Spectral Slope Measurements in Emotionally Expressive Speech

Lucas Tamarit, Martijn Goudbeek & Klaus Scherer
Introduction

- Spectral slope is a measure of voice quality
- Perceptually, voice qualities include harsh, tense, breathy, creaky voice and whisper
- Reflected in the intensity of the harmonics and more generally in the shape of the power spectrum
Plan

• Overview of existing voice quality indices used in emotion discrimination
• Integration of speaker-dependent information in voice quality indices
• Proposition of new indices
• Results & discussion
• Future plans
Voice quality indices in emotion discrimination

- Mostly global, power spectrum oriented measures
- Characterizations of the long-term average spectrum (LTAS)
- Share the idea of splitting the spectrum in a high and a low frequency range (notion of pivot frequency)
Widespread voice quality indices

- Hammarberg Index
- Drop-off of the spectral energy above 1000Hz
- Relative amount of energy in the high-(above 1000Hz) versus low-frequency range of the spectrum (up to 1000Hz)
Example: The Hammarberg Index

- Defined as the difference between the maximum energy in the 0..2kHz and in the 2..5kHz band
Pivot Frequency

• The lower part of the spectrum mainly contains phonetic information (F1 and F2)

• The higher part of the spectrum is known to convey more information about the voice quality
Existing voice quality indices

- **Pros**
  - Concise synthesis of the data of whole speech segments
  - Directly related to the spectral slope
- **Cons**
  - Do not take any speaker dependent information into account
  - Rough measures, do not reflect much details of the spectrum (basically only a slope)
Integrating speaker-dependent information

- Replace the fixed 2kHz pivot by a speaker-dependent pivot
Speaker dependent Pivot

• Three approaches
  – Pitch based
  – Formants based
  – Spectral Balance based
• All three approaches are based on neutral reference recordings.
Neutral reference recordings

- For each speaker, we use reference recordings of the sustained vowel /a/ when expressing interest.
- For more information, see Dr. Goudbeek’s poster presentation « Acoustic Profiles in Emotion – The GeMEP Corpus ». 

![Image of neutral reference recordings]
Pitch based pivot

• The fundamental frequency of all reference recordings is extracted then averaged.
• Any multiple of that reference frequency can be used as the pivot.
• Here, we chose the 10th harmonic as the basis for further analysis.
Formants based pivot

- The second formant of all reference recordings is extracted then averaged.
- The choice of F2 among other formants refers to the aforementioned considerations about phonetic and voice quality content.
Spectral balance based pivot

• Defined as the frequency separating the whole spectrum in two parts having a fixed ratio of energy $\lambda$:

• A value of $\lambda = 50$ yields pivot frequencies typically located between 1kHz and 2kHz.
Spectral Slope Measurements

• Once a pivot has been chosen, several spectral slope measurements can be derived from it.

• Three approaches:
  – Modified Hammargberg Index
  – Spectral Slope
  – Spectral drop-off curvature
Modified Hammarberg Index

- Same as the original Hammarberg Index, replacing the fixed 2kHz value by the chosen pivot:

- The upper limit is pushed to 8kHz.
Spectral Slope

• Apply simple linear regression to the considered spectral portion

\[ G = af + b \]

where \( a \) is the parameter of interest
Spectral drop-off curvature

• Fitting a decaying exponential to the LTAS

\[ G = Ae^{\alpha f} + b \]

(with \( \alpha < 0 \))

where \( \alpha \) is the parameter of interest
Summary

• Three pivots
  – Pitch based
  – Formants based
  – Spectral balance based

• Three spectral slope measurements
  – Modified Hammarberg Index
  – Spectral Slope
  – Spectral drop-off curvature
Results

• Indices means for all three pivots:

- Hammarberg Indices
- Spectral slope (linear)
- Spectral drop-off curvature (decaying exponential)
Results

- Significant effect of emotions on all dependent variables (MANOVA on coefficients): Hamml, mHamml, linear, exponential
- Highly similar pattern across pivots
- High correlations between linear measures
- Exponential least important in discriminant function
Discussion

• Direct comparison of different characterizations of the LTAS shape of emotionally expressive speech
• Traditional Hammarberg Index shows good performance against new indices propositions
• Spectral slope curvature might not be related to affect
Future plans

• Compare the present results with analyses on a speaker independent basis.
• Disentangle speaker specific and emotion specific effects on the shape of the spectrum
Thanks for your attention.