1. INTRODUCTION

Over the last decade, various methods for obtaining speech recordings remotely for speech science (Freeman and De Decker, 2021; Sanker et al., 2021; Zhang et al., 2021), language documentation (Bird et al., 2014; Hilton, 2021), sociolinguistic (De Decker and Nycz, 2011; Hall-Lew and Boyd, 2017; Kim et al., 2019; Leemann et al., 2020), clinical (Vogel et al., 2014), and pedagogical (Wanjema et al., 2013) purposes have been proposed and evaluated. These remote speech recording methods include direct audio recording through an internet browser (Leemann et al., 2020; Wanjema et al., 2013), direct audio recording in video conferencing applications (Freeman and De Decker, 2021; Sanker et al., 2021; Zhang et al., 2021), direct audio recording through cell phone applications (Bird et al., 2014; Hilton, 2021; Leemann et al., 2020), and local audio recordings on a computer, tablet, or cell phone, which are then transferred to the researcher (De Decker and Nycz, 2011; Kim et al., 2019; Sanker et al., 2021; Vogel et al., 2014; Zhang et al., 2021).

The primary advantage of remote collection of speech recordings is the access it affords to diverse populations, who may not be willing or able to come to a campus laboratory for a recording session (Bird et al., 2014; Kim et al., 2019). The primary disadvantage of remote collection of speech recordings is the variable recording quality of the collected speech, due to variability in both hardware (e.g., cell phone vs. computer, internal vs. external microphone) and software (e.g., compression algorithms, noise-cancelling algorithms; De Decker and Nycz, 2011; Freeman and De Decker, 2021; Sanker et al., 2021; Vogel et al., 2014; Zhang et al., 2021). Controlling the recording software to limit this source of variability across talkers is straightforward (e.g., using Zoom for all recordings); controlling the hardware is more challenging, although talkers can be required to use a particular type of device (e.g., a computer or a cell phone) and to report their device information for inclusion in statistical analyses (Freeman and De Decker, 2021; Leemann et al., 2020). However, even when the hardware and software can be controlled to enhance the potential for obtaining high-quality recordings that are comparable across talkers, these methods remain susceptible to variation due to the talkers' internet connection strength and their ability to control or limit sources of background noise (Freeman and De Decker, 2021; Kim et al., 2019; Sanker et al., 2021; Zhang et al., 2021).

The goal of the current study was to collect a corpus of regionally-diverse adult American English speech for use as stimulus materials in future studies. We expected that the benefits of remote data collection for accessing a regionally-diverse sample of talkers would outweigh the costs of the variable recording quality that we expected would be present in the materials. The final Stories and Words Online Regional Dialect (SWORD) corpus includes read words, nonwords, and short stories produced by 31 adults with one of four target American English dialects.

2. CORPUS DESIGN

The SWORD corpus was designed to include speech from three authentic regional dialects of American English, as well as one novel dialect of American English created specifically for the corpus. For each of the four target dialects, one vowel contrast was identified as a characteristic feature of the dialect. The speech materials were then selected to contain these four target vowel contrasts for all talkers.

A. TARGET DIALECTS AND LINGUISTIC VARIABLES

I. AUTHENTIC DIALECTS

The three authentic regional dialects are New England, Northern, and Southern American English, as defined by Labov et al. (2006) and shown on the map in Fig. 1. One characteristic feature of New England American English is non-rhoticity (Labov et al., 2006), so the target vowel contrast for the New England dialect was /ɑɪ ɑ/ (e.g., *card, cod*). Given non-rhoticity in New England, we expected these vowels to be more acoustically similar for New England talkers than for talkers from other regions. Northern American English is characterized by the Northern Cities vowel shift (Labov et al., 2006), including the raising and fronting of /æ/ and the lowering and backing of /ɛ/, so the target vowel contrast for the Northern dialect was /æ ɛ/ (e.g., *mass, mess*). Given the Northern Cities vowel shift, we expected these vowels to be more acoustically similar for Northern talkers than for talkers from other regions. One characteristic feature of Southern American English is /ɑj/ monophthongization (Labov et al., 2006), so the target vowel contrast for the Southern dialect was /ɑj ɑ/

(e.g., *side, sod*). Given /ɑj/ monophthongization in the South, we expected these vowels to be acoustically more similar for Southern talkers than for talkers from other regions.

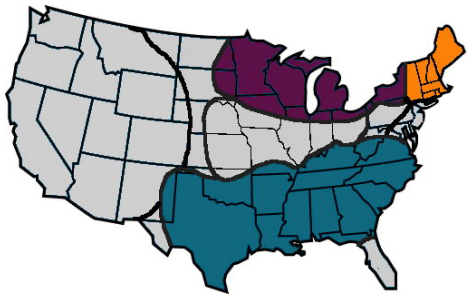

Figure 1. Map of the three authentic regional dialects in the SWORD corpus.

II. NOVEL DIALECT

The novel dialect was created to include an unfamiliar English vowel variant that would result in greater acoustic similarity for the target vowel contrast in the novel dialect than in the authentic dialects, in parallel to the target vowel contrasts selected for the authentic dialects. The target vowel contrast for the novel dialect was /ow ɔ/ (e.g., *boat, bought*). The acoustic similarity between these two vowels was increased in the novel dialect through monophthongization of /ow/. Specifically, Spanish-English bilingual linguists served as talkers for the novel dialect and were asked to produce Spanish /o/ in place of English /ow/ in all of the speech materials. All other consonants and vowels were produced in their native American English variety. The resulting novel dialect therefore sounds like an unfamiliar (i.e., novel) dialect of English.

B. SPEECH MATERIALS

I. WORDS AND NONWORDS

The first component of the SWORD corpus is a large set of isolated word and nonword utterances containing the target vowels. For each of the four vowel contrasts, we selected 24 minimal pairs (48 words per contrast). These critical minimal pair words are mostly monosyllabic, with varying syllable and morphological structures. For example, for the /æ ɛ/ contrast, we included the minimal pair *mass* and *mess* in the critical word list. For each of the four vowel contrasts, we also selected 36 words containing each vowel in the contrast (72 words per contrast) that do not have a minimal pair in English with the other vowel in the contrast. These filler words are also mostly monosyllabic, with varying syllable and morphological structures. For example, for the /æ ɛ/ contrast, we included the word *brass*, where **bress* is not a real word in English, and the word *crest*, where **crast* is not a real word in English, in the filler word list. All of the critical and filler words had a familiarity rating of at least 5.5 out of 7 in the Hoosier Mental Lexicon (Nusbaum et al., 1984). Finally, for each of the four vowel contrasts, we created 16 nonwords containing each vowel in the contrast (32 nonwords per contrast). The nonwords were created so that the intended spelling-to-sound correspondence would be clear to the talkers and so that neither the target nonword nor its minimal pair with the other vowel in the contrast would be a real word in English. These nonwords are all monosyllabic, with varying syllable and (apparent) morphological structures. For example, for the /æ ɛ/ contrast, we included the nonword **plass*, where **pless* is also not a real word in English, and the nonword **trest*, whereas **trast* is also not a real word in English, in the nonword list. A summary of the word and nonword materials is shown in Table 1.

Table 1. Summary of the word and nonword materials in the SWORD corpus.

Item type	Examples	Count per vowel contrast	Total count
Critical words	mass, mess	48	192
Filler words	brass, crest	72	288
Nonwords	plass, trest	32	128
Total		152	608

II. SHORT STORIES

The second component of the SWORD corpus is a set of read short stories. One familiar fairy tale was selected for each vowel contrast and edited so that each vowel within the target vowel contrast appeared 40 times within the story. The selected stories were *Goldilocks and the Three Bears* for the /ɑɹ ɑ/ contrast, *Little Red Riding Hood* for the /æ ɛ/ contrast, *The Pied Piper of Hamelin* for the /ɑj ɑ/ contrast, and *The Adventures of Pinocchio* for the /ow ɔ/ contrast. The short stories were 408-439 words long. The recordings of the short stories in the corpus range in length from 113 s to 181 s, with a mean of 142 s.

3. CORPUS COLLECTION

A. TALKERS

I. AUTHENTIC DIALECTS

The authentic dialect talkers were recruited using Prolific and paid for their participation. For each target dialect, potential talkers were screened based on their current state of residence, as provided in their Prolific profile. New England talkers were recruited from Connecticut, New Hampshire, Maine, Massachusetts, Rhode Island, and Vermont. Northern talkers were recruited from Illinois, Indiana, Iowa, Michigan, Minnesota, New York, Ohio, and Wisconsin. Southern talkers were recruited from Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia. A total of 85 talkers completed the study on Prolific.

Talkers were asked to self-report their age, gender, race/ethnicity, native language, residential history (city and state), and any history of speech, language, or hearing disorders. Talkers who did not report English as one of their native languages (N=1) and talkers who reported a history of a speech, language, or hearing disorder (N=1) were excluded from the corpus. Self-reported residential history was used to confirm that each talker included in the corpus is an authentic talker of their dialect region. Talkers were considered an authentic talker of their dialect region if they lived only in that region from birth to at least age 18 years. Talkers who did not meet this definition of an authentic talker of one of the three target dialects (N=19) or who did not provide sufficient demographic data to evaluate their language or residential history (N=20) were excluded from the corpus. As expected, the recording quality was highly variable across talkers. All recordings were therefore screened and talkers whose recordings were auditorily determined to have an unacceptable signal-to-noise ratio in the majority of their recordings (N=18) were excluded from the corpus.

The final set of 26 authentic dialect talkers included in the SWORD corpus all reported English as one of their native languages, reported no history of speech, language, or hearing disorders, and were identified as authentic talkers of their dialect region based on their childhood residential history. A summary of the self-reported age, gender, and race/ethnicity of the authentic dialect talkers in the corpus is shown in Table 2.

Table 2. Summary of the talker demographics in the SWORD corpus.

Dialect	Number of talkers	Age range	Gender	Race/ethnicity
New England	10	23-75 years	4 female 6 male	1 Asian 1 Black 8 white
Northern	8	22-62 years	5 female 2 male 1 non-binary	8 white
Southern	8	22-53 years	5 female 3 male	1 Black 1 Hispanic 1 more than one 5 white
Novel	5	22-34 years	2 female 2 male 1 non-binary	1 Hispanic 1 more than one 3 white

II. NOVEL DIALECT

The novel dialect talkers were Spanish-English bilingual linguists and were recruited through our professional networks. They received a gift card to Amazon or Target for their participation. Data from five novel dialect talkers are included in the SWORD corpus. The novel dialect talkers had varied residential histories: all five of them had lived in the Midland dialect region and one had also lived in the Northern dialect region, one had also lived in the Southern dialect region, and one had also lived in the Western dialect region. A summary of the self-reported age, gender, and race/ethnicity of the novel dialect talkers in the corpus is shown in Table 2.

B. RECORDING PROCEDURE

We anticipated considerable variability in the quality of the recordings that we would obtain from talkers on Prolific, due to variability in their recording hardware, internet connectivity, and ability to limit background noise in their environment (Freeman and De Decker, 2021; Kim et al., 2019; Sanker et al., 2021; Zhang et al., 2021). We also expected that Prolific users would be more likely to participate in a relatively shorter study overall. We therefore decided to obtain a subset of the full corpus materials from a larger set of authentic dialect talkers, rather than the full set of materials from a smaller set of authentic dialect talkers. From each authentic dialect talker, we recorded approximately 25% of the total set of corpus materials, with the selected materials weighted towards the vowel contrast that was selected to characterize their regional dialect. For the critical minimal pair words, each authentic dialect talker recorded 18 of the 24 minimal pairs for their vowel contrast and 6 of the 24 minimal pairs for each of the other three vowel contrasts, for a total of 36 minimal pairs (72 critical words). For the filler words, each authentic dialect talker recorded 9 of the 36 words containing each vowel for each of the four vowel contrasts, for a total of 72 filler words. For the nonwords, each authentic dialect talker recorded 8 of the 16 nonwords containing each vowel for their vowel contrast and for one other vowel contrast, for a total of 32 nonwords. The vowel contrasts were paired for the nonwords, so that the New England and Northern talkers produced /ɑɹ ɑ/ and /æ ɛ/ nonwords and the Southern talkers produced /ɑj ɑ/ and /ow ɔ/ nonwords. This pairing was selected so that no individual talker received both of the contrasts containing /ɑ/. Each authentic dialect talker produced the short story for their vowel contrast. Thus, the authentic dialect talkers each produced 144 words, 32 nonwords, and 1 short story. For each authentic dialect, four stimulus lists were created in a Latin Square design so that all materials were presented to talkers from all authentic dialect regions across the lists. Each authentic dialect talker was assigned one list. This stimulus list approach means that the exclusions from the corpus due to poor recording quality represent a much smaller total speech sample than excluding the full set of corpus materials from the same number of talkers. Thus, the time and resources that we and the authentic dialect talkers expended to collect data that we do not currently have plans to use was reduced relative to an approach in which we collected the full set of materials from every authentic dialect talker. This approach also allowed us to limit the time commitment of each individual authentic dialect talker to approximately 20 minutes. Given the specialized knowledge required for the novel dialect talkers, these talkers each produced all 480 words and 128 nonwords, as well as the short story for their vowel contrast.

Each authentic dialect talker was presented with three blocks of words and one block of nonwords, with the nonword block always presented after the three word blocks. Words and nonwords were fully randomized within blocks. The short story was presented after the nonword block. The demographic questionnaire was presented last. Talkers were permitted to take a self-timed break between each block. The target text (words, nonwords, short story) was presented in the talker's browser window in black font against a light blue background. Words and nonwords were displayed individually for 1500 ms with a 500 ms inter-stimulus interval. The short story was presented on a single page with appropriate formatting (e.g., line breaks) for the narrative. The short story reading was self-paced and the talkers were encouraged to read the entire story silently to themselves before reading it aloud. The procedure for the novel dialect talkers was the same as for the authentic dialect talkers, with two exceptions. First, the words and nonwords were blocked by vowel contrast and presented in the same fixed order to all talkers. Within each contrast, the words were presented before the nonwords. Second, the stimulus words, nonwords, and short story were provided to the novel dialect talkers in advance and they were encouraged to practice the novel dialect before starting the recording.

All of the authentic and novel dialect talkers were required to complete the study on a computer and were asked to report their operating system, browser, and microphone type (i.e., computer or laptop internal microphone, integrated microphone in earbuds/headphones, or external microphone). The data were collected using direct audio recording through the internet browser, based on the HTML, PHP, and JavaScript code developed by Wanjema et al. (2013). The audio recordings were saved directly to a secure departmental server

at a sampling rate of either 44.1 kHz or 48 kHz with 16-bit quantization. The variability in sampling rate reflects the native sampling rate of the individual talker's hardware and software. Most Windows users (15/17) had sampling rates of 48 kHz, whereas most MacOS users (10/14) had sampling rates of 44.1 kHz. The final corpus materials were segmented into individual sound files for each word, nonword, and short story and down-sampled to 22050 Hz with 16-bit quantization.

4. PRELIMINARY ACOUSTIC ANALYSIS

A preliminary acoustic analysis of the critical words was conducted to assess the success of the corpus in capturing the target regional dialect variation. We examined formant trajectory distances for the critical minimal pair words in each vowel contrast to assess our prediction that these distances would be smaller in the associated target dialect than in the other dialects. In particular, we predicted that the /ɑɹ ɑ/ distance would be smaller for the New England talkers than the other talkers, that the /æ ɛ/ distance would be smaller for the Northern talkers than the other talkers, that the /ɑj ɑ/ distance would be smaller for the Southern talkers than the other talkers, and that the /ow ɔ/ distance would be smaller for the novel dialect talkers than the other talkers.

A. ACOUSTIC ANALYSIS

The analysis comprised 72 minimal pair tokens produced by each of the 26 authentic dialect talkers and 192 minimal pair tokens produced by each of the five novel dialect talkers. Prior to the analysis, missing (N=11), misread (e.g., "soldier" for *solder*, N=22), and noisy (N=44) tokens, along with their minimal pairs (N=77), were excluded. The analysis therefore included a total of 2678 tokens, representing 1339 minimal pairs.

The target vowel in each token was segmented by hand, following the conventions described by Peterson and Lehiste (1960), except that coda /ɹ l/ were segmented with the target vowel to capture variation in rhoticity and /l/-vocalization across dialects. Onset approximants were segmented separately from the target vowel. The first three formant frequencies were estimated from each vowel token at 10% temporal intervals over the middle 80% of the vowel (i.e., at 10%, 20%, …, 90% of the vowel duration) using a 12^{th}-order LPC analysis over the frequency range of 0-5500 Hz. All formant estimates were converted to the Bark scale (Traunmüller, 1990) for analysis.

B. VOWEL DISTANCE

Vowel distance was defined as the root-mean-square distance (RMSD) of the Bark formant trajectories for each minimal pair (Cole et al., 2023; Heeringa et al., 2009), calculated separately for each formant for each talker. For each vowel contrast, we selected one formant as the focus of the analysis, based on the predicted dialect differences. For the /ɑɹ ɑ/ contrast, we selected F3 to capture variation in rhoticity across dialects. We predicted a smaller F3 RMSD for the New England talkers than for the other talkers, reflecting non-rhoticity in New England. For the /æ ɛ/ contrast, we selected F1 to capture variation in vowel height across dialects. We predicted a smaller F1 RMSD for the Northern talkers than for the other talkers, reflecting the Northern Cities vowel shift in the North. For the /ɑj ɑ/ contrast, we selected F2 to capture variation in the /ɑj/ offglide across dialects. We predicted a smaller F2 RMSD for the Southern talkers than for the other talkers, reflecting /ɑj/ monophthongization in the South. For the /ow ɔ/ contrast, we selected F1 to capture variation in the /ow/ offglide across dialects. We predicted a smaller F1 RMSD for the novel dialect talkers than for the authentic dialect talkers, reflecting /ow/ monophthongization in the novel dialect.

C. RESULTS AND DISCUSSION

The mean formant trajectory RMSDs are shown in Fig. 2 for the selected formant for each vowel contrast for each talker dialect. Larger differences are observed across dialects for the /ɑɹ ɑ/ and /ɑj ɑ/ contrasts than for the /æ ɛ/ and /ow ɔ/ contrasts overall. As expected, the New England talkers have the shortest F3 RMSD for the /ɑɹ ɑ/ contrast, the Northern talkers have the shortest F1 RMSD for the /æ ɛ/ contrast, and the Southern talkers have the shortest F2 RMSD for the /ɑj ɑ/ contrast. However, contrary to expectations, the Northern talkers have the shortest F1 RMSD for the /ow ɔ/ contrast, rather than the novel dialect talkers.

A linear mixed-effects model on the formant trajectory RMSDs with vowel contrast, talker dialect, and their interaction as fixed effects, as well as random by-talker and by-minimal-pair intercepts and a random by-talker slope for vowel contrast, revealed a significant main effect of vowel contrast ($F(3, 39.0) = 40.66$, $p < .001$), but no main effect of talker dialect and no interaction. Thus, some of the vowel contrasts are intrinsically more similar in the selected formant trajectories than others. In particular, the /æ ɛ/ contrast, which involves two short,

lax vowels has the shortest overall RMSD, whereas the /ɑj ɑ/ contrast, which involves one diphthong and one monophthong, has the longest overall RMSD. However, within each vowel contrast, the differences between talker dialects are not statistically robust. This lack of clear dialect effects on the formant trajectory differences reflects variability within and across the dialects, as well as the relatively small sample within each dialect. For example, auditory inspection of the corpus materials suggests that some, but not all, of the New England talkers are non-rhotic, and that some, but not all, of the Southern talkers are also non-rhotic. As a result of this variability, both the New England and the Southern dialects show relatively shorter mean F3 RMSDs for the /ɑɹ ɑ/ contrast in Fig. 2 than the Northern and novel dialects. This variability within and across dialects is consistent with documented dialect variation in the United States (Labov et al., 2006) and suggests that the SWORD corpus reflects both authentic cross-dialect variation and authentic within-dialect variation.

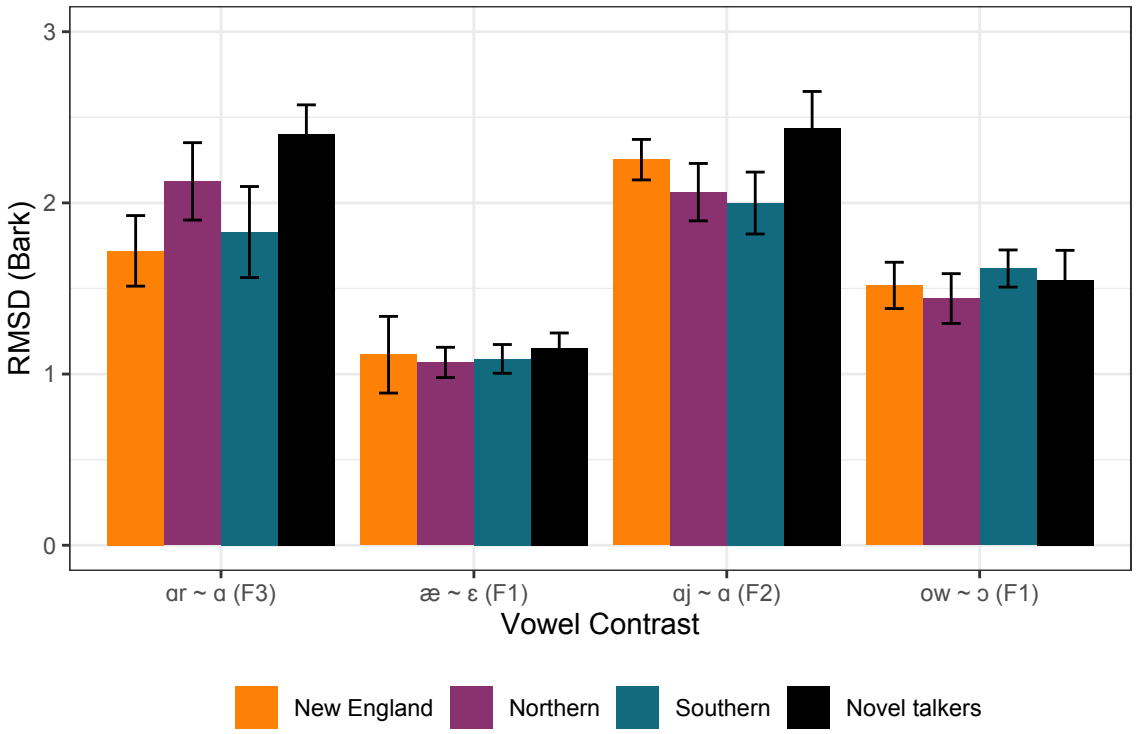

Figure 2. Selected formant trajectory RMSDs in Bark for each vowel contrast for each talker dialect in the SWORD corpus.

5. CONCLUSION

We recruited and recorded a regionally-diverse set of American English adult talkers online through Prolific for the authentic dialect component of the SWORD corpus. In addition to the target regional diversity, the talkers included in the SWORD corpus are more diverse in terms of age and race/ethnicity than corpora that we have previously collected in university labs in the American Midwest, including the Nationwide Speech Project corpus (Clopper and Pisoni, 2006) and the Ohio State Stories corpus (Burdin et al., 2015). Thus, as in previous studies employing online speech recordings (Bird et al., 2014; Kim et al., 2019), we were able to obtain data from a more diverse population than we typically have access to in our university lab. Our preliminary acoustic analysis of the materials further suggests that the recordings we collected reflect expected patterns of variation within and across regional dialects of American English (Labov et al., 2006), although our sample is too small to explore this variability in detail.

As expected, the recordings were also variable in overall quality, especially with respect to background noise (Hall-Lew and Boyd, 2017; Kim et al., 2019; Sanker et al., 2021). We excluded 21.2% of the talkers from the corpus for excessive background noise. However, we also excluded slightly larger percentages of talkers who either did not fully complete the demographic questionnaire (23.5%) or who did not meet our language and residential history requirements (24.7%). Thus, variable recording quality was roughly equivalent in cost to both

talkers' incomplete compliance with task instructions and the limited talker screening we employed in Prolific. Our overall inclusion rate was 30.6% and this rate could likely be improved with changes to our recruitment and data collection protocol to more effectively pre-screen potential talkers and encourage completion of all task components.

In summary, the final SWORD corpus demonstrates that online data collection can be a reasonable strategy for the collection of speech corpora from diverse populations when some degree of variability in recording quality can be tolerated. The SWORD corpus is available to the scholarly community for research and pedagogical projects: https://u.osu.edu/swordcorpus/.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (BCS-1843454). We would like to thank Margot Hare, Jonathan Kalala, and Kritsia Owens for research assistance.

REFERENCES

Bird, S., Hanke, F. R., Adams, O., and Lee, H. (**2014**). "Aikuma: A mobile app for collaborative language documentation," *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Language*, 1–5.

Burdin, R. S., Turnbull, R., and Clopper, C. G. (**2015**). "Interactions among lexical and discourse characteristics in vowel production," *Proc. Mtgs. Acoust.* **22**, 060005.

Clopper, C. G., and Pisoni, D. B. (**2006**). "The Nationwide Speech Project: A new corpus of American English dialects," *Speech Comm.* **48**, 633–644.

Cole, J., Steffman, J., Shattuck-Hufnagel, S., and Tilsen, S. (**2023**). "Hierarchical distinctions in the production and perception of nuclear tunes in American English," *Lab. Phonol.* **14**(1), 1–51.

De Decker, P., and Nycz, J. (**2011**). "For the record: Which digital media can be used for sociophonetic analysis?" *Univ. Penn. Working Papers Linguist.* **17**(2), 51–59.

Freeman, V., and De Decker, P. (**2021**). "Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps." *J. Acoust. Soc. Am.* **149**, 1211–1223.

Hall-Lew, L., and Boyd, Z. (**2017**). "Phonetic variation and self-recorded data," *Univ. Penn. Working Papers Linguist.* **23**(2), 86–95.

Heeringa, W., Johnson, K., and Gooskens, C. (**2009**). "Measuring Norwegian dialect distances using acoustic features," *Speech Comm.* **51**, 167–183.

Hilton, N. H. (**2021**). "Stimmen: A citizen science approach to minority language sociolinguistics," *Linguist. Vanguard* **7**(s1), 20190017.

Kim, C., Reddy, S., Stanford, J. N., Wyschogrod, E., and Grieve, J. (**2019**). "Bring on the crowd! Using online audio crowd-sourcing for large-scale New England dialectology and acoustic sociophonetics," *Am. Speech* **94**, 151–194.

Labov, W., Ash, S., and Boberg, C. (**2006**). *Atlas of North American English* (Mouton de Gruyter, New York).

Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., and Messerli, J. (**2020**). "Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing," *Linguist. Vanguard* **6**(s3), 20200061.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). "Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words," *Research on Speech Perception Progress Report No. 10* (Speech Research Laboratory, Indiana University, Bloomington, IN), pp. 357–376.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.

Sanker, C., Babinski, S., Burns, R., Evans, M., Johns, J., Kim, J., Smith, S., Weber, N., and Bowern, C. (**2021**). "(Don't) try this at home! The effects of recording devices and software on phonetic analysis," *Lang.* **97**, e360–e382.

Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," *J. Acoust. Soc. Am.* **88**, 97–100.

Vogel, A. P., Rosen, K. M., Morgan, A. T., and Reilly, S. (**2014**). "Comparability of modern recording devices for speech analysis: Smartphone, landline, laptop, and hard disc recorder," *Folia Phoniatr. Logop.* **66**, 244–250.

Wanjema, S., Carmichael, K., Walker, A., and Campbell-Kibler, K. (**2013**). "The OhioSpeaks project: Engaging undergraduates in sociolinguistic research," *Am. Speech* **88**, 223–235.

Zhang, C., Jepson, K., Lohfink, G., and Arvaniti, A. (**2021**). "Comparing acoustic analyses of speech data collected remotely," *J. Acoust. Soc. Am.* **149**, 3910–3916.