Voice Quality of Hesitations: Acoustic Measures and VPA ratings

Amanda Cardoso¹, Paul Foulkes¹, Peter French¹², Philip Harrison¹², Vince Hughes¹, Colleen Kavanagh¹³, and Eugenia San Segundo¹

¹Department of Language and Linguistic Science, University of York, UK
{amanda.cardoso|paul.foulkes|peter.french|vincent.hughes|colleen.kavanagh|eugenia.sansegundo}@york.ac.uk
²J P French Associates, York, UK
philip.harrison@jpfrench.com
³RCMP, Canada
collen.kavanagh@rcmp-grc.gc.ca

Voice quality (VQ) provides a great deal of speaker-specific information and is widely regarded as a useful tool to discriminate speakers by forensic speech scientists (Gold and French 2011, San Segundo et al. in press). However, the range of anatomical, motoric, and socio-cultural factors that affect a speaker’s VQ makes it difficult to analyse (Kreiman et al. 1993). All specific vocal settings are intermittent and not permanent, in that they are manifested only in certain segments (e.g. phonation is only manifested in voiced segments). Given these complications, is it possible to reduce auditorily-assessed VQ to a set of acoustic measures?

The acoustic analysis of VQ metrics is complicated by the common practice to apply those metrics to sustained vowels, which are infrequent in natural speech and only partially representative of a speaker’s vocal range. However, an exception to this is vocalic hesitations (uhhs and ums), which occur frequently in speech, are often longer than lexical vowels, and often occur adjacent to unfilled pauses. The acoustics of hesitations also provide good speaker discriminatory information (Tschäpe et al. 2005, Hughes et al. 2016). Therefore, hesitations may help to answer whether and which acoustic correlates of VQ can be compared with perceived VQ settings.

San Segundo et al. (in press) produced auditory ratings of VQ settings for each speaker in the DyViS corpus (Nolan et al. 2009), using a modified version of the vocal profile analysis (VPA) scheme (Laver et al. 1981). Our data consists of 33 speakers who received the highest and lowest scalar degree ratings across the VQ dimensions and who, therefore, represent extreme speakers within this corpus. We focus here only on creaky, breathy, and neutral (i.e. VPA settings rated 0).

Reliable acoustic measures to distinguish breathy and creaky phonation have been widely debated (e.g. Gobl and Ni Chasaide 1992, Klug 2014, Keating et al. 2015, Klug et al. 2018). Breathy and creaky voice have lower f0 values than modal voice. Creaky voice is also claimed to have higher jitter, i.e. frequency variation between consecutive f0 cycles. Breathy voice has higher positive values of H1-H2 (amplitude difference between first and second harmonic) and H1-A3 (amplitude difference between first harmonic and third formant) than creaky voice. Acoustic measures of f0, jitter, H1-H2, and H1-A3 were taken from the sustained vowel portions for each of the 33 speakers’ hesitations (mean/median=18, range=11–39) from DyViS Task 1.

We found that some acoustic measures do align with the speakers’ VPA ratings (Table 1). Speakers who are perceived to have creaky phonation tend to have lower f0 than other speakers (Figure 1), and higher jitter (Figure 2). Conversely, H1-H2 and H1-A3 do not correspond well to VPA ratings, with little difference between speakers classified as breathy and creaky, and breathy values lower than those for creaky - contra previous studies (e.g. Ni Chasaide and Gobl 1997, Esposito 2010).

VQ of hesitations are comparable to overall VQ, but there are also VQ patterns specific to hesitations, which further supports the use of hesitation for speaker discrimination.
Table 1. Average acoustic measures of voice quality for speakers rated as creaky, breathy or neutral.

<table>
<thead>
<tr>
<th>VPA VQ</th>
<th>f0 (Hz; mean)</th>
<th>Jitter (mean)</th>
<th>H1-H2 (dB; mean)</th>
<th>H1-A3 (dB; mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creaky</td>
<td>99</td>
<td>0.92</td>
<td>3.98</td>
<td>9.64</td>
</tr>
<tr>
<td>Breathy</td>
<td>109</td>
<td>0.85</td>
<td>3.27</td>
<td>9.34</td>
</tr>
<tr>
<td>Neutral</td>
<td>109</td>
<td>0.82</td>
<td>1.69</td>
<td>6.10</td>
</tr>
</tbody>
</table>

Figure 1. Distribution of f0 (Hz) values across speakers rated as creaky, breathy, or neutral.

Figure 2. Distribution of jitter values across speakers rated as creaky, breathy, or neutral.

References


