

HQSYN16 - Task #4176

Task # 3677 (New): RA3b - Phonetically justified parameters (spectral tilt, ...)

Task # 3970 (Closed): Formant-based join cost computation

Quick analysis of formants use failures

07.03.2017 14:42 - Tihelka Dan

Status:	Closed	Start date:	07.03.2017
Priority:	Normal	Due date:	31.03.2017
Assignee:	Tihelka Dan	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	RA3: Phonetically justified parameters for speech synthesis		
Description			
<p>The experiment with the replacement of MFCC by Formant trajectories was not very successful (well, it was not successful at all ...).</p> <p>Thus, we have selected clear examples, where the use of formant trajectories was clearly worse than the use of MFCC. These are attached in <i>examples.zip</i> stored on the experiment wiki. Could you, please, look at the examples and try to determine what went wrong? Also, if you have any suggestions of experiment re-design, (i.e. measure trajectories for particular phones only?), suggest it here.</p> <p>If required, we can provide the trajectories of formant frequencies used.</p> <p>For the detailed description see experiment wiki, or write me an e-mail.</p>			

History

#1 - 17.04.2017 12:00 - Skarnitzl Radek

- File *notes.docx* added

- Status changed from Assigned to Resolved

- % Done changed from 0 to 100

I have listened to the MFCC and formant versions of the sentences, and the results are not surprising (see the attached document). Formant frequencies are only one aspect in the spectral domain in which a discontinuity may arise. The distribution of energy in the spectrum is not accounted for, and differences in that also leads to audible discontinuities. MFCC coefficients, on the other hand, should account for the frequencies of the main spectral components (formants), as well as spectral energy distribution. It would therefore be interesting to compare MFCCs against a combination of formant frequencies and spectral slope information.

#2 - 17.04.2017 12:00 - Skarnitzl Radek

- Assignee changed from Skarnitzl Radek to Tihelka Dan

#3 - 18.04.2017 09:17 - Tihelka Dan

Thank you for the revision. We have also suspected that the problems may be related to the discontinuities in higher spectral frequencies not handled by *formants* which are somehow taken into the consideration when using MFCC. The surprising fact is the discontinuities in F0, as F0 was included in the cost computation (although as a "static" feature only) - see the description [of F0 computation](#) (and also [#3674](#)).

So, if you suggest to add a spectral slope measure, we can experiment with it, but **we need it to be computed from the files somehow** (i.e. Praat?).

Let me also note that we can join MFCC with formant frequencies in the CC for experimental purposes. However, in production version of ARTIC, I would rather look for a simpler scheme, since the computation of both formants and MFCC costs is quite expensive.

#4 - 20.04.2017 11:38 - Skarnitzl Radek

We can try experimenting with some spectral slope measures. These typically compare energy in specific spectral bands. I would try the measure which Honza Volín introduced, which calculates the ratio of a low and high frequency band energy. The low band is defined as 350-1100 Hz (so it excludes the band corresponding to F0), the high band includes the 2300-5500 Hz frequency range (excluding the F2 range, which is important to convey phonemic differences).

In Praat, you can use the command `Get band energy difference... 350 1100 2300 5500`.

And of course, trying to combine this with formants or MFCCs is a good idea!

#5 - 02.06.2017 09:30 - Tihelka Dan

Adding spectral slope will continue in [#4211](#). Once results are obtained, this task will be used for re-analysis.

#6 - 20.01.2020 08:49 - Tihelka Dan

- *Status changed from Resolved to Closed*

Files

notes.docx	14 KB	17.04.2017	Skarnitzl Radek
------------	-------	------------	-----------------