Secure account-based data capture with smartphone – preliminary results from a study of articulatory precision in clinical depression

Erin Miley[1], Felix Schaeffler[1,2], Janet Beck[1], Matthias Eichner[2], Stephen Jannetts[1]

[1] Clinical Audiology, Speech and Language Research Centre, Queen Margaret University Edinburgh

[2] FitVoice C.I.C., Musselburgh

Corresponding author contact: emiley@qmu.ac.uk

Secure account-based data capture with smartphones – preliminary results from a study of articulatory precision in clinical depression

1 Introduction

Smartphones provide an attractive way of sourcing data for linguistic researchers, potentially collecting recordings in more natural environments and thus with higher ecological validity than recordings made in a recording studio or similar setting. The hardware itself offers several advantages over standardized recording equipment, including a wide range of embedded sensors, intuitive user interface, large memory storage and computational power for recording and processing information (Dogan et al. 2017; Lane et al. 2010). In addition to this, the ubiquity and popularity of these devices allow unparalleled access to individuals and can be easily integrated into daily lives as a non-invasive, cost-efficient method of sampling and monitoring objective data. This article introduces a smartphone app for the secure collection of phonetic data and demonstrates its use in a pilot study of articulatory precision of depressed speech.

2 Reliability of smartphone recordings

Many branches of linguistic research require the use of high-quality speech data that preserves acoustic detail for analysis (Lin, Hornibrook and Ormond 2012). Recording and processing of audio data allows for the entry of error, and particular attention must be paid to the microphones embedded in smartphones (Grillo et al. 2016; Jannetts et al. 2019; Kardous and Shaw 2014; Lin et al. 2012; Manfredi et al. 2017; Uloza et al. 2015). In order to achieve the most reliable voice measures for analysis, recording equipment should maintain the original characteristics of the speech signal as much as possible. Voice measures can be influenced by many different sources of variation, including biological factors which cause natural changes in the voice over time (e.g. health status, mood or age) (Decoster and Debruyne 2000) as well as artefacts like environmental characteristics, technology and equipment (Vogel and Maruff 2014).

At present there is a diverse range of smartphone types , recent reports include 22 major vendors within the smartphone industry (Sui 2018) and device microphones and other audio components can vary considerably between these. As manufacturers tend not to publish their exact specifications, their use in linguistic research requires further comparative testing by reference to established, standardised recording equipment.

Several recent studies have compared studio and smartphone recording practices, demonstrating variable robustness of acoustic parameters under smartphone recording conditions (E Lin, Hornibrook and Ormond 2012; Vogel and Maruff 2014; Uloza et al. 2015; Grillo et al. 2016; Manfredi et al. 2017; Maryn et al. 2017; Kojima et al. 2018; Jannetts et al. 2019).

> **Commented [SF1]:** Add Jannetts et al 2019 here.
>
> **Commented [ME2R1]:**

Jannetts et al. (2019) compared smartphone recordings to studio equipment recordings and observed small but statistically significant systematic errors for a range of acoustic parameters (e.g. fundamental frequency [F0] and cepstral peak prominence [CPPS] [Heman-Ackah et al. 2003]) in sustained vowels and connected speech.. Furthermore, voice parameters like Jitter % and Shimmer % showed a random error for smartphone recordings that was deemed too high for reliable use with voice analysis.

These findings highlight the fact that the use of smartphone recordings for detailed acoustic assessment is not without problems, and care needs to be taken to determine which acoustic parameters can reliably be used when working with smartphone recordings.

3   Smartphone recordings - ethical considerations

Using smartphone applications for data collection is an efficient method of capturing recordings from participants in real-world settings, but many of the strengths of smartphone data collection also drive privacy and security concerns (Arora, Yttri and Nilsen 2014). Ethical issues arise with respect to the security of sensitive participant information, particularly the physical location of data storage and user access, as well as compliance with regulations governing personal information (e.g. the EU General Data Protection Regulation [GDPR]).

Ethical guidelines typically dictate that any research data is stored securely and accessibly by authorised researchers only. Data collected via smartphones is generally stored on a server, the location and access of which varies between server providers. To protect participant information, transfer of data from application to server should be completed via a secure connection, with identifying information being stored in encoded format. Data backup should also be enabled on separate hosts from the original dataset. Lastly, the server itself should have clearly managed access rights so that unauthorised access to the data is avoided.

Ethical considerations are especially important when recording data from health service patients. Collection of potentially sensitive or identifying information should be minimised and safeguards must be in place to protect data transfer between devices and servers. The collection of standardised speech tasks (e.g. Jannetts et al. 2019; Leemann et al. 2016; Lin, Hornibrook and Ormond 2012; Uloza et al. 2015) may reduce the possibility of recording identifying information through an app, but voice data is always potentially identifiable by those who know the individual, and extra care should therefore be taken to limit the risk of data breaches.

Smartphones have great potential to collect data from individuals at various time points (e.g. for longitudinal studies), and to collect sets of various types of data (e.g. survey data) alongside linguistic recordings. The consequent need to keep track of multiple submissions per user can be managed through accounts-based applications. However, accounts-based applications will require additional security measures to protect the integrity of user accounts and the data stored within them.

4   Fitvoice Application

The Fitvoice system has been developed for longitudinal monitoring of voices, initially from a voice health perspective with occupational voice users as target audience. We are currently piloting its extension to other health related uses, e.g. voice monitoring in clinical depression. The system is currently tested in various field and reliability studies and has around 200 test users.

The mobile app has been implemented using the cross-platform Ionic framework (ionicframework.com) and is currently available for iOS and Android in the UK Apple app store and Google Play. The app has been designed to record three types of speech prompts a) sustained vowels, b) a set of sentences targeting specific aspects of voice and c) a short connected speech sample (a modified version of the dog and duck story, [Brown and Docherty 1995]).

Alongside each recording the app prompts a short voice health questionnaire that records self-assessment of the user's voice (e.g. vocal load, current ease of use, sensations like cattargh or pain). The audio signal is recorded in 1 channel (i.e. mono). The app accesses the device microphone using

audio recording APIs that are available in both Android and iOS operating systems. The raw audio samples are stored uncompressed with 16bit signed integer resolution ("wav" format), and 44.1 kHz sampling rate. The wav files are not compressed for transmission to the backend server.

To address the various ethical and confidentiality issues raised in section 2, the app has been implemented with a range of security features. After a recording has been completed on the app, audio recordings and survey data are securely transmitted to a backend server using an SSL/TLS secure channel. The SSL/TLS secure channel guarantees full encryption of the data between device and server. This technology makes it virtually impossible to read or alter the data during transfer, e.g. when passing through intermediate data servers. The backend server runs on a virtual machine from cloud service provider linode.com. The physical location of the server is currently London, and holds various security-relevant certificates, e.g. ISO 27001. The Fitvoice system is account-based, i.e. each Fitvoice user creates a password-protected account before using the system and submitting recordings. User account management is provided by GDPR compliant authentication management provider Auth0 Europe (auth0.com). In this setup, all confidential information (e.g. user name, email and password) are stored on Auth0 servers. The Fitvoice backend only stores audio recordings, survey data and encrypted links to the Auth0 data. As individuals are potentially identifiable through voice recordings, the voice data is stored on secure servers (see above), and research participants are alerted to this fact and are asked to provide explicit consent to various types of data use (e.g. storage beyond the end of specific projects, use for perception experiments etc.).
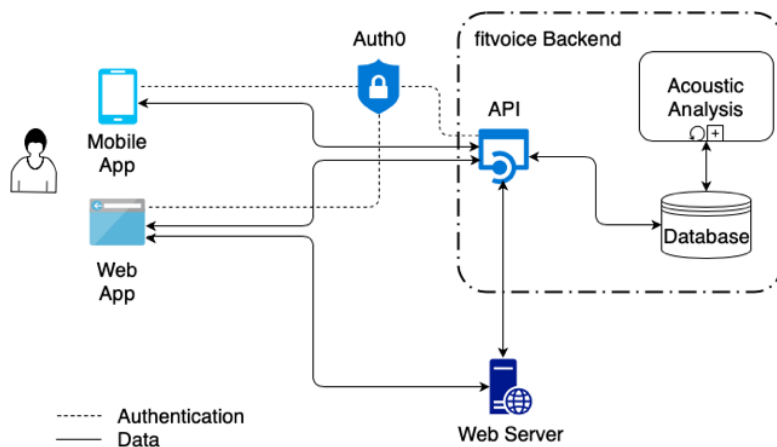

Figure 1: Fitvoice system setup, including data and authentication flow

The Fitvoice backend has APIs implemented so that other applications can access the data after authorisation. In its current implementation, the Fitvoice web app allows users to access their own data through a web interface, review survey data, play back audio recordings, and inspect a range of acoustic parameters individually for each recording as well as longitudinally over several submissions. Longitudinal analysis is facilitated by displaying a multivariate parameter (Hotelling's T^2, Schaeffler and Beck, 2017). Web app authorization is again managed by Auth0, and ordinary users can only access data from their own account. The Fitvoice web app offers two additional levels of access, administrators can access all recordings while operators can only access data for specific

projects. The Fitvoice system is currently maintained and commercialised through a Community Interest Company (Fitvoice C.I.C., registered in Scotland).
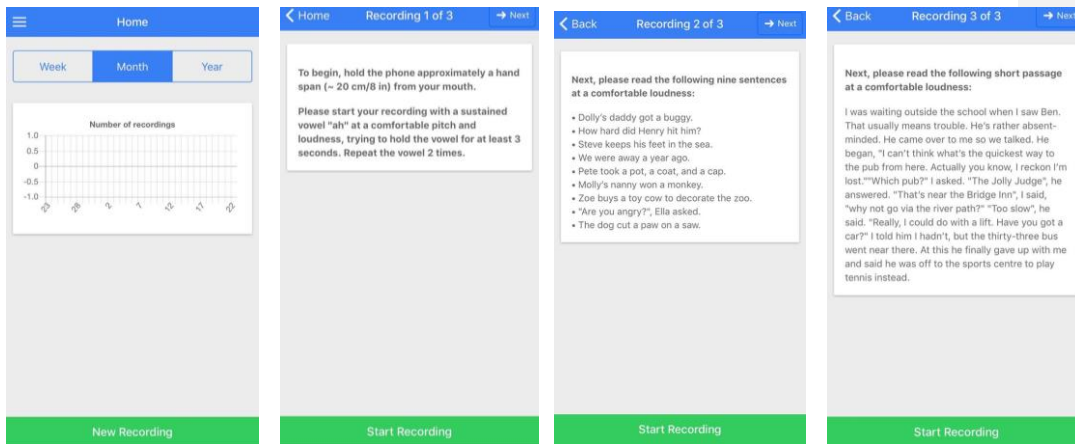

Figure 2: Screenshots of home screen and recording screen prompts for the Fitvoice mobile application.

5 Pilot Study: Speech Features in Clinical Depression

The search for more effective treatment strategies for depression includes the identification of patients who are most likely to benefit from a given therapy – this is usually done through the discovery of biological and clinical markers of response (Papakostas and Fava 2008). The human voice may be a highly sensitive indicator of physiological and psychological state; reported links between the voice and psychomotor activity point to the potential value of voice variation as a clinical marker of symptom severity and treatment response in antidepressant trials (Cummins et al. 2015; Mundt et al. 2012; Quatieri and Malyska 2012; Stassen, Kuny and Hell 1998). These studies provide preliminary evidence to suggest that the neurobiological mechanisms responsible for voice and speech production may be affected in depression, and inversely these changes may revert as an individual responds to antidepressant treatment.

From a cognitive psychology perspective, it appears that cognitive impairments associated with depression may affect an individual's phonological loop, resulting in impaired phonation and articulation (Christopher and MacDonald 2005). Research does appear to support this notion, as speech in depressed individuals demonstrates poor planning (as seen in difficulty with word choice or initiation of speech; Alpert, Pouget and Silva 2001; Mundt et al. 2012) and articulatory precision (e.g. shortened voice onset time, decreased second formant transition and increased spirantization (Flint et al. 1993) possibly due to impaired neuromuscular coordination processes (cf. Cummins et al., 2015).

Articulation can be thought of as a specific component in the chain of speech production, an integral part of the motor programming stage where the speech mechanism is prepared just prior to speaking (van der Merwe 2009). In depression, this process can be disrupted due to biological changes within the brain (e.g. atrophy and neuronal loss in the hippocampus and prefrontal cortex [Racagni and Popoli 2008] and changes in dopaminergic tone in the basal ganglia and nucleus accumbens [Pallis, Thermos and Spyraki 2001]). As such, articulation and other motor planning are affected by measurable changes within the speech signal. One such method of assessing articulatory proficiency (and consequently, levels of psychomotor retardation (Flint et al. 1993), is through

Commented [SF3]: Maybe: … resulting in impaired phonation and articulation.

Commented [ME4R3]: That makes more sense. Changed

changes in spirantization, or frication throughout speech production. Frication, or high-frequency energy visible in spectrograms, can denote a leakage of air from the oral cavity – this is particularly salient when the oral cavity should be closed (i.e. just prior to the release of stop plosives). So far there is little research examining articulatory precision and motor planning in depressive speech – a gap that our pilot study aims to fill.

5.1 Pilot Aims

The primary aim of this pilot study is to assess the effect of depression on articulatory control from a phonetic perspective. The study compares this phenomenon between depressed and healthy participant groups, taking note of durational measures and frequency of frication occurring prior to a plosive release, as a potential indicator of decreased articulatory control and psychomotor retardation.

Data collection utilises the Fitvoice app, and a secondary aim is to assess the app's ability to preserve sufficient acoustic detail in a depressive individuals' speech for analysis. This is an important consideration, as few clinical rooms provide a good recording environment; minimal furnishings and carpet often result in a high level of reverberations. The app also may aid in collecting speech data outwith clinical settings, avoiding the need for participants to travel when experiencing a depressive episode. The ability to record speech within patients' own homes is yet another advantage of smartphone data collection, but one that requires additional consideration of room size, acoustics, etc.

5.2 Method

Ethical approval for the completion of this study was granted by the Health Research Authority North East – Newcastle and North Tyneside 1 Research Ethics Committee (Ref: 18/NE/0129). Twelve speakers (6 healthy controls; 6 depressed subjects, aged 26 – 59 years) produced two trials of the phonetically balanced text 'The Dog and Duck Story' (Brown and Docherty, 1995). The vocal health and diagnostic status of each participant was determined by a pre-recording interview using a demographic questionnaire, Mini-International Neuropsychiatric Interview (MINI; [Sheehan et al. 1998]) and Hamilton Depression Rating Scale (17-item; [Hamilton 1960]) scales.

5.3 Recording Procedure

Recordings were made in a clinical interview room. Participants were asked to speak simultaneously into the smartphone and microphone (Neumann U89i model which has a flat frequency response curves up to 10 kHz within a pickup angle of ± 100° and a dynamic range of 117 dB [A-weighted]). Each was set up 20 cm away from the subject's mouth. The smartphone device used the Fitvoice app for all recordings and all microphone recordings were completed using recording and editing software Audacity™ (Version 2.3.0; Mazzoni 2018). It was considered necessary to test the application in comparison to standard equipment within a controlled setting prior to its utilisation in field recordings; this was done using a direct comparison between the Fitvoice app and microphone.

5.4 Articulatory Closure Proficiency Analysis

A total of 24 recordings (12 subjects x 2 recording sessions) underwent a phonetic markup procedure where all productions of word-initial voiceless plosives in stressed words were identified (Karlsson et al. 2014). Utterance initial plosives were excluded due to complications with separating the initial silence portion of the stop from the preceding silent segment. Additionally, plosives produced before or after a fricative were also excluded due to difficulties in segmenting phonemic boundaries.

To avoid reliability problems (as outlined in Section 2), this study relied on relatively simple spectrogram based analysis. Manual segmentation of the structure of the plosive (i.e. start and end times, release, etc.) was completed in Praat (Boersma and Weenink 2019), with the release being distinguished as the portion of the signal with the strongest amplitude and widest distribution across frequencies. Stemming from a previous acoustic analysis study investigating closure proficiency (Karlsson et al. 2014), extracted acoustic parameters included the following:

- Duration of the plosive (in ms).
- Location of the plosive release, identified as the strongest peak in signal amplitude combined with high energy across the spectrogram output.
- 'Pre-release frication' or high-frequency energy occurring just prior to the release of a plosive (see Figure 3). The number of occurrences per participant as well as length of this frication (in ms) were both recorded for further analysis.
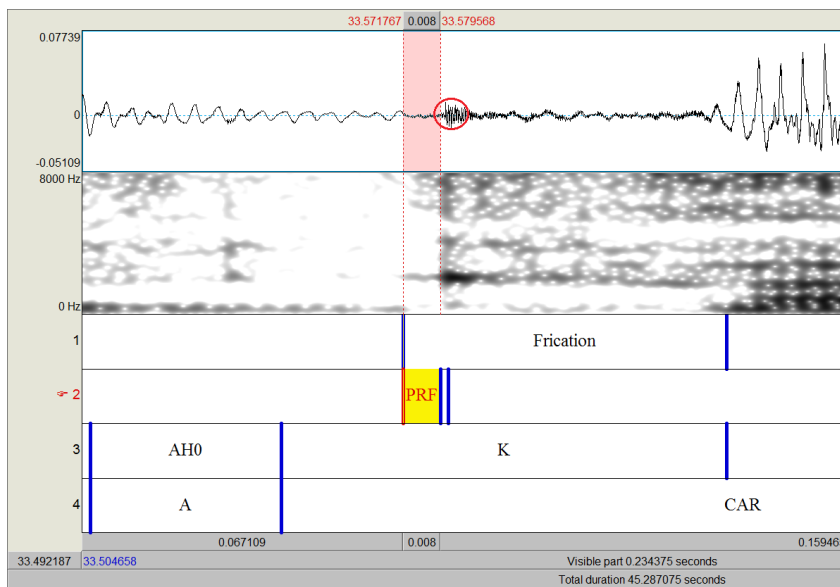

Figure 3: Example of pre-release frication (PRF) identified prior to release (circled in red).

Preliminary analysis demonstrates that the smartphone model is less affected by environmental noise and produces a 'cleaner' spectrogram for analysis than the omnidirectional microphone (See Figure 4). To quantify environmental noise levels between the smartphone and studio microphone recordings, a random subset of the smartphone recordings (7 participants x 2 sessions x 2 recording conditions) was selected to complete a signal-to-noise ratio (SNR) analysis.
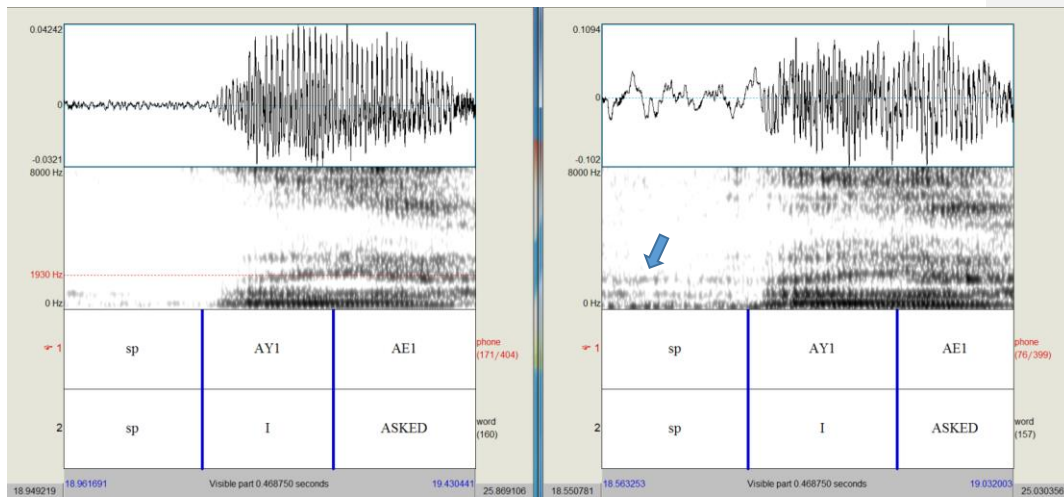
Figure 4: Comparison of spectrograms from the smartphone (left) and microphone (right) recordings made simultaneously. The spectrogram resulting from the smartphone has picked up less background noise, as evidenced by less 'noise' above the space (sp) phoneme. The blue arrow demonstrates an example of where the microphone picks up background noise (i.e. darker area) that is occurring outwith the phonemic segment; the corresponding space on the smartphone spectrogram has less of this background noise denoted by less dark colouring.

We measured signal-to-noise ratio by comparing signal intensity levels in parts of the recordings with and without speech. This analysis was performed with a Praat script that extracted intensity levels for speech and non-speech parts from the sound file, based on manually corrected forced-aligner segmentations of the audio signal, using the Penn Phonetics Lab forced aligner (Yuan and Liberman, 2008). The results provide an indication that smartphone recordings might have advantages over microphone recordings in clinical settings. Figure 5 shows that smartphone recordings had consistently higher SNR values than the microphone recordings for all participants, with an average difference of 0.5 dB. Higher SNR values reflect a lower relative amount of noise in the recordings.
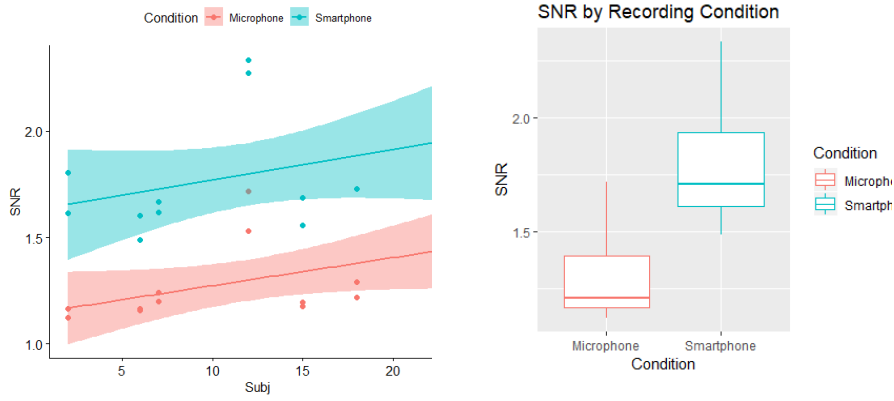
Commented [SF5]: Add ref here.

Figure 5: A comparison of Signal to Noise Ratio (SNR) between recording conditions (microphone versus smartphone). The left graph demonstrates the higher smartphone SNR values across both recording sessions for all participants. On the right, the mean difference between microphone and smartphone conditions is graphed in a boxplot.

Commented [SF6]: I think that's implicit in the above.

5.5 Exploratory Analysis and Results

All analysis was completed using R version 3.5.1 (R Core Team, 2019). One instance of plosive production for a healthy participant was excluded due to background noise, leaving a total of 120 plosive productions for the depressed group and 119 for the control group. Initially, duration of plosives was compared between groups (healthy vs. depressed) with a linear mixed effects model analysis with speaker as random factor and group as fixed factor. A likelihood ratio test using ANOVA was completed to test the significance of using group as a fixed effect. Results showed that the difference between groups was not significant -- healthy and depressed participants produced plosives of similar duration ((χ2 (1)=1.16, p=0.282; see Figure 6).
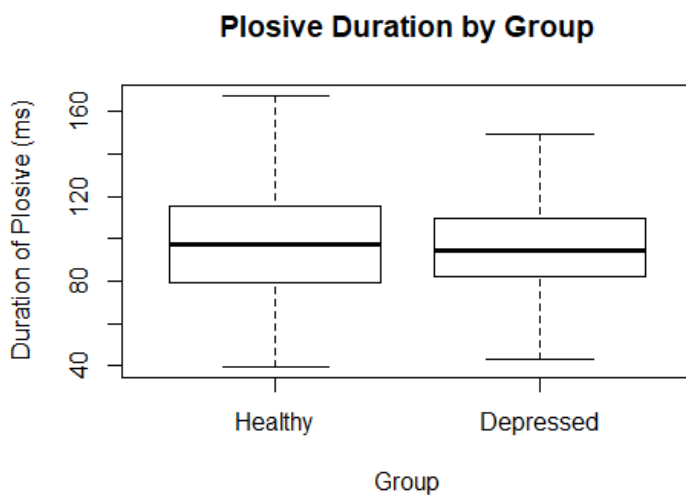
Plosive Duration by Group

Figure 6: Plosive duration(in ms) by group, demonstrating no significant group differences.

A separate analysis was completed to compare pre-release frication across groups. First, frequency of occurrence of pre-release frication was compared across groups (see Table 1). The depressed group showed significantly more plosives with pre-release frication (Fisher's exact test =239, *p* = .0082).

Table 1: Summary table of frequency of plosive productions with pre-release frication.

Group	Productions without pre-release frication	Productions with pre-release frication	Total
Healthy	91	28	119
Depressed	72	48	120
Total	163	76	239

For those plosives with pre-release frication, duration of frication was compared using a linear mixed effect model where group was included as a fixed effect and speaker was incorporated as a random effect (see Figure 7). Again, likelihood ratio test using ANOVA was completed to test the significance of using group as a fixed effect. Length of pre-release frication did not differ by group ((χ2 (1)=3.57, p=.06).
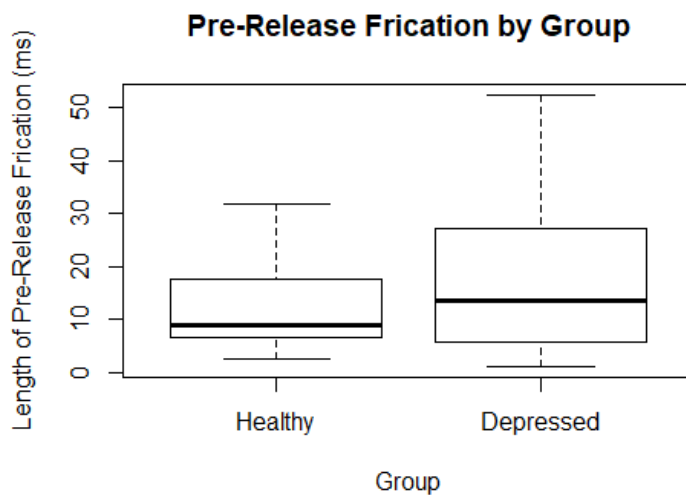
Pre-Release Frication by Group

Figure 7: Length of pre-release frication by group (in ms) demonstrating no significant group differences.

The depressed sample in this pilot study exhibited more productions of pre-release frication than the healthy controls in this sample, despite producing plosives that were nearly equal in duration. Instances of high-frequency energy preceding a stop burst may indicate a lack of adequate closure in the oral cavity, with an apparent 'leaking' of air. Karlsson et al. (2014) attribute this to a lack of 'articulator force', whereby the muscular ability of an individual results in less forceful closure action. Depressed individuals may find it more difficult to achieve adequate closure and to synchronise articulatory movements due to psychomotor retardation, a prominent feature of clinical depression (Quatieri and Malyska 2012b). When pre-release frication did occur in healthy participants, the length of these did not differ from those produced by depressed participants. Rather, the frequency of this occurrence between groups may be a distinguishing factor between diagnostic groups.

Although previous studies have found increased pausing behaviour (Alpert et al., 2001; Cannizzaro et al. 2004; Stassen, Kuny and Hell 1998; Mundt et al. 2012) and average syllable duration (Honig et al. 2014) in depressed groups, results from this pilot point to similar segment duration, at least for voiceless plosives, between clinical and healthy groups; These findings also contrast those from another previous study (Trevino, Quatieri and Malyska 2011) which demonstrated evidence for phone-specific duration measures as a potentially strong indicator of depression severity and even specific subsymptoms; in particular, slowing of production of the /t/ phoneme was one of the strongest correlators to psychomotor retardation. It may be that decreased articulatory coordination often seen in depression may lead to large variability in plosive production

Commented [SF8]: On the phonetic level I'd always avoid length and would use 'duration' instead.

and duration, making comparisons of findings between research samples difficult. Further analysis will be completed with the full sample after data collection is completed to reassess these behaviours on a wider scale.

Results suggest that utilisation of smartphone technology to capture details of the speech signal is both feasible and effective. The use of mobile interventions for depression and other mental health conditions is growing, supported by increasingly sophisticated technology (Adams et al. 2017; Depp et al. 2010; Donker et al. 2013; Reid et al. 2011). Continuing refinement of mobile-based speech monitoring of speech biomarkers has the potential to enhance the sensitivity of such approaches.

6 Recommendations and Future Directions

As mobile technology becomes more prevalent in linguistic research data collection, there is a growing need for regulation of recording conditions and parameter extraction methods. The mobile phone market is extremely dynamic, with new manufacturers, phone models and microphone settings emerging on a regular basis. In order to keep pace with these technological innovations, a potential solution could be the creation of standardised acoustic tests to evaluate hardware and software. Such tests could elucidate the unique effects of using mobile phones for data collection, such as the impact of noise cancelling microphones and beamforming.

Mobile technology facilitates the collection of supplementary information alongside recordings, a benefit that could be exploited more widely. For example, field recordings may be affected by room size, furnishings and other characteristics, so routine collection of information about recording environment may be helpful to support evaluation of the impact of such environmental factors. Fitvoice asks users to complete an integrated questionnaire alongside each recording, providing information about room size/type as well as other factors that are expected to affect vocal quality, such as level of stress, voice use and presence of cold-like symptoms. Supplementary questionnaires of this type are easily adaptable and may allow more sophisticated analysis and interpretation of data.

As ascertained from a comparison of spectrograms between microphone and smartphone conditions in our pilot, smartphones may also be a good way moving forward with recordings in clinical settings. Smartphones seem able to produce satisfactory output recordings for phonetic analysis, preserving sufficient detail to evaluate minute measures such as pre-release frication. Still, further testing will be required to understand better any artefacts that are introduced during the smartphone recording process.

For data management, account-based mobile data capture applications are useful, especially with longitudinal studies. To facilitate data sharing and support reproducibility and accuracy of findings (Houtkoop et al. 2018; Klein et al. 2018), open access databases could be created. These could be well suited to navigating around the previously discussed ethical issues (e.g. confidentiality constraints) by only providing access to extracted acoustic parameters, rather than the primary data itself. In this manner, researchers interested in sharing data may increase the transparency of their findings and also maintain the privacy of those who have been recorded.

There is no doubt that mobile technology can provide powerful tools for linguistic research, with the potential to benefit research as well as practice, e.g. in the area of health care and vocal biomarkers for mental health applications. Given the complexity of web-based technology and the very dynamic smartphone market there is an urgent need for ongoing multidisciplinary efforts and knowledge exchange to maximise the potential of mobile-mediated linguistic and phonetic research.

Commented [ME9]: No information on whether this is 'easy' so reworded.

Commented [SF10]: Not sure we need expand on this, I think the original, shorter wording is sufficiently clear.

References

Adams, Zachary W., Erin A. McClure, Kevin M. Gray, Carla Kmett Danielson, Frank A. Treiber and Kenneth J. Ruggiero. 2017. Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. *Journal of Psychiatric Research* 85. 1–14. doi:10.1016/j.jpsychires.2016.10.019.

Alpert, Murray, Enrique R Pouget and Raul R Silva. 2001. Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders* 66(1). 59–69. doi:10.1016/S0165-0327(00)00335-9.

Altman, D. G. and J. M. Bland. 1983. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician* 32(3). 307. doi:10.2307/2987937.

Arora, Shifali, Jennifer Yttri and Wendy Nilsen. 2014. Privacy and security in mobile health (mHealth) research. *Alcohol research: current reviews* 36(1). 143.

Bland, J Martin and Douglas G Altman. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8(2). 135–160. doi:10.1177/096228029900800204.

Bland, J. Martin and DouglasG Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet* 327(8476). 307–310.

Boersma, Paul and David Weenink. 2019. *Praat: doing phonetics by computer*. http://www.praat.org.

Bogue, Robert. 2013. Recent developments in MEMS sensors: a review of applications, markets and technologies. *Sensor Review* 33(4). 300–304. doi:10.1108/SR-05-2013-678.

Brown, A and G J Docherty. 1995. Phonetic variation in dysarthric speech as a function of sampling task. *European Journal of Disorders of Communication* 30(1). 17–35.

Cannizzaro, Michael, Brian Harel, Nicole Reilly, Phillip Chappell and Peter J. Snyder. 2004. Voice acoustical measurement of the severity of major depression. *Brain and Cognition* 56(1). 30–35. doi:10.1016/j.bandc.2004.05.003.

Christopher, Gary and John MacDonald. 2005. The impact of clinical depression on working memory. *Cognitive Neuropsychiatry* 10(5). 379–399. doi:10.1080/13546800444000128.

Cummins, Nicholas, Vidhyasaharan Sethu, Julien Epps, Sebastian Schnieder and Jarek Krajewski. 2015. Analysis of acoustic space variability in speech affected by depression. *Speech Communication* 75. 27–49. doi:10.1016/j.specom.2015.09.003.

Daly, Leslie and Geoffrey J. Bourke. 2000. *Interpretation and uses of medical statistics*. 5th edn. Oxford: John Wiley and Sons.

Decoster, Wivine and Frans Debruyne. 2000. Longitudinal voice changes: Facts and interpretation. *Journal of Voice* 14(2). 184–193. doi:10.1016/S0892-1997(00)80026-0.

Depp, Colin A., Brent Mausbach, Eric Granholm, Veronica Cardenas, Dror Ben-Zeev, Thomas L. Patterson, Barry D. Lebowitz and Dilip V. Jeste. 2010. Mobile Interventions for Severe Mental Illness: Design and Preliminary Data From Three Approaches. *The Journal of Nervous and Mental Disease* 198(10). 715–721. doi:10.1097/NMD.0b013e3181f49ea3.

Dogan, Ezgi, Christian Sander, Xenija Wagner, Ulrich Hegerl and Elisabeth Kohls. 2017. Smartphone-Based Monitoring of Objective and Subjective Data in Affective Disorders: Where Are We and Where Are We Going? Systematic Review. *Journal of Medical Internet Research* 19(7). e262. doi:10.2196/jmir.7006.

Donker, Tara, Katherine Petrie, Judy Proudfoot, Janine Clarke, Mary-Rose Birch and Helen Christensen. 2013. Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review. *Journal of Medical Internet Research* 15(11). e247. doi:10.2196/jmir.2791.

Flint, Alistair J., Sandra E. Black, Irene Campbell-Taylor, Gillian F. Gailey and Carey Levinton. 1993. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. *Journal of Psychiatric Research* 27(3). 309–319. doi:10.1016/0022-3956(93)90041-Y.

Grillo, Elizabeth U., Jenna N. Brosious, Staci L. Sorrell and Supraja Anand. 2016. Influence of Smartphones and Software on Acoustic Voice Measures. *International Journal of Telerehabilitation* 8(2). 9–14. doi:10.5195/IJT.2016.6202.

Hamilton, Max. 1960. A RATING SCALE FOR DEPRESSION. *Journal of Neurology, Neurosurgery, and Psychiatry* 23(1). 56–62.

Heman-Ackah, Y D, R J Heuer, D D Michael, R Ostrowski, M Horman, M M Baroody, J Hillenbrand and R T Sataloff. 2003. Cepstral peak prominence: A more reliable measure of dysphonia. *Annals of Otology Rhinology and Laryngology* 112(4). 324–333.

Honig, Florian, Anton Batliner, Elmar Noth, Sebastian Schnieder and Jarek Krajewski. 2014. Automatic modelling of depressed speech: Relevant features and relevance of gender. *Interspeech 2014*. Singapore.

Houtkoop, Bobby Lee, Chris Chambers, Malcolm Macleod, Dorothy V. M. Bishop, Thomas E. Nichols and Eric-Jan Wagenmakers. 2018. Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science* 1(1). 70–85. doi:10.1177/2515245917751886.

Jannetts, Stephen, Felix Schaeffler, Janet Beck and Cowen, Steve. 2019. Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types. *International Journal of Language and Communication Disorders*.

Kardous, Chucri A. and Peter B. Shaw. 2014. Evaluation of smartphone sound measurement applications. *The Journal of the Acoustical Society of America* 135(4). EL186–EL192. doi:10.1121/1.4865269.

Karlsson, Fredrik, Katarina Olofsson, Patric Blomstedt, Jan Linder, Erik Nordh and Jan van Doorn. 2014. Articulatory closure proficiency in patients with Parkinson's disease following deep brain stimulation of the subthalamic nucleus and caudal zona incerta. *Journal of speech, language, and hearing research: JSLHR* 57(4). 1178–1190. doi:10.1044/2014_JSLHR-S-13-0010.

Klein, Olivier, Tom E. Hardwicke, Frederik Aust, Johannes Breuer, Henrik Danielsson, Alicia Hofelich Mohr, Hans Ijzerman, Gustav Nilsonne, Wolf Vanpaemel and Michael C. Frank. 2018. A Practical Guide for Transparency in Psychological Science. *Collabra: Psychology* 4(1). 20. doi:10.1525/collabra.158.

Kojima, Tsuyoshi, Shintaro Fujimura, Ryusuke Hori, Yusuke Okanoue, Kazuhiko Shoji and Masato Inoue. 2018. An Innovative Voice Analyzer "VA" Smart Phone Program for Quantitative Analysis of Voice Quality. *Journal of Voice*.

Lane, Nicholas, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury and Andrew Campbell. 2010. A survey of mobile phone sensing. *IEEE Communications Magazine* 48(9). 140–150. doi:10.1109/MCOM.2010.5560598.

Leemann, Adrian, Marie-José Kolly, Ross Purves, David Britain and Elvira Glaser. 2016. Crowdsourcing Language Change with Smartphone Applications. *PLOS ONE* 11(1). e0143060. doi:10.1371/journal.pone.0143060.

Lin, Emily, Jeremy Hornibrook and Tika Ormond. 2012. Evaluating iPhone Recordings for Acoustic Voice Assessment. *Folia Phoniatrica et Logopaedica* 64(3). 122–130. doi:10.1159/000335874.

Manfredi, C, J Lebacq, G Cantarella and J Schoentgen. 2017. Smartphones offer new opportunities in clinical voice research. *Journal of Voice* 31(1). 111-e1.

Maryn, Youri, Femke Ysenbaert, Andrzej Zarowski and Robby Vanspauwen. 2017. Mobile Communication Devices, Ambient Noise, and Acoustic Voice Measures. *Journal of Voice* 31(2). 248.e11-248.e23. doi:10.1016/j.jvoice.2016.07.023.

Mazzoni, Dominic. 2018. *Audacity Software*. Audacity Team.

Merwe, Anita van der. 2009. A theoretical framework for the characterization of pathological speech sensorimotor control. *Clinical Management of Sensorimotor Speech Disorders*. 2nd edn. New York: Thieme Medical Publishers. https://www.researchgate.net/publication/308085824_A_theoretical_framework_f or_the_characterization_of_pathological_speech_sensorimotor_control (15 February, 2019).

Mundt, James C., Adam P. Vogel, Douglas E. Feltner and William R. Lenderking. 2012. Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry* 72(7). 580–587. doi:10.1016/j.biopsych.2012.03.015.

Pallis, Eleftherios, Kyriaki Thermos and Christina Spyraki. 2001. Chronic desipramine treatment selectively potentiates somatostatin-induced dopamine release in the nucleus accumbens. *European Journal of Neuroscience* 14. 763–767.

Papakostas, George I. and Maurizio Fava. 2008. Predictors, moderators, and mediators (correlates) of treatment outcome in major depressive disorder. *Dialogues in Clinical Neuroscience* 10(4). 439–451.

Quatieri, Thomas F and Nicolas Malyska. 2012. Vocal-Source Biomarkers for Depression: A Link to Psychomotor Activity. *Interspeech 2012*, 1059 - 1062. Portland, USA.

R Core Team. *R: A language and environment for stasticial computing*. Windows. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Racagni, Giorgio and Maurizio Popoli. 2008. Cellular and molecular mechanisms in the long-term action of antidepressants. *Dialogues in Clinical Neuroscience* 10(4). 385–400.

Reid, Sophie C, Sylvia D Kauer, Stephen JC Hearps, Alexander HD Crooke, Angela S Khor, Lena A Sanci and George C Patton. 2011. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. *BMC Family Practice* 12(1). doi:10.1186/1471-2296-12-131. http://bmcfampract.biomedcentral.com/articles/10.1186/1471-2296-12-131 (29 March, 2018).

Schaeffler, F and J Beck. 2017. Monitoring voice conditions using smartphones. *Models and Analysis of Vocal Emissions for Biomedical Applications,* 27 - 30. Firenze, Italy: Firenze University Press.

Sheehan, D. V., Y. Lecrubier, K. H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker and G. C. Dunbar. 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry* 59 Suppl 20. 22-33;quiz 34-57.

Stassen, H.H, S Kuny and D Hell. 1998. The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology* 8(4). 303–310. doi:10.1016/S0924-977X(97)00090-4.

Sui, Linda. 2018. VALUE SHARE: Global Smartphone Revenue, ASP and Profit by Vendor by Price Tier : Q4 2017. *Strategy Analytics*. https://www.strategyanalytics.com/access-services/devices/mobile-phones/smartphone/smartphones/market-data/report-detail/value-share-global-smartphone-revenue-asp-and-profit-by-vendor-by-price-tier-q4-2017#.WpC2fahuaF60.40598907602422174 (4 February, 2019).

Trevino, Andrea Carolina, Thomas Francis Quatieri and Nicolas Malyska. 2011. Phonologically-based biomarkers for major depressive disorder. *EURASIP Journal on Advances in Signal Processing* 2011(1). doi:10.1186/1687-6180-2011-42. https://asp-eurasipjournals.springeropen.com/articles/10.1186/1687-6180-2011-42 (15 February, 2019).

Uloza, Virgilijus, Evaldas Padervinskis, Aurelija Vegiene, Ruta Pribuisiene, Viktoras Saferis, Evaldas Vaiciukynas, Adas Gelzinis and Antanas Verikas. 2015. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *European Archives of Oto-Rhino-Laryngology* 272(11). 3391–3399. doi:10.1007/s00405-015-3708-4.

Vogel, Adam P. and Paul Maruff. 2014. Monitoring change requires a rethink of assessment practices in voice and speech. *Logopedics Phoniatrics Vocology* 39(2). 56–61. doi:10.3109/14015439.2013.775332.

Yuan, J. and M. Liberman. 2008. Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08.*