# Remote sociophonetic data collection: Vowels and nasalization from self-recordings on personal devices

**Article** *in* Language and Linguistics Compass · July 2021

DOI: 10.1111/lnc3.12435

CITATIONS
6

READS
144

**2 authors**, including:

Valerie Freeman
Oklahoma State University - Stillwater
**77** PUBLICATIONS **281** CITATIONS

SEE PROFILE

# Remote Sociophonetic Data Collection: Vowels and Nasalization from Self-recordings on Personal Devices

**Valerie Freeman[1] & Paul De Decker[2]**

*[1]Oklahoma State University & [2]Memorial University of Newfoundland*

valerie.freeman@okstate.edu & pauldd@mun.ca

**Abstract:**

When the COVID-19 pandemic halted in-person data collection, many linguists adopted modern technologies to replace traditional methods, including speaker-led options in which participants record themselves using their own personal computers or smartphones and then email or upload the sound files to online storage sites for researchers to retrieve later. This study evaluated the suitability of such "home-made" recordings for phonetic analysis of vowel space configurations, mergers, and nasalization by comparing simultaneous recordings from several popular personal devices (Macbook, PC laptop, iPad, iPhone, Android smartphone) to those taken from professional equipment (H4n field recorder, Focusrite with Audio Technica 2021 microphone). All personal devices conveyed vowel arrangements and nasalization patterns relatively faithfully (especially laptops), but absolute measurements varied, particularly for the female speaker and in the 750-1500 Hz range, which affected the locations (F1xF2) of low and back vowels and reduced nasalization measurements (A1-P0) for the female's pre-nasal vowels. Based on these results, we assess the validity of remote recording using these consumer devices and offer recommendations for best practices for collecting high fidelity acoustic phonetic data from a distance.

**Keywords:** sociophonetics, methods, pandemic, recording, vowels, formants, nasality

## I. INTRODUCTION

When the COVID-19 pandemic put a halt to in-person data collection, some linguistic researchers and other social scientists asked participants to record themselves on their personal mobile devices or home computers (Lupton, 2020). At that time, one of the authors was in the middle of collecting laboratory recordings to examine pre-lateral vowel mergers in Oklahoma, and the other was preparing an acoustic study of how nasalization interacts with the short-a (æ) split in Canada. Since then, dozens of Oklahoma merger participants have recorded themselves on phones and laptops. This study addresses whether such "home-made" recordings are suitable for phonetic analysis of vowel space configurations, mergers, and nasalization by comparing measurements taken from simultaneous recordings with several popular personal devices to those taken from professional equipment. Given the ubiquity of mobile devices, linguistic data collection is likely to continue expanding outside the lab and field interview going forward, necessitating an evaluation of the quality of self-recordings as a data collection method that has the capacity to include participants regardless of geographic location. The procedures are relevant to sociolinguists, language documentarians, speech therapists, and others who typically record speech outside the lab in order to prioritize naturalness, widen their access to speakers, or work with limited resources.

### A. Background

In 2007, Apple released the first iPhone in the United States and three years later began marketing the iPad tablet. The next few years were marked by a proliferation of mobile devices, and some linguists (e.g., De Decker & Nycz, 2011) wondered if those devices could be used for the remote collection of sociophonetic data. Research showed that two issues might impede such recordings. First, compression techniques that encode speech on a digital recording device tend to alter acoustic content, resulting in significant measurement or tracking errors (Bulgin, De Decker, &

Nycz, 2010; Van Son, 2005). Second, different makes and models of tablets, cellphones, and computers are built with significantly different microphone qualities, a problem for reliably quantifying acoustic parameters (Parsa, Jamieson, & Pretty, 2001).

Following the introduction of the iPhone, De Decker & Nycz (2011) tested a number of popular devices that held promise for recording spoken language data outside the constraints of a phonetics laboratory, making them arguably less susceptible to the effects of the Observer's Paradox (Labov, 1984). In their F1xF2 vowel space analysis, the authors found that compressed audio from an AVI video and the mp3 from a derived YouTube video significantly affected Praat's estimation of F1 and F2. The effects were more dramatic for their female speaker than for their male speaker. In comparison, the uncompressed m4a files made on a Macbook Pro laptop and iPhone did not contribute to the same distorted effects on a speaker's vowel space.

More recently, researchers have continued to consider the utility of mobile recording devices and their compression techniques for data collection (Bleaman & Duncan, 2016; Fuchs & Maxwell, 2016; Holland & Brandenburg, 2017; Kolly & Leemann, 2015). These studies advocated for what might be called a *pluralistic* view of audio collection techniques, using a variety of devices and archival sources, including recordings that are generated by speakers themselves, the focus of the present study.

In this paper we re-examined the reliability of self-recordings made with mass-marketed technology available over the past three years. We had in mind scenarios where speakers record themselves on their own personal device, such as a laptop computer, tablet, or mobile phone, and send the resulting files to researchers via email or cloud storage apps like Dropbox or Google Drive. The most common audio format for these files is m4a, which uses either the advanced audio coding (AAC) standard or Apple Lossless Audio Codec (ALAC)**.** This allows m4a recordings to be created

at high quality settings equivalent to lossless formats, allowing us to hypothesize that personal recordings should be comparable to those made on professional equipment -- or at least suitable for phonetic analysis.

While considering this hypothesis, we organized our study around a more general research question: How do recordings made on popular mobile devices compare to professional recordings in terms of vowel and nasality measurements? To test this, we first examined the locations and relative arrangements of vowels in each speaker's acoustic (F1xF2) vowel space, plotted in raw Hz to provide direct comparisons across devices. Second, we inspected patterns and amounts of overlap between pairs of pre-lateral vowels which may be involved in merger. Third, we compared patterns of nasalization in oral and pre-nasal /æ/ vowels by plotting spectral tilt (A1-P0) across vowel duration.

**B. Acoustic-phonetic features examined**

Plotting vowel arrangements in acoustic space is the most common method of comparing English dialects in sociophonetic studies of language variation and change, and so we examined it as our first measure. In these plots, the values of each vowel's lowest two formants (F1, F2), or frequency peaks, are plotted against each other to create a two-dimensional display resembling the traditional vowel quadrilateral representing vowel height and backness. As described below, we chose speakers from two dialect regions, which allowed us to test our hypotheses on examples of two different vowel space and merger configurations. We expected our male speaker (from Idaho and Arizona) to show the defining features of the American West, the low-back *cot-caught* merger and no Southern Shift or Northern Cities Shift (see Labov, Ash, & Boberg, 2006 for more on these mergers and shifts). Our female speaker is from Oklahoma, which is on the periphery of multiple dialect regions and shows a mix of features, including the Midland/West *cot-caught* merger and

Southern fronting of non-pre-lateral /u, o/ that we expected to see in our speaker's vowel space configuration (Bakos, 2013; Labov, et al., 2006). (Some parts of the Southern Shift can be found in older generations and the southern part of the state, but our speaker is young and from northern Oklahoma.)

Vowel mergers, or the loss of distinction over time, are a common way of distinguishing dialect regions and are typically adopted by some social groups before others, providing a snapshot of socio-regional change in progress. There are various mergers in progress around the country in which neighboring vowels are pronounced similarly before /l/, often with a tense vowel becoming lax, like front vowels *feel-fill* /il-ɪl/ and *sale-sell* /el-ɛl/ or high-back *pool-pull* /ul-ʊl/, but sometimes with a lax vowel becoming tense, like mid-back *bull-bowl* /ʊl-ol/, or with the central /ʌl/ backing to /ʊl/ *hull-bull*, /ol/ *hull-hole*, or /ɔl/ *hull-hall* (Labov, et al., 2006). These multiple possibilities allow us to see how well our devices capture merging behavior throughout the vowel space. Not much prior work covers prelateral mergers in depth, particularly the many possible configurations of *dull-pull-pole-pool* /ʌl, ʊl, ol, ul/, but near-mergers of high-back *pool-pull* /ul-ʊl/ and front *feel-fill* /il-ɪl/ and *sale-sell* /el-ɛl/ have been found in some Midland and West speakers, including Oklahomans (Bailey et al., 1996; Bakos, 2013; Labov et al., 2006).

While many sociophonetic studies of vowel quality have examined F1 x F2 properties, the harmonic structure can also be used to categorize vowels as nasalized or not nasalized. Here we looked at the effects of recording devices on the relationship between two harmonics, or frequency components, associated with anticipatory nasalization found in the low-front 'ash', (a.k.a. short-a, /æ/, or TRAP/BATH) lexical set. /æ/ was chosen because dialects of North American English exhibit /æ/-raising before nasals (Labov, Ash & Boberg 2006), a perceptual phenomenon achieved primarily through anticipatory nasalization, though some speakers have been found to use a

concomitant lingual-raising gesture (Carignan, Mielke & Dodsworth 2016, De Decker & Nycz 2012). Carignan (2021) identifies a number ways pre-nasal vowels are distinct from those in non-nasal environments, and following the work of Chen (1997) and Styler (2017), we looked at what is known as A1-P0, an acoustic measurement that compares the amplitude of the highest harmonic peak within the bandwidth of F1 (A1) with the amplitude of a low frequency harmonic peak (either H1 or H2) introduced by nasalization on the vowel. Anticipatory vowel nasalization can be partial (with a gradual increase in strength closer to the velar-lowering gesture of the nasal target) or whole (affecting the entire duration of the vowel) (Cohn 1993). This process has quantifiable acoustic properties that might be affected by recording technology. In general though, we anticipate that devices which can reliably capture the acoustics of anticipatory vowel nasalization will preserve a speaker's distinction between pre-nasal and elsewhere conditions, whether those distinctions are phonetically partial or full. Currently, we are aware of no other studies on social or regional differences in anticipatory /æ/-nasalization that can help us further refine these expectations, and so we had no specific predictions as to how anticipatory nasalization would be realized by our speakers.

In this study, we have included one male speaker and one female speaker, both for gender balance and due to the relationship between sex and spectral variation. On average, adult men tend to have lower pitched voices than women; an important consequence of this is that men exhibit smaller spacing between harmonics than women, making their vowel spaces (the range of formant values) smaller than women's (Simpson & Ericsdotter, 2007). This could also affect measurement of A1-P0 as described above. We had no *a-priori* assumptions about how differences in harmonic spacing might be affected by consumer recording devices, so we included both sexes to explore this relationship.

## II. METHODS

### A. Participants and materials

Data collection took place in the spring of 2020, while universities were closed during the COVID-19 pandemic and only lab members were allowed entry to the first author's lab. This restricted our sample to one male in his early 40s who grew up in Idaho and Arizona and one female in her early 20s from Oklahoma. Both were native speakers of English, monolingual during childhood. Following procedures approved by the Oklahoma State University institutional review board, each speaker gave written, informed consent and was compensated $5 for their 15-minute session. Each read a wordlist which included multiple instances of each North American English vowel phoneme in phonetic environments known to differentiate regional dialects through various mergers and shifts. For the present analysis, 75 words were selected, including 3 for each English monophthong before a coronal obstruent, 3-6 words with each non-low vowel before /l/, and 4 words with /æ/ before /n/.

### B. Procedure and equipment

Simultaneous recordings (Byrne & Foulkes, 2004) were made in a sound-attenuated booth on the Oklahoma State University campus using the seven devices listed in Table I. The lab booth setting was chosen to exclude environmental noise as a variable and focus on recording device differences. Each device was placed between 30-60 cm from the speaker's mouth when seated in front of the arrangement shown in Figure 1. The recorders were five popular commercial mobile devices -- two laptops (PC, Mac), a tablet (iPad), and two smartphones (iPhone, Android) -- plus two professional systems to provide "gold standard" recordings as baselines for comparison (an H4n Pro field recorder, and an AT2021 cardioid condenser microphone connected to a desktop computer outside the booth via a Focusrite audio interface). To simulate everyday speakers' usage

(and often lack of technical expertise), the consumer devices and software were not modified or calibrated for this study.

TABLE I. Recording equipment specifications.

| Device | Specs<br>(Purchase year/Update month) | Software<br>Version (Updated) | File type | Sampling |
|---|---|---|---|---|
| Pro mic + recorder | Mic: Audio Technica 2021 (2017) w/<br>Focusrite Scarlett 18i8 2nd Gen (2017)<br>Dell Optiplex 7050 (2018) w/<br>Windows 10 Pro (Mar 2018) | Audacity<br>2.2.2 (2018) | wav | 16 bit<br>44.1 kHz<br>Mono |
| Field recorder | Zoom H4n Pro Handy Recorder (2017)<br>Built-in mics set at 90-degree angles | System 1.00<br>(Mar 2018) | wav | 16 bit<br>44.1 kHz<br>Stereo |
| PC laptop | Acer Switch Alpha 12 (2017)<br>Windows 10 Home (Nov 2019)<br>Built-in mic | Voice Recorder<br>10.2004.1202.0<br>(2018) | m4a | 32 bit<br>48 kHz<br>Stereo |
| Mac laptop | MacBook Air (2020)<br>MacOS Catalina 10.15.4<br>Built-in mic | Voice Memos<br>2.1 (2019) | m4a | 32 bit<br>48 kHz<br>Stereo |
| Android phone | Pixel 3a (2019)<br>Android 10 (Apr 2020)<br>Built-in mic | Google Recorder<br>1.2.312645208<br>(May 2020) | m4a | 32 bit<br>32 kHz<br>Stereo |
| iPhone | iPhone X<br>iOS 13.4.1<br>Built-in mic | Voice Memos<br>2.1 (2019) | m4a | 32 bit<br>48 kHz<br>Stereo |
| iPad | Air (2014)<br>iOS 12.4.7 (June 2020)<br>Built-in mic | Voice Memos<br>(2019) | m4a | 32 bit<br>44.1 kHz<br>Stereo |

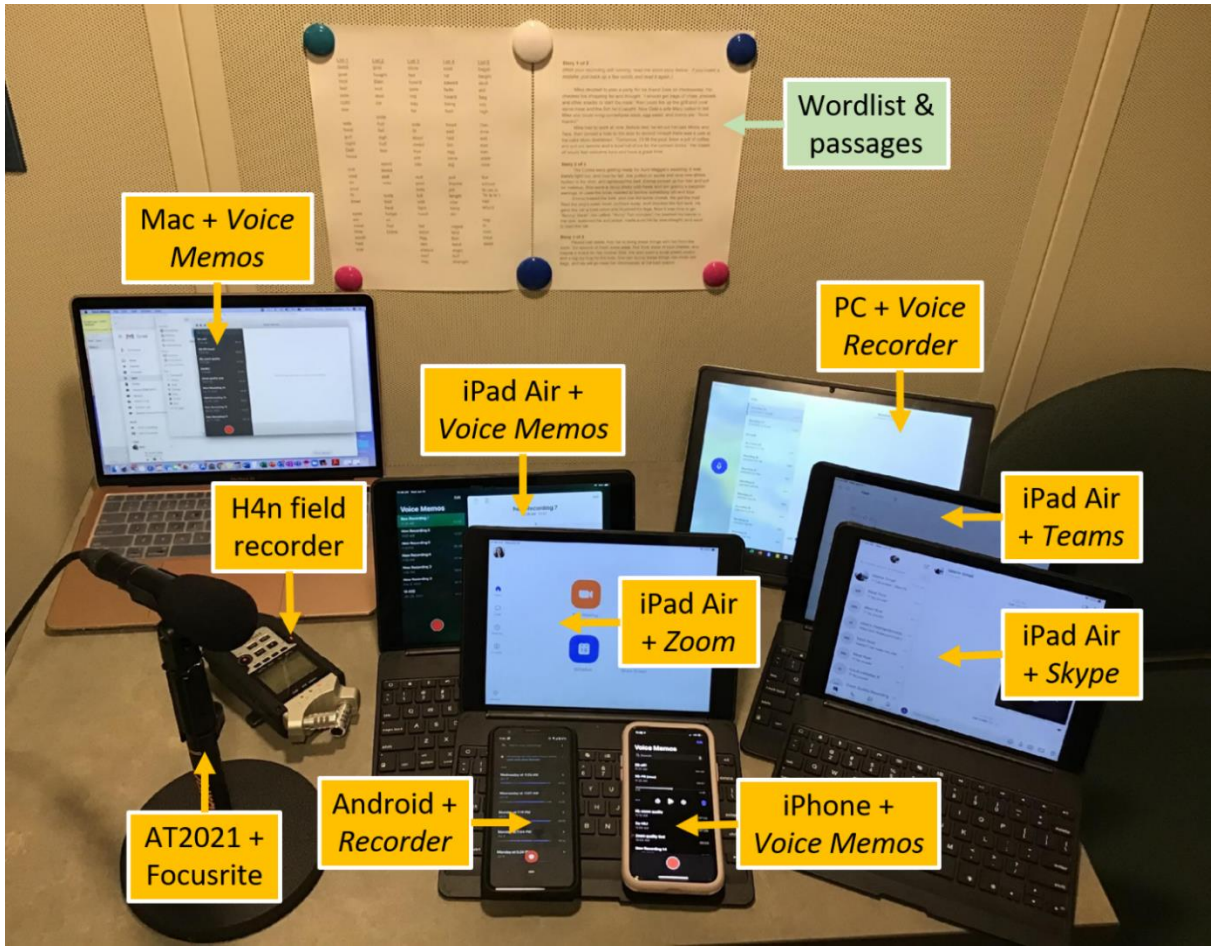


FIGURE 1. Equipment setup in recording booth. Note: Results from the iPads running Zoom, Skype, Teams appear in Freeman & De Decker (2021).

### C. Measurements and analysis

#### *1. Processing*

All m4a recordings (see Table I) were converted to WAV format using Online Audio Converter (123apps, 2020) to facilitate subsequent analyses using Praat (Boersma & Weenink, 2020). This procedure re-encoded each audio file using a sampling rate of 44.1 kHz and a 16 bit depth. WAV files were then processed with a plain text transcript through the semi-automated function of DARLA (Reddy & Stanford, 2015), which creates a Praat TextGrid with word and phone boundaries aligned to the audio. Resulting vowel boundaries were hand-corrected before measurements were taken.

## *2. Vowel formants*

While checking vowel boundaries, Praat's LPC formant trace was visually checked for accuracy, and the number of formants was adjusted to fit the speaker (defaults: 5 for the male, 4 for the female), with adjustments per vowel as necessary (typically increasing the number of formants by 1 to improve Praat's tracking of high-back pre-laterals, which often have very similar F1 and F2 values, totalling about 10% of the male's measurements and about 25% of the female's). A simple Praat script was then used to measure each vowel's formants (F1, F2) at midpoint for plain vowels or 35% of vowel duration for pre-laterals to minimize coarticulatory effects of surrounding consonants. The formant range was set to 0-5000 Hz with a window length of 25 ms and dynamic range of 30 dB, and the maximum formant was increased to 5500 Hz to improve the tracking for about half the male's vowels.

Raw formants from each recording were plotted separately with ellipses of 1 standard deviation around vowel means using NORM (Thomas & Kendall, 2007), whose user-friendly web-based interface allows quick plotting without technical expertise. Obvious outliers far from the ellipses were checked in Praat for measurement errors and corrected. Raw formants were then replotted using the phonR package (McCloy, 2016) in R (R Core Team, 2020), which requires some technical experience but allows fine tuning of graphics for clear, visually appealing plots. Separate plots were made for each speaker to show (a) the overall vowel space outlined by pre-coronal monophthong means, and (b) mergers with ellipses of +/-0.5 standard deviation around vowel means. (This ellipse size was chosen for clarity of presentation; with so few tokens, larger confidence intervals displayed very large ellipses with great overlap.) Plots were inspected visually to compare locations and relative arrangements, and ANOVAs and linear regressions with Device, Speaker, and Vowel as fixed effects

were run in R using the lme4 and afex packages (Bates et al., 2015; Singmann et al., 2020) to determine whether the recording devices affect each formant measure.

### *3. Anticipatory nasalization*

We also examined whether different recording devices presented challenges to measuring the amplitudes of individual harmonics. To test this, we looked at A1-P0, which is a spectral tilt measurement of two harmonic amplitudes found in the lower frequency range. This measurement is often examined in studies of anticipatory vowel nasalization in English and inherent vowel nasality in French (e.g., Chen,1997; Styler, 2017). While A1-P0 is extremely variable across speakers, making a comparison of absolute values impossible, previous studies have shown it is relatively higher in oral contexts and lower in nasal contexts. In this paper, we looked at the anticipatory behavior of A1-P0 over the duration of the low front lax vowel /æ/, which shows a "split" in a number of North American varieties of English. This split consists of tensed and raised variants before nasal consonants and lowered variants before oral consonants (Labov, Ash, & Boberg, 2006). Here, we assessed whether each recording device reliably captured the nasalization (A1-P0) patterns produced by our two speakers.

The A1-P0 measurement procedure was conducted using a Praat script (Styler & Scarborough, 2018) that detects and compares A1, the amplitude of the harmonic closest in value to the first formant, and P0, a low-frequency harmonic referred to as the nasal peak (Chen, 2000). In order to document the process of anticipatory nasal coarticulation, measurements were taken at ⅓ and ⅔ of each token's duration. In the results below, the direction of the A1-P0 slope is illustrated across these two timepoints under each of the recording conditions. A total of 7 tokens per speaker per device were included in this analysis, 3 from the oral context and 4 from the nasal context.

## III. RESULTS

### A. Vowel space shape

Formants measured from both laptops were very similar to those from professional devices (Figure 2), with slightly more variation across devices in the female's values than the male's. However, there was more variation in vowel formant measures across mobile devices (Figure 3)[1]. Variation was greater overall for the female than the male voice, and the alteration was greater in frequencies between about 750 Hz and 1500 Hz, which primarily affected this female's low vowels and this male's high-back vowels and low-front /æ/. The iPad showed lower F1 values for those vowels (female æ, a, ɔ; male æ, o, ʊ) compared to the professional devices, while the smartphones provided higher F1 values for low-front /æ/.

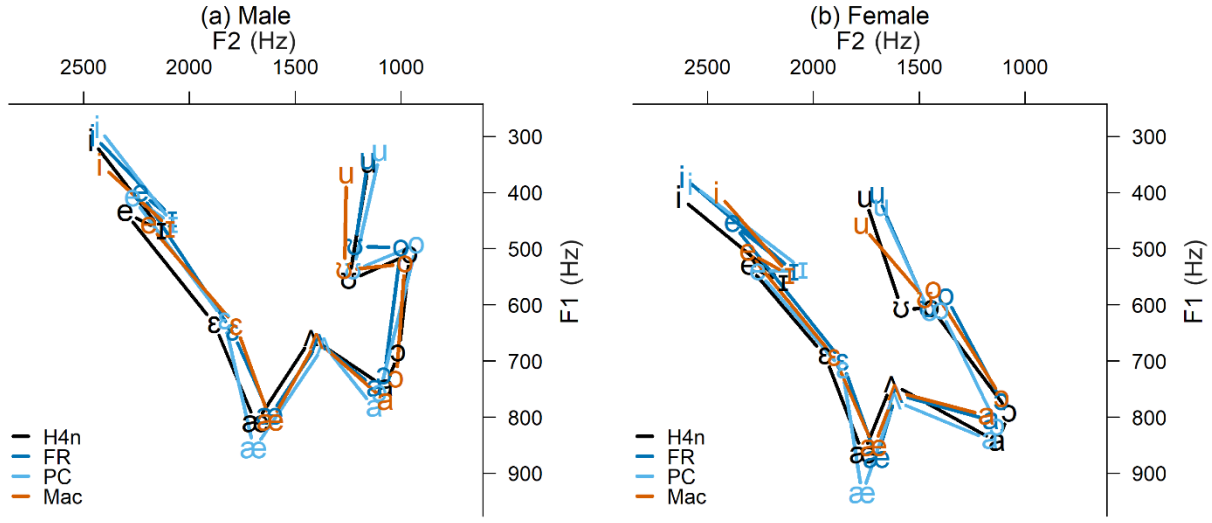


FIGURE 2. Vowel spaces recorded from laptops (Mac, PC) compared to professional devices (H4n Pro field recorder, FR: Focusrite with microphone and computer).

[1] In order to effectively illustrate any differences across devices, we have selected the H4n device as our reference point in many figures. The H4n is commonly used in the field, in many of the same environments where the other consumer devices might be used. This convenience in displaying our results should not be construed as a preference for the H4n as a recording device, nor as a standard toward which the other devices should conform.

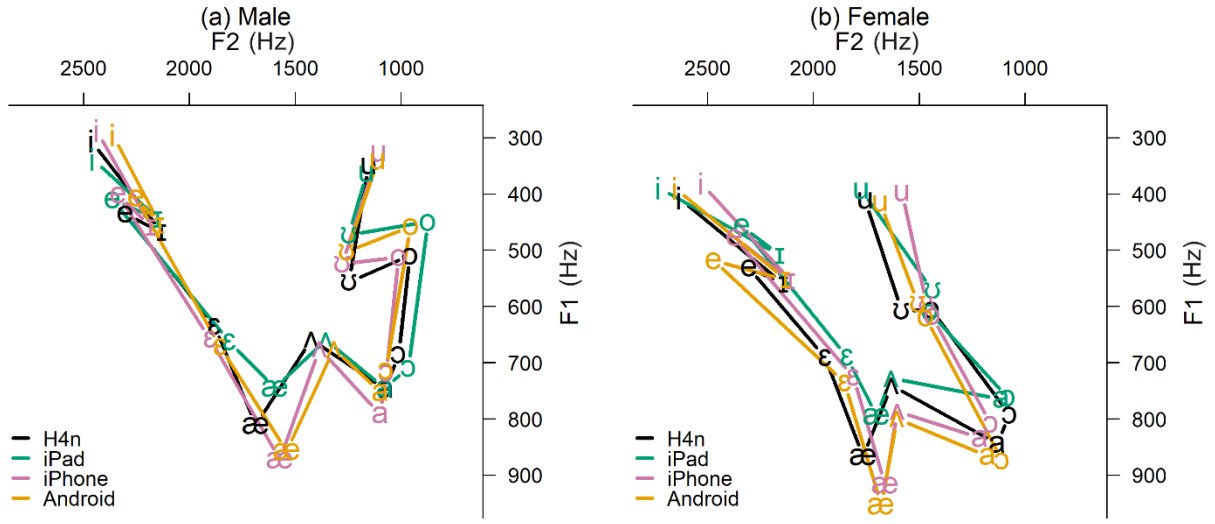


FIGURE 3. Vowel spaces recorded from mobile devices (iPad, iPhone, Android phone) compared to a handheld professional device (H4n Pro field recorder).

### B. Merger: Vowel overlap patterns

Coming from different dialect regions (male: West; female: Oklahoma, a mix of Southern and Midland; Bakos, 2013), the two speakers produced different vowel arrangements and patterns of mergers/shifts (Figure 4). For both, front vowels were slightly backed and lowered before /l/, as expected due to anticipatory coarticulation, and none were involved in pre-lateral mergers (*feel-fill, sale-sell*). This was faithfully represented by all devices. Both speakers produced overlapping low-back vowels, indicative of *caught-cot* merger, as expected for both regions. This was also faithfully represented by all devices, with the greatest separation captured by the H4n for the male in Figure 4.

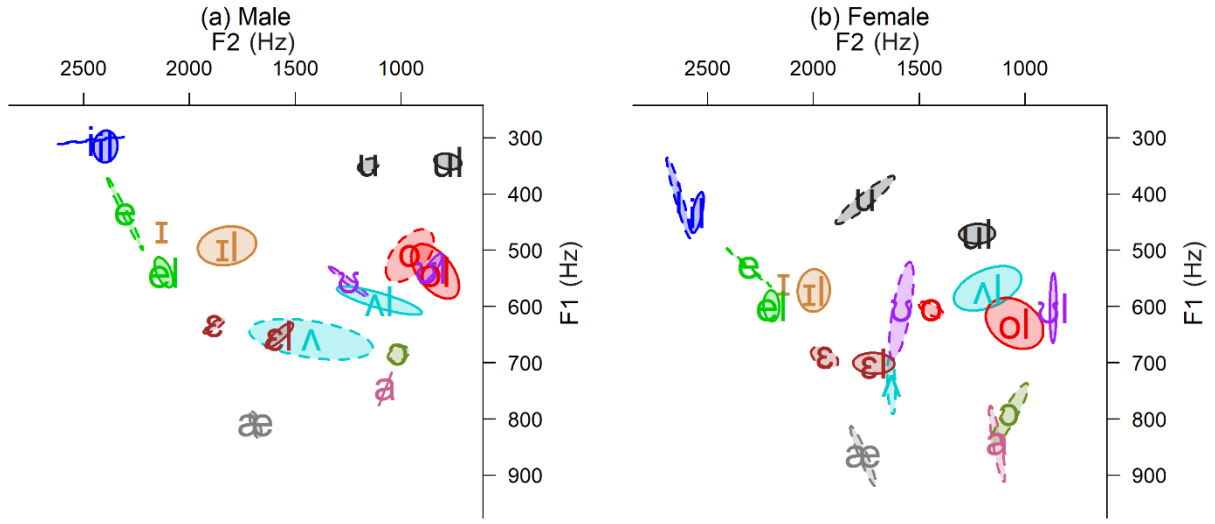


FIGURE 4. Plain and pre-lateral vowels from H4n with 0.5 SD ellipses around means.

On all devices, the female's back rounded plain vowels were fronted, especially /u/, while the male had only a slightly fronted /u/ and /ʊ/ but not /o/. For both speakers, /ʊ/ was low, appearing front of /o/. Before /l/, back rounded vowels were *not* fronted and appeared in the far back vowel space at similar heights to their plain counterparts. For the male, both/ol/ and /ʊl/ were backed to the same location, resulting in complete *bull-bowl* merger, but for the female, /ʊl/ was backed even farther than /ol/, suggesting near-merger. For both speakers, the central /ʌ/ before /l/ was shifted up and back toward these mid-back pre-laterals. For the male, /ʌl/ remained lower and more front than /o/, while the female's /ʌl/ was raised higher than her /ol/. Her /ul/ was also lowered, making it possible that /ʌl/may be in a near-merger relationship with either /ol/ (*dull~dole*) or /ul/ (*skull~school*) – or both (*cull~coal~cool*).

These pre-lateral arrangements were captured by all devices, with some variation in the relative height of the female's /ʊl/: the Focusrite, PC, iPad, and Android phone showed /ʊl/ as high as

/ʌl/ (Figure 5, first two columns), while the H4n, Mac, and iPhone showed its height between that of /ʌl/ and /ol/ (Figure 4; Figure 5, third column).

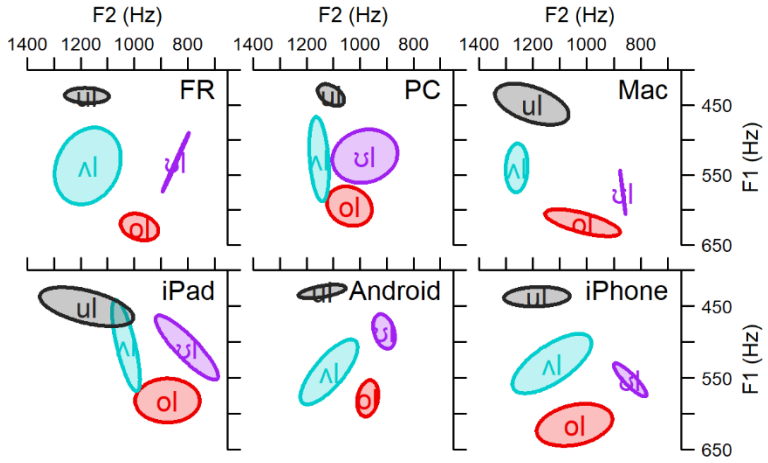


FIGURE 5. Female's back pre-lateral vowels with 0.5 SD ellipses around means, illustrating device differences in height of /ʊl/ (purple).

### C. Vowel formant faithfulness

For an overall view of how devices affected formant measures, three-way ANOVAs and linear regressions were performed for F1 and F2 separately. Three-way ANOVAs showed significant main effects and all interactions for Device, Speaker, and Vowel (with plain and pre-lateral vowels as separate qualities) for F1 and all combinations except Device*Speaker for F2 (Appendix Table I), indicating that the devices affected both formant measures and differentially affected F1 between speakers. Linear regressions with the H4n as the reference device showed that F1 was faithfully measured by the Mac, PC, iPhone, and Android phone, but it was altered by the Focusrite and iPad. Similarly, F2 was faithfully measured by the Mac, PC, and Android but not the Focusrite, iPad, or iPhone (Appendix Table II). When the reference device was set to the Focusrite instead, all other recordings exhibited distorted F1 but not F2 (Appendix Table III).

Finally, we offer a side note on variation in formant tracking accuracy via the number of outliers that were corrected after initial formant measurements were taken for each recording (Table II). The male's initial measurements produced very few outliers: only 4-5 vowels were corrected for each device except the iPad, which had 8 initial outliers. Praat tracked the female's voice a little less accurately overall, with most devices producing 8-9 outliers, the H4n producing only 5, and the iPhone and iPad producing 11 each.

TABLE II. Number of formant outliers (F1 and/or F2) corrected per speaker and device.

| Device | Male | Female |
|---|---|---|
| H4n | 5 | 5 |
| Focusrite | 4 | 8 |
| PC | 5 | 9 |
| Mac | 5 | 8 |
| iPad | 8 | 11 |
| iPhone | 5 | 11 |
| Android | 5 | 9 |

### D. Anticipatory nasalization

Figure 6 illustrates the spectral tilt measure A1-P0 in productions of /æ/ as recorded using the professional devices (H4n, Focusrite) and laptops (Macbook Air, Acer PC). Both nasal and oral contexts are shown. The male speaker showed an expected distinction between oral and nasal contexts with high A1-P0 values in the former and lower in the latter. A1-P0 also diverged over the course of the vowel approaching the following oral or nasal consonants, showing the expected strengthening of nasalization in the latter. This pattern was reproduced by each of the devices, yet each differed in the absolute values and ranges of A1-P0.

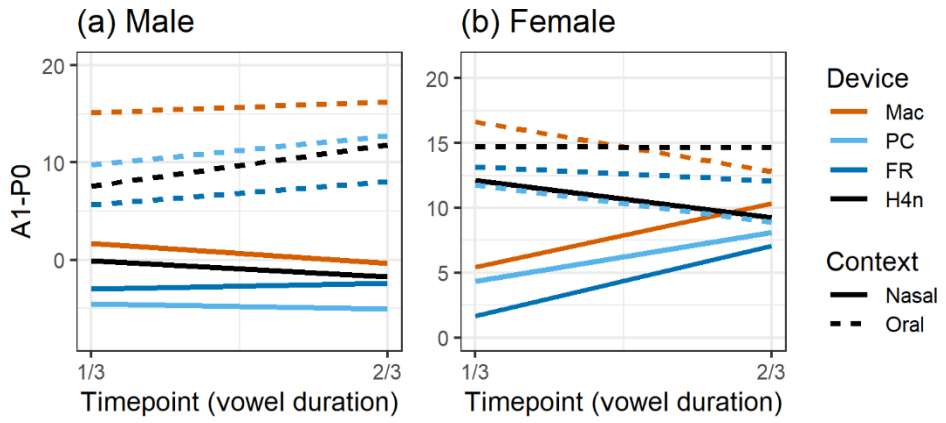


FIGURE 6: Nasalization across vowel duration, laptops (Mac, PC) compared to professional devices (H4n Pro field recorder, FR: Focusrite with microphone and computer).

For the female speaker (Figure 6b), each device also recorded the expected lower values in the nasal context, exhibiting a broad range of variation. However, unlike that reported for the male, the general trend showed a lowering of A1-P0 in the oral context and rising in the nasal context. This was found for the Macbook, PC, and Focusrite, but the H4n bucked this trend, showing a relatively stable A1-P0 in the oral context and the declining A1-P0 in the nasal context that we would expect for a measure of anticipatory nasality. The other devices, while consistent with each other, did not capture this pattern for the female speaker.

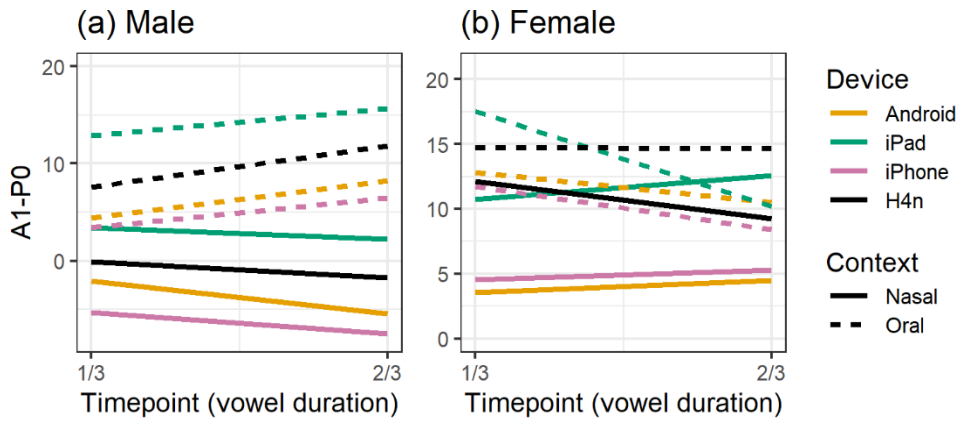


FIGURE 7: Nasalization across vowel duration, mobile devices (iPad, iPhone, Android phone) compared to professional device (H4n Pro field recorder).

The results for A1-P0 taken from recordings made on mobile devices are found in Figure 7, which includes the H4n for ease of comparison with the larger devices in Figure 6. Again, we find speaker-specific differences. The male speaker showed the expected rise/decline by the appropriate context with variation in absolute A1-P0 values introduced by different devices. For the female speaker, the mobile devices seemed to capture a different pattern: lowering of A1-P0 in oral contexts and rising in nasal contexts, with the exception of the iPad, which showed anticipatory nasalization increasing over the two timepoints. The lowering pattern was not expected (though some research has confirmed it as a possible realization of the oral-nasal contrast (Kim & Kim, 2019)), nor was it consistent with that found in H4n recordings. In all cases, the mobile devices introduced variation in the A1-P0 values but remained consistent with the Focusrite pattern described above.

## IV. Discussion

### A. Summary

This paper examined the utility of a number of devices currently accessible to individuals outside of the phonetics lab that could be suitable for remote data collection. Our motivation to undertake this study stemmed from physical distancing practices currently in place that prevent researchers from collecting data using traditional face-to-face methods. Building on previous studies of recording devices, we examined the effects of current personal devices on the acoustic properties of two speakers' vowel spaces and their anticipatory nasalization of /æ/ using A1-P0.

Regarding vowel formants, we conclude that most current personal devices are likely capable of faithfully conveying the relative arrangements of vowels within each speaker, but the distances/amounts of overlap between neighboring classes can vary. If our results generalize beyond our specific recording conditions, laptops may record formants more accurately than mobile devices

like tablets and smartphones, and females' vowel spaces may be more vulnerable to distortion than males' (a finding also reported by De Decker & Nycz, 2011). Mobile devices may also be associated with larger distortions in frequencies between about 750 and 1500 Hz, which especially affects high-back vowels and females' low vowels while leaving the front vowel space less affected.

Regarding nasalization, our finding that frequencies between 750 and 1500 Hz were selectively altered by these devices leads us to expect some effect on A1-P0, since A1 of /æ/ is close in value to 750 Hz, depending on the speaker and dialect. P0, however, typically falls outside this range. We found this to be true for nearly all devices in terms of the absolute A1-P0 values -- no two devices produced the same results. However, for the male speaker, the distinction between nasal and oral contexts was consistently maintained across devices. This is an important finding that suggests anticipatory nasalization may be studied via user-generated recordings given that relative patterns, not absolute values, are typically compared in analyses. Recordings made by the female speaker were less consistent, however, in maintaining the relative oral-nasal distinctions. In addition to this, the two professional devices used in this study (H4n and Focusrite) showed strikingly different patterns for her recordings only.

**B. Questions for future work**

For lossless recordings, the variation between devices, particularly for the female's nasalization measures, was unexpected and introduces several questions for further research. Most striking was the differential effect of devices on our male and female speakers, leading us to ask if spectral properties of certain voices - or genders - are recorded or processed more accurately than others. Future work should include more speakers to determine whether particular speaker or vocal traits can predict variation between device recordings.

In terms of hardware, while we took care to position each device roughly the same distance from the speakers' mouths, the number of devices we included prevented us from adhering to this strictly. Does microphone placement distance affect recordings from some devices more than others? What role does microphone quality play, and how might this interact with speech codecs, the compression technology used to code and decode speech on these devices? That is, do speech codecs that introduce some loss in acoustic detail further attenuate properties altered by the microphone at the audio capture stage? Future work could improve upon our design by playing the recording from a professional device through a high-quality loudspeaker to each test device, one at a time, placed at the same distance and angle.

Relatedly, we observed a bandwidth effect in the range of 750 to 1500 Hz, and we wonder about the purpose or cause of this distortion or whether it might help explain the differential treatment of the two voices. With respect to nasalization, it is known that oro-nasal coupling results in acoustic energy lost in the lower frequency range, including the harmonics that make up the first formant. Further research might consider the unexpected rising of A1-P0 in the nasal context in the female recordings and whether the speech codecs examined here are designed to correct for contextual linguistic changes.

While some recordings exhibited distorted spectral measurements, it should be made clear that these are the product of a combination of factors associated with the recording and measuring process. This includes hardware, such as microphone quality, as well as software, from the codecs used by each device to acoustic analysis parameters selected in Praat. Where exactly in the digital signal landscape the cause, or even if there is a single cause, is something to be determined through further research.

Finally, we note that self-recordings are only one way of employing current technology for linguistic research. As the popularity and accessibility of video conferencing apps like Zoom, Skype, and Teams have increased immensely since the onset of the COVID-19 pandemic, they have become viable tools for remote data collection as well. In another part of our project, we examine the suitability of recordings made through these apps using the same spectral measures reported here (Freeman & De Decker, 2021).

### C. Conclusions and recommendations

With similar performance across our test devices, self-recordings on current personal devices should be suitable for research questions involving relative arrangements of vowels and categorical determinations of merger (e.g., separate, approaching, near/merged), but caution is warranted for questions that rely on small differences to determine distance or amount of overlap, particularly among low and back vowels.

We should be more cautious of these devices for studies of nasalization involving A1-P0. Our male speaker's nasalization patterns were consistent across devices (with variation in absolute values), but further study is needed to determine the cause of variation between devices in our female speaker's patterns.

Due to variation across devices, we make the following recommendations for researchers using self-recordings for data collection:

1. Ask participants about the devices they use to make their recordings and include this as a factor in post-hoc statistical analysis to determine whether device type affects the outcome measures of interest.
2. If participants are likely to be frequent laptop users, consider asking them to use a laptop instead of a tablet or phone to make their recordings (but otherwise consider participant

experience; e.g., for populations who are far more experienced with their phones than computers, trying to use an old or unfamiliar laptop may be so cumbersome as to frustrate them or waste enough time in preparation that they withdraw).

3. If the participant pool is likely to use multiple types of devices, and the research question relies on a measure that may be affected by device type, consider running a quick test like we did here, with a few speakers and simultaneous recordings with the most popular devices among participants, either before or after receiving their recordings. If the measure of interest is systematically affected by a certain type of device, adjustments could be calculated on participant data to improve comparability across speakers.
4. If speakers make multiple recordings in different sessions, ask them to use the same device each time, and instruct them on device placement and minimizing background noise to increase both audio quality and comparability across sessions. Environmental noise was not a factor in the current study, but it varies considerably in natural settings and could be problematic for many types of acoustic measurements (e.g., De Decker, 2016; Rathke, Stuart-Smith, Torsney, & Harrington, 2017).
5. Consider other creative solutions like mailing participants equipment, instructing them on how to disable speech-processing apps on their device before recording, asking them to record on two devices simultaneously, or sending them a small device like an inexpensive keychain loaded with a short audio clip containing multiple variations of the measurements of interest and asking them to play it aloud before recording. While some participant populations may have limited time or technical expertise to aid in recording preparation, others may be receptive to training in microphone placement, environmental noise

reduction, software use, and device modification, which could greatly benefit recording quality and comparability.

Self-recordings have great potential for collecting large natural language samples quickly and cheaply. Researchers save time in scheduling and conducting sessions, and travel time and expenses are eliminated for both researchers and participants. As an example, the first author's Oklahoma vowels study had been scheduling about three participants a week to make recordings in the lab with one of three research assistants, resulting in 10 recordings over three weeks. After switching to self-recordings, we received 27 files in one week and 20 more in the two weeks following a reminder email. We began with the same recruitment pool of local Oklahomans, but asking participants to share the survey/upload link with friends and family vastly increased the response pool beyond the usual campus community, both geographically and demographically. (However, note that males and older respondents were in short supply regardless of recording method. Some older speakers struggled with recording themselves and wished they could come to the lab instead; some persevered and some gave up, indicating that an in-person option should be available when possible. When not possible, consider alternatives like a guided walk-through of the setup over a Zoom call.)

Even with the eventual return of in-person interaction, we can continue to leverage the ubiquity of personal devices to expand our reach, diversify our toolkits, and consider pluralistic methods of data collection.

## ACKNOWLEDGMENTS

Special thanks to Molly Landers (Oklahoma) for her work on data processing and measurement, and to Molly, Peter Richtsmeier (Oklahoma), and Laura Tulk (Newfoundland) for assistance with data collection.

**REFERENCES**

123apps, LLC. (2020). Online Audio Converter. https://online-audio-converter.com.

Bakos, J. (2013). *Comparison of the speech patterns and dialect attitudes of Oklahoma*. Doctoral dissertation, Oklahoma State University.

Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1996). The linguistic consequences of catastrophic events: An example from the American Southwest. In J. Arnold et al. (Eds.), *Sociolinguistic Variation: Data, Theory, and Analysis* (pp. 435-451).

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01.

Bleaman, I. L., & Duncan, D. (2016). The Gettysburg Corpus: Testing the proposition that all tense/æ/s are created equal. *American Speech*, 1-46.

Boersma, P. & Weenink, D. (2020). Praat: doing phonetics by computer, version 6.1.16. http://www.praat.org/.

Bulgin, J., De Decker, P., & Nycz, J. (2010). Reliability of formant measurements from lossy compressed audio. British Association of Academic Phoneticians Colloquium, University of West Minister, London, United Kingdom. Available at https://research.library.mun.ca/684/1/Bulgin_De_Decker_Nycz_2010.pdf

Byrne, C., & Foulkes, P. (2007). The 'mobile phone effect' on vowel formants. *International Journal of Speech Language and the Law, 11*(1), 83-102.

Carignan, C. (2021). A practical method of estimating the time-varying degree of vowel nasalization from acoustic features. The Journal of the Acoustical Society of America, 149(2), 911-922.

Carignan, C., Mielke, J., & Dodsworth, R. (2016). Tongue trajectories in North American English/æ/-tensing. *The future of dialects: Selected papers from Methods in Dialectology XV,* 1, 313.

Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America,* 102(4), 2360-2370.

Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *Journal of Phonetics, 28*(1), 43–67. https://doi.org/10.1006/jpho.2000.0106

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology*, 10(1), 43-81.

De Decker, P. (2016). An evaluation of noise on LPC-based vowel formant estimates: Implications for sociolinguistic data collection. *Linguistics Vanguard, 2*(1).

De Decker, P., & Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis?. *University of Pennsylvania Working Papers in Linguistics, 17*(2), 51–59. Available at https://repository.upenn.edu/pwpl/vol17/iss2/7

Freeman, V., & De Decker, P. (2021). Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *Journal of the Acoustical Society of America, 149*(2), 1211–1223. https://doi.org/10.1121/10.0003529.

Fuchs, R., & Maxwell, O. (2016). The effects of mp3 compression on acoustic measurements of fundamental frequency and pitch range. International Symposium on Computer Architecture (ISCA).

Holland, C., & Brandenburg, T. (2017). Beyond the front range: The Coloradan vowel space. *Publication of the American Dialect Society, 102*(1), 9-30.

Kim, D., & Kim, S. (2019). Coarticulatory vowel nasalization in American English: Data of individual differences in acoustic realization of vowel nasalization as a function of prosodic prominence and boundary. *Data in Brief, 27,* 104593. https://doi.org/10.1016/j.dib.2019.104593

Kolly, M. J., & Leemann, A. (2015). Dialäkt Äpp: Communicating dialectology to the public–crowdsourcing dialects from the public. *Trends in phonetics and phonology. Studies from German-speaking Europe*, 271-285.

Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In J. Baugh & J. Sherzer (eds.), *Language in Use*. Englewood Cliffs: Prentice Hall.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter.

Lupton, D. (Ed.) (2020). Doing fieldwork in a pandemic. [Crowd-sourced document]. https://docs.google.com/document/d/1clGjGABB2h2qbduTgfqribHmog9B6P0NvMgVuiHZCl8/edit

McCloy, D. (2016). phonR: tools for phoneticians and phonologists. R package version 1.0-7. http://drammock.github.io/phonR/

Parsa, V., Jamieson, D. G., & Pretty, B. R. (2001). Effects of microphone type on acoustic measures of voice. *Journal of Voice, 15*(3), 331-343.

R Core Team (2020). R: A language and environment for statistical computing, version 4.0.2. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rathcke, T., Stuart-Smith, J., Torsney, B., & Harrington, J. (2017). The beauty in a beast: Minimising the effects of diverse recording quality on vowel formant measurements in sociophonetic real-time studies. *Speech Communication, 86*, 24-41.

Reddy, S., & Stanford, J. (2015). A Web application for automated dialect analysis. In *Proceedings of NAACL-HLT* (pp. 71–75). IEEE. doi: 10.3115/v1/N15-3015

Simpson, A., & Ericsdotter, C. (2007). Sex-specific differences in f0 and vowel space. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 933-936).

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2020). afex: Analysis of Factorial Experiments. R package version 0.27-2. https://CRAN.R-project.org/package=afex

Styler, W. (2017). " On the acoustical features of vowel nasality in English and French," J. Acoust. Soc. Am. 142(4), 2469–2482. https://doi.org/10.1121/1.5008854

Styler, W., & Scarborough, R. (2018). Nasality automeasure script package. [Praat script]. Available at: https://github.com/stylerw/styler_praat_scripts

Thomas, E. R., & Kendall, K. (2007). NORM: The vowel normalization and plotting suite, version 1.1. http://lingtools.uoregon.edu/norm/norm1.php

**APPENDIX**

TABLE I. Results of three-way ANOVAs with Device, Speaker, and Vowel as fixed effects.

| *F1* | *df* | *F* | *p* |
|---|---|---|---|
| Device | 6 | 19.71 | < .001 |
| Vowel | 18 | 1498.18 | < .001 |
| Speaker | 1 | 1890.53 | < .001 |
| Device:Vowel | 108 | 3.34 | < .001 |
| Device:Speaker | 6 | 4.72 | < .001 |
| Vowel:Speaker | 18 | 13.42 | < .001 |
| Device:Vowel:Speaker | 108 | 1.56 | < .001 |
| Residuals (df denominator) | 691 | | |
| *F2* | *df* | *F* | *p* |
| Device | 6 | 5.43 | < .001 |
| Vowel | 18 | 2916.77 | < .001 |
| Speaker | 1 | 1323.82 | < .001 |
| Device:Vowel | 108 | 2.05 | < .001 |
| Device:Speaker | 6 | 0.90 | ns |
| Vowel:Speaker | 18 | 55.77 | < .001 |
| Device:Vowel:Speaker | 108 | 1.31 | < .05 |
| Residuals (df denominator) | 691 | | |

TABLE II. Fixed effects results (Device, Speaker) of linear regressions with the H4n as the reference device. (Not shown: all Vowels differed from the reference /ɔ/ ($p < .05$) except F2 for /a/, suggesting low-back near-merger.)

| *F1* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept: H4n, female) | 806.70 | 10.31 | 54.53 | 78.23 | < .001 |
| Device: Android | -6.59 | 3.77 | 881.79 | -1.75 | ns |
| Device: FR | -18.43 | 3.77 | 881.71 | -4.88 | < .001 |
| Device: iPad | -28.84 | 3.77 | 881.79 | -7.66 | < .001 |
| Device: iPhone | -5.82 | 3.77 | 881.71 | -1.54 | ns |
| Device: Mac | -5.53 | 3.76 | 879.76 | -1.47 | ns |
| Device: PC | -3.46 | 3.77 | 881.71 | -0.92 | ns |
| Speaker: male | -74.97 | 2.02 | 879.34 | -37.18 | < .001 |
| *F2* | *Estimate* | *SE* | *df* | *t* | *p* |
| (Intercept: H4n, female) | 1179.39 | 27.32 | 60.79 | 43.16 | < .001 |
| Device: Android | -10.83 | 12.32 | 883.82 | -0.88 | ns |
| Device: FR | -16.64 | 12.34 | 883.70 | -1.35 | ns |
| Device: iPad | -29.11 | 12.32 | 883.82 | -2.36 | < .05 |
| Device: iPhone | -31.83 | 12.34 | 883.70 | -2.58 | < .05 |
| Device: Mac | -19.13 | 12.31 | 881.67 | -1.55 | ns |
| Device: PC | -14.86 | 12.34 | 883.70 | -1.20 | ns |
| Speaker: male | -168.34 | 6.60 | 881.25 | -25.51 | < .001 |

TABLE III. Fixed effects results (Device, Speaker) of linear regressions with the Focusrite as the reference device. (Not shown: all Vowels differed from the reference /ɔ/ ($p < .05$) except F2 for /a/, suggesting low-back near-merger.)

| *F1* | *Estimate* | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| (Intercept: Focusrite, female) | 788.27 | 10.32 | 54.60 | 76.42 | < .001 |
| Device: Android | 11.84 | 3.77 | 879.19 | 3.14 | < .01 |
| Device: H4n | 18.43 | 3.77 | 881.71 | 4.88 | < .001 |
| Device: iPad | -10.41 | 3.77 | 879.19 | -2.76 | < .01 |
| Device: iPhone | 12.61 | 3.78 | 879.11 | 3.33 | < .001 |
| Device: Mac | 12.90 | 3.78 | 879.77 | 3.42 | < .001 |
| Device: PC | 14.97 | 3.78 | 879.11 | 3.96 | < .001 |
| Speaker: male | -74.97 | 2.02 | 879.34 | -37.18 | < .001 |
| *F2* | *Estimate* | *SE* | *df* | *t* | *p* |
| (Intercept: Focusrite, female) | 1162.75 | 27.34 | 60.92 | 42.53 | < .001 |
| Device: Android | 5.81 | 12.36 | 881.11 | 0.47 | ns |
| Device: H4n | 16.64 | 12.34 | 883.70 | 1.35 | ns |
| Device: iPad | -12.47 | 12.36 | 881.11 | -1.01 | ns |
| Device: iPhone | -15.19 | 12.38 | 880.99 | -1.23 | ns |
| Device: Mac | -2.49 | 12.36 | 881.68 | -0.20 | ns |
| Device: PC | 1.78 | 12.38 | 880.99 | 0.14 | ns |
| Speaker: male | -168.34 | 6.60 | 881.25 | -25.51 | < .001 |