The Suitability of iPhone Recordings for the Acoustic Measures of Speech and Voice Quality

Emily Lin, PhD* and Jeremy Hornibrook, BSc, BMedsC, MBChB, FRACS**

*Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand
**Department of Otolaryngology, Christchurch Hospital, Christchurch, New Zealand

Abstract

This study examined the quality of iPhone recordings for acoustic measurements of speech and voice quality. A selection of acoustic measures were extracted from voice samples recorded using the “Voice Memos” application in an iPhone and compared with those derived from signals directly digitized (DD) on a laptop via a 12-bit A/D converter. Participants were 11 healthy adults, including six females and five males, aged between 27 to 67 years (Mean = 41.8 years, SD = 16.7). Four participants were native and seven were non-native English speakers. All participants were asked to read the first six sentences in the “rainbow passage” (Farbkind, 1981), one sentence at a time. Additionally, two of the participants (Participants 10 and 11) were asked to read the sentences “We saw two cars” 10 times and sustain each of the isolated vowels, /a/, /u/, 10 times. Participant 10 was a 63-year-old female native speaker of American English and Participant 11 was a 32-year-old male non-native English speaker. For Participants 10 and 11, the order of the 30 sustained vowel productions (3 vowels X 10 trials) was randomized, with three sustained vowel productions followed by one sentence.

Procedure Each participant was seated in a sound-treated room, which was monitored to ensure that the ambient noise level did not exceed 30 dB. The simultaneously recorded three signal productions (iPhone vs. directly digitized) were saved in separate digital audio files.

Experimental Measures: Based on previous research, the Spectral tilt (ST) and Vowel-based (VB) procedures described below were used to extract voice quality measures.

1. Spectral tilt (ST): amplitude difference between the highest spectral peak at 0 and 1 kHz and that between 1 and 5 kHz.
2. Vowel-based (VB): examination of the vowel speech measures, specifically those relevant to voice quality.

Participants and Participant’s Task A total of 11 healthy adults, including six females and five males, were recruited as participants. Participants aged between 27 to 67 years (Mean = 41.8 years, SD = 16.7). Four participants were native and seven were non-native English speakers. All participants were asked to read all the sentences in the “rainbow passage” (Farbkind, 1981), one sentence at a time. Additionally, two of the participants (Participants 10 and 11) were asked to read the sentences “We saw two cars” 10 times and sustain each of the isolated vowels, /a/, /u/, 10 times. Participant 10 was a 63-year-old female native speaker of American English and Participant 11 was a 32-year-old male non-native English speaker. For Participants 10 and 11, the order of the 30 sustained vowel productions (3 vowels X 10 trials) was randomized, with three sustained vowel productions followed by one sentence.

In summary, the iPhone recording method was found to be compatible with the direct digitization method for acquiring voice samples for acoustic quality analyses. In particular, F0, F1, and F2 were found to yield minimal inter-recorder variations. The results suggest that the iPhone recording method is suitable for the assessment of voice quality measures, and the future work should be focused on the investigation of the differences in the sensitivity of the microphone used, which may have resulted in the higher standard deviations in the acoustic measures. In other words, a direct comparison between measures from different recording systems should be used to investigate the differences in the sensitivity of the microphone used, which may have resulted in the higher standard deviations in the acoustic measures.

Results

Measures from iPhone and directly digitized (DD) signals were highly correlated for F1 (r = 0.98, n = 33), F2 (r = 0.98, n = 33), F0 (r = 0.96, n = 33), %Shim (r = 0.81, n = 33), voice roughness (r = 0.94, n = 33), and moderately high for H1-H2 (r = 0.77, n = 33), %Ll (r = 0.77, n = 33), SPM (r = 0.74, n = 33), and ST (r = 0.61, n = 66).

The descriptive statistics of the “normalization measures” for each measure were summarized in Table 2 and are shown in the Table 2 for the three separate data sets (one for the “rainbow passage” production by all participants and two for the sustained vowel and sentence production by Participants 10 and 11). As shown in Table 2, the mean inter-recorder NAD is consistently low (i.e., lower than 20%) for F0, F1, and F2, suggesting that these measures are least susceptible to the recording effect and more comparable than other acoustic measures.

Table 1: Results from a series of two-way (factor by voice quality measure) ANOVAs, evaluating the significance of the differences between the iPhone and directly digitized conditions, obtained from the “rainbow passage” sentence (i.e., the first six sentences in the passage) and “We saw two cars” 10 times in each of the isolated vowels, /a/, /u/, 10 times in all 11 participants.

<table>
<thead>
<tr>
<th>Measure</th>
<th>iPhone vs. DD</th>
<th>iPhone vs. DD</th>
<th>iPhone vs. DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>F2</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>F0</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>%Shim</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Conclusion

In summary, the iPhone recording method was found to be compatible with the direct digitization method for acquiring voice samples for acoustic quality analyses. In particular, F0, F1, and F2 were found to yield minimal inter-recorder variations. The results suggest that the iPhone recording method is suitable for the assessment of voice quality measures, and the future work should be focused on the investigation of the differences in the sensitivity of the microphone used, which may have resulted in the higher standard deviations in the acoustic measures.