Assessing the validity of remote recordings captured with a generic smartphone application designed for speech research

Joshua Penney, Ben Davies, Felicity Cox

Centre for Language Sciences, Department of Linguistics, Macquarie University
joshua.penney@mq.edu.au; ben.davies@mq.edu.au; felicity.cox@mq.edu.au

Abstract

This paper introduces a generic smartphone application designed for collecting speech data remotely: Appen Research. Data collected from 24 female Australian English-speaking participants using the smartphone application were compared to data collected from the same participants with laboratory-based equipment. F1 and F2 values were extracted from the midpoints of the 11 stressed monophthongs of Australian English. While some non-low vowels showed slight raising of F1 values in the data recorded with the application, overall the results suggest that recordings collected with the smartphone application are generally comparable to recordings made in a studio for the purposes of analysing vowel formants.

Index Terms: remote data collection, smartphones, vowels, Australian English, monophthongs, recording methods

1. Introduction

1.1. Background

There has been a shift in recent years towards the use of simple personal electronic devices such as smartphones, tablets, and laptop computers to facilitate data collection for scientific research as such portable devices have become increasingly common (at least within certain sections of the population) [1, 2, 3, 4]. With regard to speech research, an approach to data collection that leverages the accessibility of such devices which typically contain an inbuilt microphone and an internet connection makes economical and practical sense. For example, in Europe a number of custom-made recording applications (apps) for smartphones that were paired with targeted (social) media campaigns have successfully exploited high levels of mobile phone ownership to crowdsource speech data from large numbers of speakers of certain languages and dialect groups [5, 6, 7, 8, 9, 10].

During the COVID-19 pandemic, public health orders resulted in restrictions on face-to-face contact as well as restraints on movement, at times with various state and national borders being closed. Under these circumstances, traditional laboratory-based data collection was generally unable to proceed, and as a result many researchers turned to remote data collection in order to continue their research. This resulted in a surge of interest in options for remote data collection, as well as a number of comparison studies examining the viability of data collected remotely for speech research and the extent to which acoustic measurements are affected by such methods [11, 12, 13, 14, 15]. These studies complement research assessing the validity of speech data collected via smartphones for clinical voice analysis [16, 17, 18]. Although recordings made with personal devices tend to show some deviation from those made with reference devices, in general most studies suggest that recordings made with modern smartphones are sufficiently similar to those made with traditional lab-based recording equipment, at least for the purposes of examination of F0 and the first two-three formants [11, 14, 18]. There are some important caveats to ensure that the quality of the recordings is sufficient for acoustic data analysis: data should be captured in lossless rather than lossy formats (e.g. wav rather than mp3), should not be recorded over an internet connection but rather locally recorded (i.e. on a participant's phone) prior to data being transferred, and deviations may be more problematic in particular vowels and in particular individuals [11, 13, 14]. Furthermore, for some fine-grained analyses, remote recordings made with smartphones may not be suitable [11, 15]; for example, [15, 18] found problematic differences for some acoustic measurements of voice quality. This was particularly the case for voice quality measurements based on harmonic amplitudes.

In the Australian context, remote data collection offers great potential to speech researchers. It may lead to increased participation, as it removes barriers such as the need to travel to and be present physically at a university lab; rather, participation can be done in the comfort of a participant's own home at a time that is convenient for them [12, 19]. This has the potential to open up participation to a wider sample of the population, rather than the overrepresentation of (mostly young, middle class, female) university students that very often form the participant base in many speech studies.

The ability to record participants remotely is particularly important given the geography of Australia, as it could enable speakers from rural and isolated communities outside of major urban centres to be better represented in speech research. For example, this would be beneficial for work examining variation within languages, and would allow comparisons to be made between speakers from more than just a subset of locations [20, 21, 22]. Similarly, the use of remote data collection methods could assist in documentation and revitalisation of under-studied languages such as Australian languages spoken in remote indigenous communities, with recording apps having been shown to be a useful tool for research on endangered languages [23, 24]. Remote data collection would not only provide benefits for researchers working with populations located within Australia; researchers who work on and with speech communities in overseas countries could also benefit, as the ability to record participants remotely could reduce the number of fieldwork trips and the associated travel costs that are generally required for in person data collection [12], and could also ease issues relating to accessibilty, which may be a challenge faced by some researchers. As alluded to above, the ability to collect speech data remotely can also serve to ensure that research can be continued in the case of renewed restrictions on movement, whether due to public health restrictions or other unforeseen events.

As mentioned above, some speech research teams have created their own custom recording apps that are specially designed to address particular research questions by eliciting speech through highly controlled tasks [5, 6, 7, 8, 9, 10, 23]. While such an approach is ideal for addressing specific research questions, designing and programming a custom made app is also costly, time-consuming, and requires advanced programming abilities (or funding to employ external programmers). This puts such an approach beyond the reach of many researchers, particularly students and early career researchers. Additionally, as smartphone operating systems are continually updated, ongoing maintenance is required to ensure compatibility with new operating system versions, without which a custom app may become deprecated over a short time period. Moreover, as such apps are designed with specific languages/dialects/research questions in mind, the result is that multiple researchers create a variation of what is, essentially, the same tool over and over again at great effort and expense.

The issues outlined above highlight the need for a generic recording app to enable remote speech data collection across many different projects. Some generic recording apps do exist; however, there is a distinct lack of simple, cross-platform tools that are able to capture the high quality, uncompressed recordings necessary for speech research. Moreover, existing apps that provide high quality recordings generally require subscriptions or come with advertisements. In most cases participants need to manage data settings (e.g. file formats, sampling rates, etc.) and then manually upload data to a repository, or in some cases email the saved files to the researcher: a dangerously insecure data management practice that inevitably leads to suboptimal data quality and missing – or in a worst case scenario, stolen – data.

In this paper, we introduce the Appen Research app, a simple, easy to use, generic recording app that has been designed specifically with the requirements of speech researchers in mind.

1.2. Appen Research

The Appen Research app has been designed to function as a simple, cross-platform smartphone recording device that can be used for remote data collection of high-quality speech data. The app is currently in beta testing mode. In its current configuration, it can record speech to wav file format with a 44.1kHz sampling rate and 16-Bit resolution with a maximum file duration of 15 minutes. In addition to the recorded files, metadata related to smartphone make and model, and operating system and version are also collected with each recording.

Importantly, the app is not linked to a particular experiment or project, nor does it include specific instructions for a participant to follow or consent/questionnaire forms to complete. This enables it to be used for a variety of data collection purposes on any number of different projects. For example, the app could be used to capture data from participants in tasks that are supervised by a researcher, in which supervision could be carried out by video/phone call (on a separate device). In another case, a participant may be given instructions for a simple reading task that should be completed at regular intervals without researcher supervision. The app could be used for highly controlled experiments, for recording conversations between two interlocutors, or for capturing infant direct speech. In some cases, both local and non-local participants may take part in the same study, and recording via the app would ensure a consistent recording methodology between participants. It is entirely up to the researcher how to implement their task; the app is envisaged merely as a 'portable' recording device.

Prior to data collection beginning, the researcher needs to establish a project identification code and a set of participant identification numbers to allocate to participants. Additionally, the researcher needs to designate a secure storage location for the data to be transferred to. Currently, data is configured to be transferred directly to a user's CloudStor account using the FileSender API. The recorded file is encrypted on the participant's device, temporarily staged on Appen's servers and uploaded to CloudStor within a 15 minute window. The use of end-to-end encryption is vital when collecting data remotely. It gives researchers control over their data and ensures participants can trust their recordings are not accidentally released to the wider public or stolen by those with malicious intent. AARNet's CloudStor/FileSender environment was selected to perform this function within the Appen Research app because the company is owned by a consortium of Australian universities, making it highly sensitive to the evolving data and privacy concerns of Australian research institutes.

The app is designed for simple participant usability. It has cross-platform compatibility with both Apple iOS and Android operating systems, and will be available for free for participants to download onto their own smartphone. Once a participant enters a project identification code and their participant identification number, no settings need to be adjusted; rather, the participant's interface consists of a record/pause/stop button to begin, pause, and end the recording. Once the recording is completed, the participant is prompted to upload and transfer the file, after which the task is comlete. Data transfer only takes place when a wifi connection is available.

In this paper, we present the preliminary results of a study designed to compare vowel realisations of the Australian English (AusE) monophthongs in recordings captured with the Appen Research app against recordings of the same population recorded in a laboratory setting with professional recording equipment.

2. Methods

2.1. Participants

We recruited 25 female speakers (aged between 19 and 64; mean age: 28) to take part in this study. Of these, 20 were undergraduate students in the Department of Linguistics at Macquarie University, who received course credit for their participation. The remaining five participants were researchers in the Department of Linguistics at Macquarie University, who were not compensated for their time. All participants were L1 or early L2 speakers of AusE who had completed all of their schooling in Australia. Data for one participant were excluded due to a technical issue that resulted in a partial loss of data, leaving data for 24 speakers remaining for analysis.

2.2. Procedure

All participants completed the same task in two successive sessions recorded in a sound-treated room in the Department of Linguistics at Macquarie University: in one of the sessions, participants were recorded with a Neumann TL103 condenser microphone using open-source recording software Audacity (https://www.audacityteam.org/) with a sampling rate of 44.1 kHz and 16-Bit resolution. Recordings from this session will hereafter be referred to as *Studio* data. In the other session, participants were recorded through the Appen Research app

with a sampling rate of 44.1 kHz and 16-Bit resolution, which they were instructed to install on their own personal smartphone prior to the session. Recordings from this session will hereafter be referred to as *App* data.

Participants were recorded as they read aloud 90 sentences presented orthographically on a computer monitor. Each sentence contained a monosyllabic target word with the standard /hVd/ structure, where /V/ comprised the 18 stressed vowels of AusE [23]. The target words were embedded within a carrier sentence with the form: *say <TARGET> again*, with speakers instructed to read the sentences casually as if speaking to a friend. For each participant, five repetitions of each of the 18 stressed vowels were sampled. The order of presentation was randomised for each participant (with the same presentation order in both sessions). It was our intention to counterbalance the order of sessions; however, as data collection is ongoing and the analysis here is based on an initial subset of collected data, the data examined here are not equally balanced (*Studio* prior to *App*: 20; *App* prior to *Studio*: 5). For the purposes of this analysis, only data relating to the 11 AusE monophthongs were examined.

2.3. Acoustic analysis

The collected files were orthographically transcribed and subsequently automatically segmented and force-aligned through WebMAuS [26] utilising an AusE model. The resulting textgrids and corresponding sound files were converted to an emu database using emuR [27].

Formant frequencies were calculated with Praat [28] and imported into the database via PraatR [29]. Formant frequencies were calculated for all back and central vowels with the default settings: Max. formant: 5500Hz; No. formants: 5, Window length: 0.025s; front vowels were estimated with the following settings as these resulted in improved formant tracking (and consequently fewer outliers): Max. formant: 6600Hz; No. formants: 5, Window length: 0.025s.

All files were inspected and segment boundaries for the vowels were hand corrected where necessary. In some cases, participants misread the sentence and produced an incorrect target word: 37 files were excluded from analysis due to such errors. F1 and F2 measurements (in Hertz) were then extracted at the temporal midpoint of each vowel. Outliers of each vowel were subsequently trimmed using the modified Mahalanobis distance method [30]. This resulted in 2472 files remaining for the analysis (*App*: 1243; *Studio*: 1229).

2.4. Statistical analysis

Potential differences between the *Studio* and *App* files were examined using linear mixed effects regression models with the lme4 [31] package in R [32], with *p*-values calculated by likelihood ratio tests with the afex [33] package. Separate models were fitted for F1 and F2, in each case with the formant in question included as the dependent variable. Fixed factors were Recording method (i.e. *Studio* vs *App*) and Vowel. We also included an interaction term between these fixed factors. Random intercepts were included for Participant. Random slopes were included for Recording method by Participant in the F1 model. The inclusion of Recording method by Participant in the F2 model and the inclusion of random slopes for Vowel by Participant in both F1 and F2 models resulted in singular model fits; hence these were not included in the final models. Post-hoc pairwise comparisons were conducted with the emmeans package [34] with Tukey HSD corrections for multiple comparisons.

As participants utilised their own personal smartphone for the *App* recordings, it was not possible for us to control for or balance the operating system and version of the operating system that was used. The majority of participants (n = 19) completed the *App* session using an Apple iPhone and a version of the Apple iOS operating system. The remaining participants used a Samsung (n = 4) or Google (n = 1) smartphone and a version of the Android operating system. Despite operating system not being well sampled, we included this as a covariate in our initial modelling. However, this was not a significant effect for either F1 or F2 and did not improve the model fit and was therefore removed from the final models. In addition, a subsequent analysis of the data excluding participants who used the Android operating system did not show substantial differences to the analysis that included all participants.

3. Results

Figure 1 illustrates the distribution of vowels according to the two recording methods, as presented in the traditional F1/F2 vowel plane. Means of each of the vowels are shown by the vowel labels, and the ellipses represent 95% confidence intervals. As can be seen, the monophthongal space appears similar in the data captured by the two recording methods and the vowels appear to be comparably dispersed. However, some small differences can also be noticed; in particular, some of the vowels in the *Studio* data – particularly the non-low vowels – appear to be slightly raised, i.e. they appear to be lower in F1.

The linear mixed effects model for F1 showed significant effects for Recording method ($F(1, 22) = 8.11$; $p = 0.009$) and Vowel ($F(10, 2405) = 3456.67$; $p < 0.001$), as well as their interaction ($F(10, 2406) = 4.66$; $p < 0.001$). Post-hoc pairwise comparisons revealed that the only vowel that differed significantly between the recording methods was /iː/ ($p = 0.041$), which had a mean F1 value in the *Studio* data that was 28.4 Hz lower than in the *App* data. Pairwise comparisons further showed that within each recording method each vowel differed significantly in F1 from all other vowels ($p = 0.0079$ or below) with the exception of /iː/ and /ʉː/ and /ɪ/ and /ʊ/ in the *App* data, and /iː/ and /ʉː/, /ɪ/ and /ʊ/, and /e/ and /ɔ/ in the *Studio* data. That is, /e/ and /ɔ/ differed significantly from one another in the App data (with a difference of 29.4 Hz), but not in the Studio data (with a difference of 20.7 Hz), where mean values for /ɔ/ were marginally higher.

The linear mixed effects model for F2 showed a significant effect for Vowel ($F(10, 2427) = 6678.36$; $p < 0.001$). There was no effect of Recording method or the interaction between Recording method and Vowel for F2. Post-hoc pairwise comparisons showed that each of the vowels differed significantly in F2 from all other vowels (all $p = 0.0013$ or below), apart from /ʊ/ and /ɔ/.

4. Discussion

Overall, the results above show that differences in formant measurements between the two recording methods are relatively minimal, which is consistent with previous findings [14, 18]. Small differences in F1 were observable for some vowels, with lower mean F1 values (corresponding to higher vowel realisations) in the *Studio* data than in the *App* data. However, this effect was only found to be significant for /iː/, and even in that case the size of the effect was not considerable. Given the fact that /iː/ displays variable onglide in AusE [20], it may be possible that the greater difference in this vowel could be due to measurements being taken at the vowel midpoint

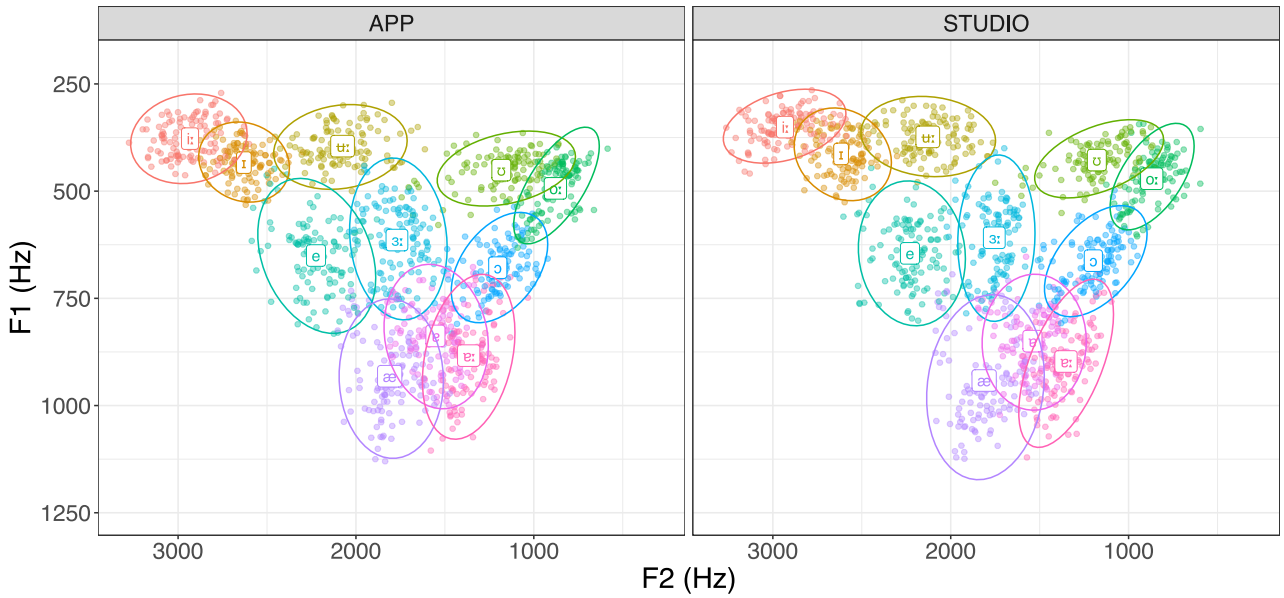

Figure 1. F1 and F2 values of all monophthongs in the *App* (left panel) and *Studio* (right panel) data. Text labels represent mean values. Ellipses represent 95% confidence intervals.

rather than identifying the vowel target. In both recording methods, vowels were found to be similarly dispersed, with the only difference found relating to /e/ and /ɔ/, which showed a significant difference in F1 in the *App* data, but not in the *Studio* data, where there was a higher realisation (lower mean F1 value) of /ɔ./ Taken together, these results suggest that F1 values in the recordings made with the Appen Research app may be affected by a slight raising; that is, some vowels may be measured as marginally lower when captured with the app.

This apparent effect of F1 raising on some (non-low) vowels in the *App* recordings may be the result of internal digital signal processing within the smartphones used by the participants. Most smartphones employ a range of algorithms to digitise and improve the audio signal, for example to reduce background noise and enhance the clarity of the speech that is captured [35, 36]. These processes, and the extent of processing, vary between different makes and models of smartphone, with little specific information available publicly about exactly what processes are applied in a particular phone. Nevertheless, such processes may very well impact upon acoustic measurements to some extent [37]. Band-pass filters that have traditionally been used in the transmission of speech by telephone are also known to inflate F1 values in non-low vowels in some cases [38, 39], among other effects. However, the effects seen here do not appear to be as substantial as those previously reported, and appear to be limited to F1: there were no significant differences according to F2.

Note that it was not our intention here to compare recordings of the same utterances captured simultaneously with different recording devices, as is often the case in studies comparing speech recorded with personal devices. Rather, in this study it was our intention to assess whether recordings of the same general population would yield effectively comparable results in non contemporaneous recordings. Apart from the small differences discussed above, this appears to be the case. We therefore suggest that remote data collection with the Appen Research app may be an effective means for conducting speech research in the Australian context, at least for the examination of F1 and F2 values in AusE monophthongs. Of course, recordings captured with the app may be more suitable for some lines of enquiry rather than others, as some acoustic measurements are likely to be more susceptible to deviations in recordings made with smartphones [11, 13, 15].

It should be pointed out that this preliminary study was based on a relatively small sample size. As noted above, utilising smartphones for remote recordings may facilitate greater participation, which in turn would lead to larger sets of data for analysis. Further research will determine whether the findings shown here also hold over a greater number of participants and with speakers from more heterogeneous backgrounds.

5. Conclusion

This paper introduced the Appen Research app, a generic smartphone based recording application for speech research, and has shown that recordings from participants' smartphones captured with this smartphone application are generally comparable to recordings made in a studio for the purposes of analysis of the first two formants taken from the midpoints of AusE monophthongs. The use of this application may therefore be of benefit to researchers interested in collecting data remotely with the intention of examining measurements of vowel formants. Future work will address the extent to which other acoustic measurements are also comparable between remote and studio-based recordings.

6. Acknowledgements

We thank members of the MQ phonetics lab and participants in the Challenges for Change workshop at LabPhon18 for their comments and suggestions. This work was supported by ARC Future Fellowship Grant FT180100462 to the third author.

7. References

[1] Birenboim, A. and Shoval, N., "Mobility research in the age of the smartphone," Ann. Am. Assoc. Geogr., 106(2):283–291, 2016.

[2] Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., and Gosling, S. D., "Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges," Perspect. Psychol. Sci. 11, 838–854, 2016.

[3] Miller, G., "The Smartphone Psychology Manifesto," Perspect. Psychol. Sci. 7(3): 221–237, 2012.

[4] Seifert Alexander, Hofer Matthias, Allemand Mathias., "Mobile data collection: Smart, but not (yet) smart enough," Frontiers in Neuroscience, 12, 2018. DOI=10.3389/fnins.2018.00971

[5] Leemann, A., Kolly, M.-J., Goldman, J.-P., Dellwo, V., Hove, I., Almajai, I., and Wanitsch, D., "Voice Äpp: A mobile app for crowdsourcing Swiss German dialect data," in Proc. INTERSPEECH 2015, Dresden, 2804–2808, 2015.

[6] Leemann, A., Kolly, M.-J., and Britain, D., "The English Dialects app: The creation of a crowdsourced dialect corpus," Ampersand, 5: 1–17, 2018.

[7] Entringer, N., Gilles, P., Martin, S., and Purschke, C., "Schnëssen: Surveying language dynamics in Luxembourgish with a mobile research app," Linguistics Vanguard, 7(s1): 1–15, 2021.

[8] Gittelson, B., Leemann, A., and Tomaschek, F., "Using crowd-sourced speech data to study socially constrained variation in nonmodal phonation," Frontiers in Artificial Intelligence, 3: 1–9, 2021.

[9] Hilton, N. H., "Stimmen: A citizen science approach to minority language sociolinguistics," Linguistics Vanguard, 7(s1):1–15, 2021.

[10] Leemann, A., "Apps for capturing language variation and change in German-speaking Europe: Opportunities, challenges, findings, and future directions," Linguistics Vanguard, 7(s1): 1–12, 2021.

[11] Freeman, V., P. DeDecker, and M. Landers, "Suitability of self recordings and video calls: Vowel formants and nasal spectra," J. Acoust. Soc. Am, 148: 2714, 2020.

[12] Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M. and Messerli, J., "Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and video conferencing," Linguistics Vanguard, 6(s3): 1–16, 2020

[13] Sanker, C., Babinski, S., Burns, R., Evans, M., Johns, J., Kim, J., Smith, S., Weber, N., Bowern, C., "(Don't) try this at home! The effects of recording devices and software on phonetic analysis," Language, 97(4): e360-e382, 2021.

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

[15] Penney, J., Gibson, A., Cox, F., Proctor, M., & Szakay, A., "A comparison of acoustic correlates of voice quality across different recording devices: A cautionary tale," in Proc. INTERSPEECH 2021, Brno, 1389–1393, 2021.

[16] Manfredi, C., Lebacq, J., Cantarella, G., Schoentgen, J., Orlandi, S., Bandini, A., and DeJonckere, P. H., "Smartphones offer new opportunities in clinical voice research," J. Voice, 31(1): 111.e1–111.e7, 2016.

[17] Grillo, E. U., Brosious, J. N., Sorrell, S. L., and Anand, S.,"Influence of smartphones and software on acoustic voice measures," Int. J. Telerehabilitation, 8: 9–14, 2016.

[18] Jannetts, S., Schaeffler, F., Beck, J., and Cowen, S., "Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types," Int. J. Lang. Commun. Disord., 54(2): 292–305, 2019

[19] Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., and Lawless, M., "Using Zoom video conferencing for qualitative data collection: Perceptions and experiences of researchers and participants," Int. J. Qual. Methods, 18: 1–8, 2019.

[20] Cox, F., and Palethorpe, S., "The border effect: Vowel differences across the NSW/Victorian border," In C. Moskovsky [Ed.], *Proc. Aust. Ling. Soc 2003,* 1–14, 2004.

[21] Burnham, D., Estival, D., Fazio, S., Viethen, J., Cox, F., Dale, R., Cassidy, S., Epps, J., Togneri, R., Wagner, M., Kinoshita, Y., Göcke, R., Arciuli, J., Onslow, M., Lewis, T., Butcher A., and Hajek, J., "Building an audio-visual corpus of Australian English: Large corpus collection with an economical portable and replicable black box," in Proc. INTERSPEECH 2011, Florence, 841–844, 2011.

[22] Cox, F., and Palethorpe, S., "Vowel variation across four major Australian cities," in Proc. ICPhS, Melbourne, 577–581, 2019.

[23] Bird, S., Hanke, F. R., Adams, O., and Lee, H., "Aikuma: A mobile app for collaborative language documentation," Proc. Workshop on the Use of Computational Methods in the Study of Endangered Languages, 1–5, 2014.

[24] Bird, S., "Designing Mobile Applications for Endangered Languages," in K. L. Rehg and L. Campbell [Eds], The Oxford Handbook of Endangered Languages, Oxford University Press, 2018.

[25] Cox, F., and Palethorpe, S., "Australian English", J. Int. Phon. Assoc., 37(3): 431–350, 2007.

[26] Kisler, T., Reichel, U., and Schiel, F., "Multilingual processing of speech via web services," Comput. Speech Lang., 45: 326–347, 2017.

[27] Winkelmann, R., Harrington, J., and Jänsch, K., "EMU-SDMS: Advanced speech database management and analysis in R," Comput. Speech Lang., 45: 392–410, 2017.

[28] Boersma, P., and Weenink, D., "Praat: Doing phonetics by computer," version 6.1.16, 2020 [Computer program]. Available: http://www.praat.org/

[29] Albin, A. L. "PraatR: An architecture for controlling the phonetics software "Praat" with the R programming language," J. Acoust. Soc. Am., 135(4): 2198, 2014.

[30] Stanley, J. A. "The Absence of a Religiolect among Latter-Day Saints in Southwest Washington," In V. Fridland, A. Beckford Wassink, L. Hall-Lew, and T. Kendall [Eds], Speech in the Western States: Vol. 3, Understudied Varieties, 95–122. Duke University Press, 2020.

[31] Bates, D., Mächler, M., Bolker, B., and Walker, S., "Fitting linear mixed-effects models using lme4," J. Stat. Softw., 67(1), 2015.

[32] R Core Team, "R: A language and environment for statistical computing," version 4.0.2, 2020 [Computer program]. Available: https://www.r-project.org/

[33] Singmann, H., Bolker, B., Westfall, J., Aust F., and Ben-Shachar, M. S., "afex: Analysis of Factorial Experiments," version 1.0-1, 2021 [R package]. Available: https://CRAN.R-project.org/package=afex

[34] Lenth, R., "emmeans: Estimated marginal means, aka least-squares means," version 1.4.8, 2020 [R package]. Available: https://CRAN.R-project.org/package=emmeans

[35] Tan, K., Zhang, X., and Wang, D., "Real-time speech enhancement using an efficient convolutional recurrent network for dual-microphone mobile phones in close-talk scenarios," in Proc. ICASSP 2019, 5751–5755, 2019.

[36] Teutsch, H., "Audio and Acoustic Signal Processing's Major Impact on Smartphones," IEEE Signal Processing Society, Online: https://signalprocessingsociety.org/publications-resources/blog/audio-and-acoustic-signal-processing%E2%80%99s-major-impact-smartphones, Accessed 20 June, 2022.

[37] Faber, B. M., "Acoustical measurements with smartphones: Possibilities and limitations," Acoustics Today, 13(2): 10–17, 2017.

[38] Künzel, H.J., "Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies," Int. J. Speech Lang. Law, 8(1): 80–99, 2001.

[39] Byrne, C., and Foulkes, P., "The 'mobile phone effect' on vowel formants," Int. J. Speech Lang. Law, 11(1): 83–102, 2004.