Sensor Data Challenge: Speech quality sensor

Annette Scherpenzeel, Jeny Philip, Marina Aoki

Aim

This sensor would be a device intended for a one-time measurement of speech quality that has the potential to be used at the time of the SHARE interview.
The sensor should be able to detect:

- The respondent's ability to speak fluently and coherently
- Slow speech, slurring, etc.
- Abnormal, varied speech rhythm

The measurement is not intended to judge the respondent's knowledge or their elocution/language. Rather it would be used to evaluate their articulation skills and whether they have any difficulties in speech or any deficiency in speech quality. As such the actual content/topic of what the respondent says will not be assessed.

Relevance

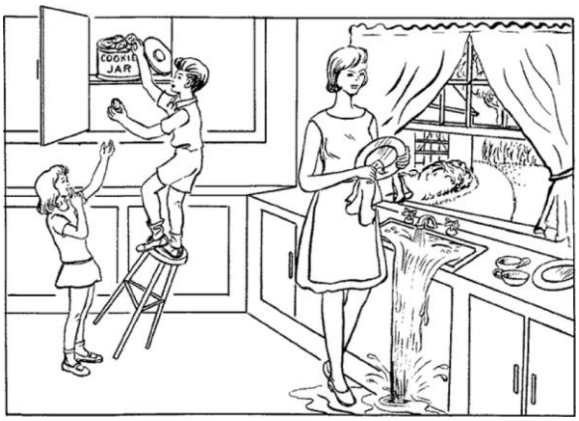
The Survey of Health, Ageing and Retirement in Europe (SHARE) is a research infrastructure for studying the effects of health, social, economic and environmental policies over the life-course of European citizens and beyond. From the time of its inception in 2004, 480,000 in-depth interviews with 140,000 people aged 50 or older from 28 European countries and Israel have been conducted making it the largest pan-European social science panel study providing internationally comparable longitudinal microdata on public health and socio-economic living conditions. For more reading about SHARE, see: The Survey of Health, Ageing and Retirement in Europe (SHARE): Home (share-project.org). In an ageing Europe, the increase in dementia is a major health theme. SHARE investigates the early onset of cognitive decline and relates that to life circumstances and life style. Speech impairment could be indicative of medical conditions and could also be correlated with stroke, brain disease, neuromuscular disease, cognitive impairment, alcohol/drug abuse or overall poor health.

Measurement Procedure

Participants are encouraged to use one or more of the measurement options below or any other alternative they may come up with, that is equally effective.

1. The sensor could pick up the speech patterns in reading a standard text during the interview so as to provide an indication of true verbal fluency.
2. Another option would be to ask the respondent to deliver a few sentences extemporaneously on a topic chosen from a list of 4 or 5 topics or of their own choice.
3. A picture description task:

One of the most used tools to elicit discourse is the Cookie Theft picture (CTP) from the Boston Diagnostic Aphasia Examination. This picture shows two children trying to remove cookies from a jar placed on top of a cupboard as their mother is washing the dishes. The picture depicts a familiar domestic scene with basic key vocabulary learned in childhood with distinct characters, time, and place contrasts.



Desirable Sensor Features

- The speech quality sensor should be portable, easy to use and cheap to facilitate implementation on a large scale in social science surveys.
- Recording can be done with a smartphone or other speech recording hardware capable of high-quality audio recordings.
- After recording, the audio data can be processed with the help of libraries/corpora that can be found online (metrics every 20-30ms; many different variables are measured)
- Participants are encouraged to look into relevant literature in order to find which variables are the most important/most relevant. The next step after selection would be to analyze these variables.
- The final step, in which the data is analyzed, should provide an indicator of speech fluency/coherence. A certain degree of subjectivity cannot be avoided but as many measures as possible should be implemented to prevent the analysis from being too subjective in nature. Considerations have to be made to account for GDPR regulations while handling personal data..
- A graphical representation of the data in order to clarify the results could be a first output for interpretation.
- This can be followed by other measures of Speech Fluency which could include but need not be restricted to those outlined in the next section. This should be stored in an interoperable database (perhaps in a de-identified form) so as to be available for further analysis by researchers.
- Note: There does not yet appear to exist any application of this type which can process this data directly on the mobile phone

Optional Features

- An added optional bonus would be the possibility to extend it for use across more than one language.
- The sensor may also pick up and process facets of the sound waves like amplitude, pitch, frequency and phase if these are needed to compute relevant speech quality measures.

Additional Tips: Speech Quality Metrics

- It would be ideal if the sensor could provide a measure of the Articulation Index (AI) where an AI of less than 0.3, generally suggests unintelligible speech and one over 0.7 indicates excellent intelligibility.
- We could otherwise use a similar variable that is considered a successor to the Articulation Index, the Speech Intelligibility Index (SII).

  The SII is interpreted as the proportion of total speech information available to the listener's ear for a given speech material.

  The general formula for calculating the SII is:

$$SII = \sum_{i=1}^{n} I_i A_i$$

In this formula,

- *n refers to the number of individual frequency bands used for the computation.*
- *The values for I_i, also known as the frequency importance function (FIF), are based on specific speech stimuli, and when summed across all bands are equal to approximately 1.0.*
- *A_i is the band audibility function, which represents the audibility of speech signal in the I'th frequency band and the value of Ai ranges between 0 and 1.*
- *Therefore, the SII is constrained to be between 0 and 1.*

Participants are welcome/encouraged to do their background research into the most accurate way to calculate the SII.

- More detailed measurement indicators (optional) are mentioned below.

Reference:

1. *Themistocleous C et al (2020) Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. PLOS ONE 15(7): e0236009. doi:10.1371/journal.pone.0236009.*
2. *Hornsby, Benjamin. (2004). The Speech Intelligibility Index: What is it and what's it good for?. The Hearing Journal. 57. 10-17.*

<u>Measures of Speech Fluency</u>

i. *Average Syllable Duration*: Is the mean syllable duration estimated as a measure of the overall speaking time divided by the number of syllables (3).

$$Averaged\ Syllable\ Duration = \frac{Overall\ Speaking\ Time}{Number\ of\ Syllables} \tag{3}$$

ii. *Articulation Rate*: Articulation rate considers phonation time, which is a measure of phonation times and excludes pauses and silences (4).

$$\text{Articulation rate} = \frac{Number\ of\ Syllables}{phonation\ time} \tag{4}$$

iii. *Speech Rate*: Is a measure of the number of syllables divided by the overall duration, which includes pauses and silences (5):

$$\text{Speech rate} = \frac{Number\ of\ Syllables}{Total\ Duration} \tag{5}$$