**Database overview for the Perceptual Voice Quality Database (PVQD) v.4 (September 2022 update)**

Patrick R. Walden, Ph.D., CCC-SLP
pwalden@monmouth.edu

## Introduction and General Methods

This database was created through generous funding from *The Voice Foundation*'s Advancing Scientific Voice Research Grant and contains voice samples which have been rated by experienced voice professionals (at least 3 different raters with a minimum of 2 years' clinical experience) in order to provide educators with standardized materials to better train pre-service clinical voice professionals. It contains 296 audio files consisting of the sustained /a/ and /i/ vowels and the sentences from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). All recordings were made in a quiet clinical environment (not sound treated) using a head-mounted condenser microphone at a 6-centimeter distance from the corner of the mouth and the Computerized Speech Lab (CSL) using 16-bit quantization and a sampling rate of 44.1k. Audio recordings have been edited as best as possible to remove all clinician instructions. However, please listen to and look at each file carefully just in case there was simultaneous clinician-client talk.

A published description of the database's characteristics can be found by accessing the article:
Walden, P. R. (in press/online first in 2020). Perceptual Voice Qualities Database (PVQD): Database Characteristics. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2020.10.001.

Listeners rated approximately 50 files each and each file was rated twice for reliability measurement (for a total of approximately 100 ratings per rater). Raters used a computer to listen to the samples and rate voice quality via a web-based system that included custom-made electronic scales for the CAPE-V (Kempster, 2007) and the GRBAS (Hirano, 1981) using Qualtrics survey software. Listeners rated each file on a 100-point visual analogue scale (VAS) to mimic the paper-based CAPE-V protocol. **Please note that severity markers (mild, moderate, severe) were not included on the 100-point VAS to avoid influencing the concurrent rating using the GRBAS scale.** Raters were urged to rate the samples over several days to avoid fatigue.

## Results

Data include a spreadsheet containing the name of each audio file, the age and sex of the speaker, and some of the audio files contain a yes/no indication of patient voice complaint. This file is labeled "Demographics.xlsx". The "Y" and the "N" indicate an extant voice complaint (per the speaker) and not necessarily a diagnosed dysphonia (or the absence of dysphonia). Individuals with no voice complaints were purposively included in the database to achieve a sizeable representation of normal variability. In cases where a dysphonia diagnosis was extant and the etiology had been established, the etiology is provided in the spreadsheet. For example, a speaker may be labeled with a "Y" which means the speaker presented to the clinical environment with a voice complaint. If an etiology is not listed, the dysphonia may not have been diagnosed and/or the etiology may not have been determined. Similarly, a speaker labeled "N" only means that the speaker did not express concern/complaint with their voice upon collection of the voice sample.

Other spreadsheets are also included. One spreadsheet (Ratings_both_scales.xlsx) provides all raw data for the CAPE-V and the GRBAS scales (each scale is located in a different tab in this spreadsheet). Other spreadsheets separate out each voice quality by scale. In these spreadsheets, raw rater data are included as well as calculated average ratings, minimum/maximum values, and the standard deviation of ratings

across all ratings. According to Nagle (2016), ratings within one centimeter (10 units) on the CAPE-V is considered to be adequate agreement. Please take the standard deviation of ratings into account when looking at each file individually.

Severity "categories" are also included (normal, mild, moderate, severe) for the GRBAS scale. Because averages across all raters were used to categorize each sample, the following thresholds were used: 00.5=normal; 0.6-1.5=mild; 1.6-2.5=moderate; 2.6-3=severe. Any user is free to change these thresholds as necessary. To help a user change the thresholds, I have provided the formula I used to automatically label each sample in the spreadsheet. Formula:
“=IF(L2<0.6,"Normal",IF(L2<1.6,"Mild",IF(L2<2.6,"Moderate",IF(L2<3.1,"Severe"))))” -please remove quotation mark at beginning and end of the formula.

**Rater (Listener) Reliability**
Inter- and intra-rater reliability calculations are provided for the sample as a whole using Intraclass Correlations. Most files were rated by three experienced voice clinicians. Some files were rated by four experienced voice clinicians. Intra-rater reliability for audio files rated by four listeners was calculated by randomly selecting three of the four raters for each file since the number of raters was not equal across the sample.

**Interrater Reliability Results Results for CAPE-V:**
Overall Intraclass Correlation for Interrater reliability = .**860** (averages used as ratings)

Intraclass Correlations by Feature (averages used as ratings):

| | |
|---|---|
| CAPE-V Severity: | .918 |
| CAPE-V Roughness: | .789 |
| CAPE-V Breathiness: | .827 |
| CAPE-V Strain: | .829 |
| CAPE-V Pitch: | .856 |
| CAPE-V Loudness: | .870 |

**Results for GRBAS:**
Overall Intraclass Correlation for Interrater reliability = **.859** (averages used as ratings)

Intraclass Correlations by Feature (averages used as ratings):

| | |
|---|---|
| GRBAS Grade: | .911 |
| GRBAS Roughness: | .787 |
| GRBAS Breathiness: | .844 |
| GRBAS: Asthenia: | .843 |
| GRBAS Strain: | .845 |

**Intrarater Reliability Results Results for CAPE-V:**

Overall Intraclass Correlation for Intrarater reliability = **.912** (assuming averages used)

Intraclass Correlations by Feature (assuming averages used):

| | |
|---|---|
| CAPE-V Severity: | .943 |
| CAPE-V Roughness: | .896 |
| CAPE-V Breathiness: | .911 |
| CAPE-V Strain: | .908 |
| CAPE-V Pitch: | .878 |
| CAPE-V Loudness: | .905 |

Overall Pearson Correlation between Trials 1 & 2 = **.839**

Pearson Correlations between Trials by Feature:

| | |
|---|---|
| CAPE-V Severity: | .890 |
| CAPE-V Roughness: | .814 |
| CAPE-V Breathiness: | .833 |
| CAPE-V Strain: | .828 |
| CAPE-V Pitch: | .772 |
| CAPE-V Loudness: | .824 |

**Results for GRBAS:**

Overall Intraclass Correlation for Intrarater reliability = **.889** (assuming averages used)

Intraclass Correlations by Feature (assuming averages used):

| | |
|---|---|
| GRBAS Grade: | .905 |
| GRBAS Roughness: | .846 |
| GRBAS Breathiness: | .884 |
| GRBAS: Asthenia: | .892 |
| GRBAS Strain: | .862 |

Overall Pearson Correlation between Trials 1 & 2 = **.800**

Pearson Correlations between Trials by Feature:

| | |
|---|---|
| GRBAS Grade: | .827 |
| GRBAS Roughness: | .734 |
| GRBAS Breathiness: | .793 |

GRBAS: Asthenia: .804
GRBAS Strain: .757

**Note**

The audio files can be downloaded directly from this database. To help users download all files simultaneously, please use the following link to my online storage. Given that technology changes rapidly, the link may not work in perpetuity in which case files will need to be downloaded directly from this database. Link: https://app.box.com/s/yj4o8zzxt45e8yqleqpwlbb69jo25kq5

**References**


1. Hirano M. *Clinical Examination of Voice*. Springer-Verlag; 1981.
2. Kempster G. CAPE-V: Development and Future Direction. *Perspect Voice Voice Dis*. 2007;17(2):1113. doi:10.1044/vvd17.2.11
3. Nagle KF. Emerging Scientist: Challenges to CAPE-V as a Standard. *Perspectives of the ASHA Special Interest Groups*. 2016;1(3):47–53.


Please direct questions to me at pwalden@monmouth.edu.


Patrick R. Walden, Ph.D., CCC-SLP
Monmouth University
Department of Speech-Language Pathology
West Long Branch, NJ, USA