining two had the participants rate dialect detectability first. To control for effects of sentence presentation order, the sentences read by each speaker were randomized for each presentation.

Data Collection

Presentation of stimuli. The experiment took place in a public school office space with individual computers at different cubicles. The participants attended to a PowerPoint presentation with embedded sound files. They viewed the presentations individually, and no more than 4 listeners participated at one time over 3 weeks. Each listener was randomly assigned to one of the four compilations such that each compilation was rated by 4 listeners. Each participant listened to the stimulus items through earphones and indicated his or her judgment on a computer-readable response form.

Rating tasks. Before each rating task, the participants viewed a set of instructional slides that oriented them to the tasks. Each rating task was preceded by a familiarity slide in which participants listened to a set of stimulus sentences with no AAE features present before listening to the various pronunciation patterns. This provided a consistent point of reference for the participants and oriented them to each speaker's voice before rating. Participants were also presented with the following context scenario for the sentences:

> Imagine that the sentences you are about to hear were spoken by the mother of a teen-ager. The mother is talking on the phone about her upcoming weekend vacation. She and her husband are leaving their teenage son at home, and they suspect that he is planning to have a wild party. She is very concerned that when they return on Monday, their house will be in shambles. She is envisioning what their home will look like when they return.

For all ratings, the orthographic transcription of the recordings was displayed with the rating scales. When making the comprehensibility ratings, the participants were asked to imagine that the two speakers were clients. The participants gave one rating to each group of six sentences when answering the following question: "How understandable do you think that these speakers would be to people in the general population?" A rating scale of A through E was presented under the question, with *very difficult to understand* located under the *A* and *very easy to understand* located under the *E*. If the participants felt that there was a difference between the two speakers, then they were asked to provide an average rating.

When making the dialect detectability ratings, the participants were informed that the speakers spoke a dialect called "African American English." They were told to indicate how noticeable use of the dialect was for each sentence using the following scale: A = *not noticeable at all,* E = *extremely noticeable.* They were further told to base their judgments not on the message of the sentence, but on how the sentence recording sounded. After rating the sentence recordings, the participants placed their responses in a sealed envelope so the examiner could not connect responses to participants.

Intrarater Reliability

Four participants (one from each compilation) were randomly asked to rate the same stimulus compilation that they had rated previously. The second rating occurred 2 weeks after the first presentation to determine intrarater reliability for both tasks.

Data Analysis Procedures

The response sheets were computer scanned, and the data generated by the two tasks were analyzed separately. The compilations were compared first using a one-way analysis of variance (ANOVA). This was followed by multiple comparisons using paired-samples *t* tests to determine the source of the differences among the different compilations when appropriate. Type I error rates were controlled using the Bonferroni method. Ratings for identical stimuli among the compilations were matched during the paired-samples *t* tests.

Three within-subject factors were included in the omnibus analyses of the dialect detectability data: (a) speaker (2 levels), (b) frequency (3 levels), and (c) salience (2 levels). Due to differences in the methods specific to the comprehensibility ratings, the speaker factor was not included. Therefore, the dialect detectability data were analyzed using a 2 × 3 × 2 ANOVA, and the comprehensibility data were analyzed using a 3 × 2 ANOVA. Both ANOVAs also included the between-subjects factor of *compilation*. Thomas (1998) showed ANOVA tests to be robust with Likert scale data. Therefore, the ANOVA was used in lieu of nonparametric analyses.

RESULTS

Reliability

A paired-samples correlation test was used to determine the association between the first and second stimulus presentation for the reliability participants. Time 1 and Time 2 ratings were highly correlated for dialect detectability ratings, $r(144) = .92$, $p < .001$, $r^2 = .85$, and comprehensibility ratings, $r(24) = .740$, $p < .001$, $r^2 = .55$. The effect sizes (r^2) accounted for more than half the variances for both tasks.

Compilation Comparisons

To determine where experimental differences existed in the data, if any, it was necessary to investigate whether aspects of the compilations themselves caused the ratings to differ. Therefore, the data associated with the compilations were compared for the different tasks. The four compilations did not elicit statistically different ratings for comprehensibility, $F(3, 92) = 2.462$, $p = .068$. Therefore, the six comprehensibility ratings generated from all 16 participants were pooled to create 96 total ratings, which were analyzed together. Thus, the between-subjects factor of compilation had four levels in the comprehensibility analysis ($N = 24$ for each level).

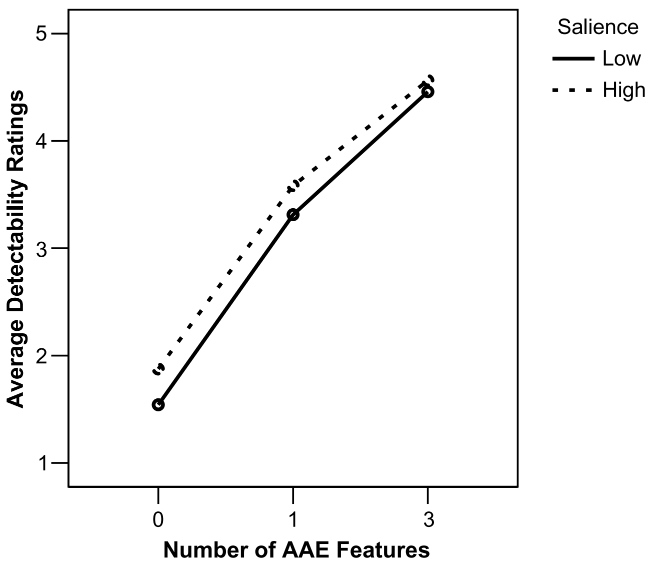
Some of the four compilations did elicit different ratings for dialect detectability, $F(3, 572) = 5.246$, $p = .001$, $\eta^2_p = .025$. Specifically, Compilations 1 and 2 ($N = 144$ ratings in each compilation) were significantly different, $M(\text{difference}) = .576$, $SD = 1.06$, $t(143) = -6.515$, $p < .001$, $d = .54$, and Compilations 1 and 4 ($N = 144$ ratings in each compilation) were significantly different, $M(\text{difference}) = .576$, $SD = 1.24$, $t(143) = -5.562$, $p < .001$, $d = .46$.

The data associated with the four compilations were grouped according to statistical similarity, which was determined by a hierarchical cluster analysis. The squared Euclidian distance between the average dialect detectability ratings created two groups. Compilations 2 and 4 received the highest ratings and comprised the *high detectability group*; Compilations 1 and 3 received the lowest ratings and comprised the *low detectability group*. These two groups of 288 ratings (36 ratings from 8 participants) were analyzed separately during the analysis of dialect detectability. In the separate analyses for these two data groupings, the between-subjects factor of stimulus compilation had two levels rather than four ($N = 144$ for each level).

Dialect Detectability Ratings

High detectability group results. As hypothesized, there was a significant main effect for frequency in the high detectability group, $F(2, 22) = 155.057$, $p < .001$, $\eta^2_p = .876$, indicating that as the number of AAE features increased, so did the dialect detectability ratings (see Figure 1). As hypothesized, the stimuli containing three AAE features were rated as more detectable than were the stimuli with just one feature, $t(95) = -9.356$, $p < .001$, $d = .96$, and stimuli containing one feature were significantly more

Figure 1. Average dialect detectability scores in the high detectability group obtained in relation to the frequency and salience of AAE features (N = 288).



Note. Dialect detectability rating scale: 1 = *not noticeable at all,* 5 = *extremely noticeable.*

AAE detectable than were those with no feature, $t(95) = -11.675$, $p < .001$, $d = 1.19$. Although both comparisons had large effect sizes, the effect of the 0–1 frequency difference was larger than that of the 1–3 frequency difference.

Although there was a significant main effect for salience, $F(1, 22) = 8.008$, $p = .01$, $\eta^2_p = .267$ (see Figure 1), the effect is qualified by a significant Salience × Speaker interaction, $F(1, 22) = 20.129$, $p < .001$, $\eta^2_p = .478$. Specifically, Speaker 2 elicited a significant salience effect, $t(71) = -4.512$, $p < .001$, with a medium effect size, $d = .53$, whereas Speaker 1 did not, $t(71) = .462$, $p = .645$.

Low detectability group results. The low detectability group had a significant main effect for salience, $F(1, 22) = 14.749$, $p = .001$, $\eta^2_p = .401$, and a significant main effect for frequency, $F(2, 44) = 159.135$, $p < .001$, $\eta^2_p = .879$. This frequency effect was qualified by a significant interaction with the variables *speaker,* $F(2, 44) = 6.306$, $p = .004$, $\eta^2_p = .223$, and *compilation,* $F(1, 22) = 6.709$, $p = .003$, $\eta^2_p = .234$. These interactions, however, did not affect the overall direction of the frequency effects. In all cases, increases in the frequency of AAE features resulted in significant increases in the dialect detectability scores. The differences between methodological variables concerned size rather than direction or significance of the effects.

Item Analyses

The post hoc item analysis compared the dialect detectability scores associated with the six individual AAE dialect features used in the detectability stimuli: (a) final consonant deletion of /k/, (b) final consonant deletion of /t/, (c) final cluster reduction of /k/, (d) final cluster reduction of /t/, (e) final t/θ substitution, and (f) final f/θ substitution. It aimed to answer the following two questions:

- What was the ranking of the dialect features with regard to each one's influence on dialect detectability ratings?
- Was *each* individual feature that was included in the high salience group more detectable than *each* of the features that was included in the low salience group?

Dialect detectability ranking. A mean *D* score was calculated for each feature. The mean *D* score was a comparison between the detectability ratings that were elicited by sentences containing just one feature and ratings that were elicited by sentences that had an unfulfilled opportunity for that same feature to occur (i.e., the control stimulus). This new calculation was called the *mean D score*. For example, the mean *D* score for final consonant deletion of /k/was obtained by subtracting the perceived dialect detectability ratings of sentences that had an unfulfilled opportunity for final consonant deletion of /k/ (and no other experimental features) from the ratings of sentences that had one instance of final consonant deletion of /k/ (and no other experimental features). The mean *D* score associated with each phonological feature was ranked. Paired-samples *t* tests matched for stimulus similarity were used to determine if the differences between the zero-frequency levels and the one-frequency levels for each dialect feature were significant. In addition to these values, the effect sizes for the differences between the two compared frequency levels for each dialect feature were determined and are presented in Table 3. The Bonferroni-adjusted alpha was .008. In this ranking of phonological features, it is particularly interesting to compare the top three and bottom three ranked features. The bottom three features all involved a /t/ (final cluster reduction of /t/, final t/θ substitution, and final consonant deletion of /t/). The top three features did not involve a /t/ (final consonant deletion of /k/, final f/θ substitution, and final cluster reduction of /k/).

Salience classification verification. The next analysis aimed to determine if the salience classification was accurate for each

Table 3. Ranking of features by mean *D* score (N = 32 for each feature).

Ranking	*Feature*	*Mean* D *score*	SD	t*(31)*	p	*Cohen's* d *value* (M/SD)
1	FDk	2.19	1.091	–11.346	<.001	2.01 (large)
2	f/th	1.94	1.014	–10.809	<.001	1.91 (large)
3	FRsk	1.66	.902	–10.388	<.001	1.84 (large)
4	FRst	1.06	.948	–6.338	<.001	1.12 (large)
5	t/th	1.56	1.480	–5.973	<.001	1.06 (large)
6	FDt	0.91	1.729	–2.964	.003	.52 (medium)

feature. The mean *D* scores for each high salience feature were matched to the mean *D* scores for a phonetically comparable low salience feature. The comparable features matched in this analysis were (a) final consonant deletion of /k/ (High)—final cluster reduction of /k/ (Low), (b) final consonant deletion of /t/ (High)—final cluster reduction of /t/ (Low), and (c) final t/θ substitution (High)—final f/θ substitution (Low). Paired-samples *t* tests were used to determine if there were significant detectability differences between the components of each pair (see Table 4). The Bonferroni-adjusted alpha was .02.

Only the final consonant deletion of /k/ and final cluster reduction of /k/ comparison reached significance. The other two comparisons, although not significant, tended in the opposite direction than was expected. In both cases, the features that were classified as low salience had higher mean *D* scores than did the phonetically comparable high salience features (cf. 1.94 and 1.56 for final f/θ substitution and final t/θ substitution, respectively).

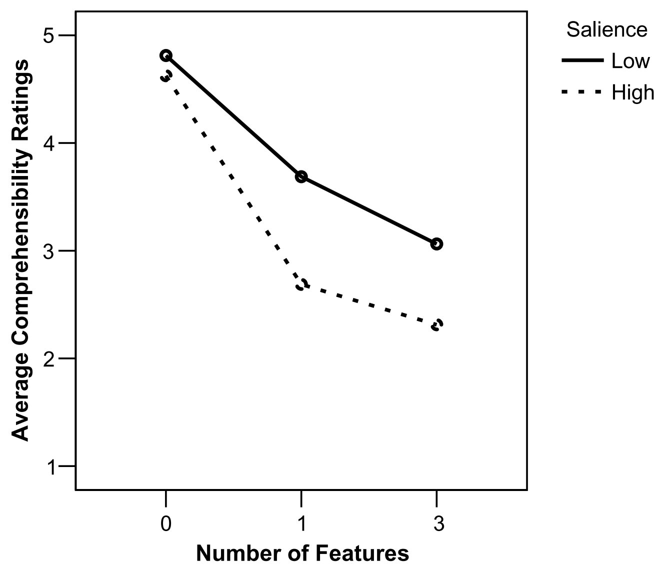
Comprehensibility Ratings

There was a main effect for frequency in the comprehensibility data, $F(2, 24) = 50.356$, $p < .001$, $\eta^2_p = .808$. Figure 2 shows that as the number of features increased, the perceived comprehensibility ratings decreased. A paired-samples *t* test was used to compare the stimuli associated with the different frequency levels. The Bonferroni-adjusted alpha was .025. The comparisons revealed that stimuli containing three features were rated significantly less comprehensible than were those with just one feature, $t(31) = 3.215$, $p < .001$, $d = .57$. The stimuli containing one feature were significantly less comprehensible than were those containing no feature, $t(31) = 8.530$, $p < .001$, $d = 1.51$. The effect size of the 0–1 difference was larger than the effect size of the 1–3 difference.

There was a significant main effect for salience, $F(1,12) = 35.593$, $p < .001$, $\eta^2_p = .748$. Figure 2 shows that those features that were classified as low salience elicited higher comprehensibility ratings than did those features that were classified as high salience, $t(47) = 5.511$, $p < .001$, $d = .86$.

The comprehensibility ratings revealed a significant Frequency × Salience interaction effect, $F(2, 24) = 4.926$, $p = .016$, $\eta^2_p = .291$, as shown in Figure 2. To analyze this interaction, paired-samples *t* tests were used to make the following comparisons: Pair (1) high salience/zero features with low salience/zero features; Pair (3) high salience/three features with low salience/three features. The

Figure 2. Average comprehensibility scores elicited from stimuli obtained in relation to the frequency and salience of AAE features ($N = 16$).



Note. Comprehensibility rating scale: 1 = *very difficult to understand,* 5 = *very easy to understand.*

Bonferroni-adjusted alpha was .02. There was no significant difference between the components of Pair 1, $t(15) = 1.861$, $p = .083$, indicating that the mere opportunity for high or low salience features was not enough to generate a salience effect. In Pair 2, a salience effect was elicited when one AAE feature was included, $t(15) = 4.472$, $p < .001$, $d = 1.12$. Specifically, the higher salience features generated lower ratings of comprehensibility. In Pair 3, the salience effect was maintained when three AAE features were included, $t(15) = 3.503$, $p = .003$, $d = .88$.

Written Comments Regarding the Study

After completing all of the ratings, the participants were asked to write “any feelings, comments, or questions about the study.” Six

Table 4. Comparison of mean *D* scores for comparable features in the high and low salience groups ($N = 32$).

Salience group	*Feature*	*Mean* D *score*	SD	*Difference*	SD	t*(31)*	p
High	FDk	2.19	1.091	.531	1.107	2.715	.010*
Low	FRsk	1.66	.902				
High	FDt	.91	1.721	–.156	1.936	–.456	.651 (ns)
Low	FRst	1.06	.948				
High	t/th	1.56	1.480	–.375	1.385	–1.531	.136 (ns)
Low	f/th	1.94	1.014				

Note. Mean *D* scores were derived by subtracting the detectability ratings for the zero-frequency level from the detectability ratings for the one-frequency level.

*Significant at $p < .02$ (.05 alpha adjusted per the Bonferroni method).

of the sixteen (37.5%) did so. Specific written responses will be addressed when appropriate in the Discussion section.

DISCUSSION

Perceptual Contributions of Frequency and Perceptual Salience

This study aimed to provide a better understanding of how AAE use is perceived by Caucasian American SLPs who are unfamiliar with the dialect. Their perceptions were of interest because of their high degree of professional power in determining who is a typically developing speaker versus one with a speech-language impairment. The stimuli were varied along two dimensions: frequency and perceptual salience.

Frequency of dialect features. This study supported the hypothesis that an increased number of AAE dialect features in an utterance will elicit increased ratings of dialect detectability and decreased ratings of comprehensibility from non-AAE-speaking SLPs. In other words, if many AAE features (of any type) are used within sentences and the person's speech is evaluated by an SLP with little exposure to the dialect, the speaker will likely be identified as a highly detectable AAE speaker with compromised comprehensibility by the general population. The threshold for a significant change in the rating of either dialect detectability or comprehensibility was a single feature. That is, it only took one feature in an eight-word sentence to increase dialect detectability ratings and decrease the perceived comprehensibility of the sentence. A significant although not as large of a difference was observed between one feature and three features. We conclude, therefore, that the mere *presence* of a single detectable AAE feature may have greater perceptual weight than does the cumulative *number* of AAE features present.

Perceptual salience of dialect features. Based on the expectations of Steriade (2004), perceptual salience was projected to influence both the comprehensibility and dialect ratings alike. In the current study, perceptual salience did not appear to influence the ratings that were elicited from the two tasks in the same way. This finding suggests that judging comprehensibility and dialect detectability are distinct processes.

Comprehensibility judgments. Perceptual salience appeared to accurately classify the features for the comprehensibility ratings (i.e., the group of features with high perceptual salience elicited lower comprehensibility ratings than did the features with low perceptual salience). However, given that the sentences were rated as a group for comprehensibility, a true feature-by-feature comparison could not be made. Nevertheless, the data suggested that SLPs may have considered the perceptual salience of the features in determining which sentences would be comprehensible to the general population. The orthographic transcription of the auditory stimuli could have aided these results.

Recall that the SLPs were presented with an orthographic transcription that corresponded to the auditory stimuli when rating the projected comprehensibility of the sentences. This was done because SLP listeners in this study may have been less linguistically naïve than listeners from the general population and may have participated in perception studies involving the speech of second language learners. SLPs have specific training in phonetics and phonology, which the participants in the other studies would not have had. This additional knowledge created several questions regarding what the SLPs would base their judgments on. Therefore, some added stability was required. However, the stability that was added was not unlike the inherent stability in the SLPs' actual practice. Specifically, it was similar to the relational analysis (comparing the client's productions to an intended target or model) that many of them would use in assessing the phonological patterns of their clients.

The use of the orthographic transcription in the current study created a need to modify the procedures that have been used in many perceptual studies of second language accent. That is, the SLPs were not asked to base their comprehensibility ratings on whether or not they themselves understood the sentence, but on whether or not the people speaking would be understandable in the "general population." Presumably, this modification required each SLP to base her judgments on a self-determined paradigm—perhaps the very paradigm that she would have used with her clients. In fact, one participant remarked that "final consonant deletion is always difficult to understand." Such a statement reveals a portion of the paradigm on which she based her ratings. Indeed, one popular phonological disorders assessment, the Hodson Assessment of Phonological Processes—Third Edition (HAPP–3; Hodson, 2004), is scored with the assumption that final consonant deletions will have the greatest impact on intelligibility. It is possible that the paradigms similar to the one used in the HAPP–3 were used to make comprehensibility judgments in the current study.

Dialect detectability judgments. The category of perceptual salience did not explain the dialect detectability scores as consistently as they did the comprehensibility scores. As expected, the high salience features tended to elicit higher dialect detectability scores than did the low salience features overall. However, when the dialect detectability ratings were examined for each feature individually, their ranking suggested that perceptual salience did not entirely explain the ratings (refer to Table 3). In fact, the feature that received the lowest perceptual salience ratings in nonsense words (f/th substitution) elicited the second highest dialect detectability ratings in the main study.

Although comprehensibility could be determined without any specific knowledge of common AAE features, the dialect detectability task required some awareness of phonological features associated with the AAE dialect. One participant commented on her own lack of AAE awareness, "It was a different experience to think about it as 'African American English' and not 'we've got to enroll this kid!'" In fact, the SLP participants were selected based on their *lack* of familiarity with AAE. If they only heard AAE spoken infrequently, then those features used most commonly by AAE speakers in the general population may have been rated highest in detectability. Irvine (2001) described features such as these as *iconic*. For example, if SLPs view the f/th substitution as particularly characteristic of AAE, then this one feature might elicit high detectability ratings, regardless of its low perceptual salience in nonsense words.

Perceptual salience did seem to affect the dialect detectability ratings. Final consonant deletion of /k/ received the highest mean *D* score, and it was also judged to be the most perceptually salient in the nonsense words. This feature appears to be less common in AAE than is final /t/ deletion. It was identified less frequently than th-substitutions and cluster reduction in the study by Craig et al. (2003). Stockman (2006) also reported that final /k/ deletion occurred significantly less frequently than did final /t/ deletion in language samples of young AAE-speaking children.

It appears that SLPs' perceptual judgments were related to a combination of factors: (a) the ubiquitous use of a feature by AAE speakers and (b) the perceptual salience of a given feature. That is, those features that are commonly used by AAE speakers received inflated detectability ratings due to their heightened iconicity with the dialect for the listeners. In this case, the f/th substitution pattern as the highest ranked of the low salience features might be the most iconic. In contrast, the perceptual salience of the deleted final /k/ was so high that it did not matter whether the SLPs associated it with AAE or not. Its perceptual salience in our experimental stimuli was enough to increase dialect detectability ratings and produce an overall perceptual salience effect for much of the dialect detectability data.

As for the three features receiving the lowest dialect detectability ratings, they all involved a final /t/ (t/th substitution, deletion of /t/, and final cluster reduction of /st/). Final /t/ is produced variably in the final position of words for many dialects of English. These variations include reduction to a glottal stop or omission in many dialects of English, even among its standard varieties (Guy, 1980; Stockman, 2006). It is regularly deleted in the final position of words when occurring in a consonant cluster (i.e. fast → /fæs/) in standard varieties of American English, particularly when a consonant follows, as it did in this study. Taking these facts into account, it is understandable that features involving final /t/ would receive lower dialect detectability ratings than different features involving different consonants. It is possible that when a feature involved a final /t/, the listeners had a wider range of acceptance due to the great variability of final /t/ production among speakers of many dialects of English.

Practical Implications

Many aspects of this study were designed to replicate in a controlled, experimental way the scenario of an AAE-speaking student being evaluated by an SLP with little exposure to AAE. The study is limited in its practical application in two ways: (a) It did not involve natural, spontaneous speech, and (b) it did not include school-age children. Nonetheless, its results are generally relevant to the basic practical question of who should be labeled as possessing a speech impairment. This question is relevant not only to school-based SLPs, but also to those who diagnose speech delay among AAE speakers in other professional work settings. At issue is whether the notions of dialect detectibility and comprehensibility should and do influence professional judgments concerning nonmainstream speakers. Although the recent revision of IDEA (IDEA, 2004, Sec. 300.8(c) (11)) defined a "speech impairment" in a way that does not implicate these terms directly, a "child's educational performance" may be adversely affected by the way school professionals perceive his or her dialect. If professionals judge a student's dialect as highly noticeable and not easily understandable, then these very factors could place an emotional burden on the student. All three of these elements have long been implicit aspects of a speech-impairment diagnosis by SLPs (Van Riper, 1978).

The current study's findings reinforce the expectation that typically developing AAE speakers may be differentially at risk for a diagnosis of speech delay. The likelihood of misdiagnosis appears to increase as the number and/or perceptual salience of dialect features used increases, particularly when SLPs are unfamiliar with the dialect. This is because such speech patterns are perceptually detectible (i.e., call attention to themselves), and the cumulative perceptual effect of multiple dialect patterns in a given stretch of speech is likely to be less comprehensible to the unfamiliar than the familiar listener. Thus, this study suggests that speech perception variables may play an inescapable role in SLPs' judgments about speakers. To counter natural perceptual tendencies to judge nonmainstream speakers as different enough to warrant a clinical diagnosis, clinicians should be empowered with two types of information: (a) social and cultural aspects of communication, generally (Payne & Taylor, 2005), and (b) the speech community that a speaker may represent, specifically. Payne and Taylor suggested that alternative, less biased assessment procedures may be used to rule out a communication disorder (IDEA, 2004, Sec. 300.8(c) (11)) as the chief reason for a student's failing educational performance, and thereby create a more effective and appropriate solution to the student's educational problems.

Implications for Future Research

The current study is one of the first to examine factors that might influence how a particular group of SLPs perceives the AAE dialect. Its investigative framework could be applied in future research to determine if the same perceptual variables of dialect detectability and comprehensibility are related to SLPs' perception of other nonmainstream English dialects. Future research also may extend what we know by considering four factors when designing such studies: (a) listener variation, (b) stimulus variation, (c) task variation, and (d) speaker variation, as discussed below.

Listener variation. The listeners in this study were Caucasian American SLPs from Michigan with little prior exposure to AAE. It is unknown how well the results would generalize to other listeners, such as SLPs who speak AAE or other stigmatized English varieties and SLPs who do not speak AAE but are familiar with it.

Stimulus variation. The speech stimuli used in this study included just a few commonly cited AAE features in the final position of words. Further research should focus on if and how other dialect patterns influence perceptual judgments of comprehensibility and dialect detectability. Other features to be explored might include d/ð substitutions in the initial position of words or derhoticization of final /ɚ/ (Pollock et al., 1998).

Furthermore, future studies may yield different results by using a different presentation modality than the strictly auditorally presented stimuli that were used in the current study. For example, the t/th substitution pattern was judged to have higher perceptual salience than the f/th substitution in nonsense words. The relative perceptual salience of these features may be diminished with videotaped stimuli. The t/th substitution, for example, involves only tongue tip variation and has subtle visual differences. The f/th substitution, on the other hand, involves variation of lip movement, which creates a more visually salient difference. Given the fact that SLPs generally assess their clients face-to-face, the addition of the visual modality would be an important modification of subsequent studies.

Task variation. The current study examined SLPs' perceptions of how understandable *others* might find the stimuli as a measure of comprehensibility. Future research should focus on how SLPs would judge the comprehensibility or intelligibility of AAE speech samples from their own perspectives as opposed to those of other people. Information gained by such studies could tap directly into the perceptual experiences of SLPs; however, caution should be taken in interpreting the data. Given the proliferation of information about assessing AAE speakers, SLPs may be savvy about the legitimacy of AAE and reluctant to admit to personal biases that conflict with the best practice policies for assessing AAE, as outlined by ASHA (1983, 1985). Therefore, more objective measures of intelligibility could be used to more directly access the personal perceptual experiences of SLPs with the AAE dialect.

Speaker variation. The stimuli in this study were spoken by two adult female AAE speakers. Future studies may focus on the perception of naturalistic speech from younger students as well as the perception of male voices, which could yield different results than were obtained in this study.

CONCLUSION

The data from this study suggest that AAE-inexperienced SLPs will judge speakers who use more AAE features in a linguistic unit as more AAE detectable and less comprehensible to the general population than those who use fewer features. This study was among the first to examine SLPs' perceptions of dialect use. More research needs to be done to further reveal the nature of dialect density perception by SLPs and other professionals who have the power to evaluate the communicative adequacy of individuals.

ACKNOWLEDGMENTS

This study would not have been possible without the generous support of several individuals. First, the project was supported by a Graduate Student Dissertation Completion Fellowship from Michigan State University. Furthermore, we would like to thank Dennis Preston, Peter Lapine, and Brad Rakerd for their assistance in designing the study and reviewing the work. Additionally, we would like to thank the following individuals for assisting in participant recruitment, stimulus construction, and data collection: Monica Clark-Robinson, Johanna Boult, Derrick Boult, Cheryl Granzo, Mary Jo Hidecker, and Diane Ogiela.

REFERENCES

Adger, C. T., Wolfram, W., & Christian, D. (2007). *Dialects in schools and communities* (2nd ed.). Mahwah, NJ: Erlbaum.

American Speech-Language-Hearing Association. (1983). Social dialects [Position statement]. Available from www.asha.org/policy.

American Speech-Language-Hearing Association. (1985). *Clinical management of communicatively handicapped minority language populations* [Position statement]. Available from www.asha.org/policy.

American Speech-Language-Hearing Association. (2005). *Highlights and trends: ASHA member counts for 2005*. Retrieved February 25, 2006, from http://www.asha.org/members/research/reports/member-counts.htm.

Annie, E. Casey Foundation. (2003). *KIDS count data*. Baltimore: Author.

Bailey, G., & Thomas, E. R. (1998). Some aspects of African American Vernacular English phonology. In S. Mufwene, J. Rickford, G. Bailey, & J. Baugh (Eds.), *African American English* (pp. 85–109). London: Routledge.

Baugh, J. (2000). *Beyond Ebonics: Linguistic pride and racial prejudice*. Boston: Oxford University Press.

Brennan, E., & Brennan, J. (1981a). Accent scaling and language attitudes: Reaction to Mexican-American English speech. *Language and Speech, 24,* 207–221.

Brennan, E., & Brennan, J. (1981b). Measurements of accent and attitude toward Mexican American speech. *Journal of Psycholinguistic Research, 10,* 487–501.

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. (2003). Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills, 97*(1), 11–20.

Craig, H. K., Thompson, C. A., Washington, J. A., & Potter, S. L. (2003). Phonological features of child African American English. *Journal of Speech, Language, and Hearing Research, 46*(3), 623–635.

Craig, H. K., & Washington, J. A. (2004). Grade-related changes in the production of African American English. *Journal of Speech, Language, and Hearing Research, 47*(2), 450–463.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustic Society of America, 91*(1), 370–389.

Gass, S., & Varonis, M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning, 34,* 65–90.

Graff, D., Labov, W., & Harris, W. (1983). Testing listeners' reactions to phonological markers of ethnic identity: A new method for sociolinguistic research. In D. Sankoff (Ed.), *Diversity and diachrony* (pp. 45–58). Amsterdam/Philadelphia: Benjamins.

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York: Academic Press.

Hodson, B. W. (2004). *Hodson Assessment of Phonological Patterns—Third Edition*. Austin, TX: Pro-Ed.

Individuals With Disabilites Education Improvement Act of 2004, 20 U.S.C., §1400 *et seq*.

Irvine, J. T. (2001). "Style" as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 21–43). Cambridge, England: Cambridge University Press.

Johansson, S. (1978). *Studies of error gravity—Native reactions to errors produced by Swedish learners of English* (Gothenburg Studies in English No 44). Gothenburg, Sweden: Department of English, University of Gothenburg.

Klein, E. S., & Flint, C. B. (2006). Measurement of intelligibility in disordered speech. *Language, Speech, and Hearing Services in Schools, 37,* 191–199.

Labov, W. (1966). *The social stratification of English in New York City.* Philadelphia: The Center for Applied Linguistics.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*(Suppl. 1), 285–310.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(Pt 3), 289–306.

Oetting, J. B., & McDonald, J. L. (2002). Methods for characterizing participants' nonmainstream dialect use in child language research. *Journal of Speech, Language, and Hearing Research, 45*(3), 505–518.

Payne, K. T., & Taylor, O. L. (2005). Multicultural influences in human communication disorders. In N. B. Anderson & G. H. Shames (Eds.), *Human communication disorders* (7th ed., pp. 65–78). Boston: Pearson.

Pollock, K. E., Bailey, G., Berni, M. C., Fletcher, J. M., Hinton, L. N., Johnson, I., & Weaver, R. (1998). *Phonological features of African American Vernacular English (AAVE)*. Retrieved October 24, 2004, from http://www.rehabmed.ualberta.ca/spa/phonology/features.htm.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*(1), 10–30.

Ramirez, J. D. (2005). *Ebonics: The urban educational debate*. Clevedon, UK: Multilingual Matters.

Robinson, G. C. (2006). *Perceptions of African American English dialect density by Anglo-European American speech-language pathologists*. Unpublished doctoral dissertation, Michigan State University, Lansing, MI.

Ryan, E., Carranza, M., & Moffie, R. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish–English bilinguals. *Language and Speech, 20,* 267–273.

Sjolander, K., & Beskow, J. (2004). Wavesurfer (Version 1.7.1) [Computer software]. Stockholm, Sweden: Centre for Speech Technology.

Steriade, D. (2004). *The phonology of perceptibility effects: The P-map and its consequences for constraint organization*. Retrieved November 20, 2004, from www.linguistics.ucla.edu/people/steriade/papers/P-map_for_phonology.doc.

Stockman, I. J. (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27,* 355–366.

Stockman, I. J. (2004). Clinical phonology issues in serving speakers of African-American Vernacular English. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 158–160). Cambridge, MA: MIT Press.

Stockman, I. J. (2006). Alveolar bias in the final consonant deletion patterns of African American children. *Language, Speech, and Hearing Services in Schools, 37,* 85–95.

Stockman, I., Guillory, B., Newkirk, B., & Seibert, M. (in preparation). *Relationship between standardized language test scores and dialect density in African American children*.

Thomas, J. E. (1998). *Properties of the ANOVA F Test as applied to ordered categorical data: A simulation-based study.* Unpublished doctoral dissertation, Baylor University, Waco, TX.

Thompson, C. A., Craig, H. K., & Washington, J. A. (2004). Variable production of AAE across oracy and literacy contexts. *Language, Speech, and Hearing Services in Schools, 35,* 269–282.

U.S. Department of Special Education Programs. (2004). *IDEAdata.org*. Retrieved October 10, 2004, from http://www.ideadata.org/index.html.

Van Riper, C. (1978). *Speech correction*. Englewood Cliffs, NJ: Prentice Hall.

Washington, J. A., & Craig, H. K. (1994). Dialectal forms during discourse of poor, urban, African American preschoolers. *Journal of Speech and Hearing Research, 37*(4), 816–823.

Received August 10, 2007
Accepted March 1, 2008
DOI: 10.1044/0161-1461(2008/07-0063)

Contact author: Gregory Robinson, University of Arkansas at Little Rock, University of Arkansas Medical Sciences, Audiology and Speech Pathology Department, Little Rock, AR 72204. E-mail: gcrobinson@ualr.edu.