HQSYN16 - Task #3972

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

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

Use formants instead of MFCCs in join cost computation

04.07.2016 23:02 - Matoušek Jindřich

Status:	Closed	**Start date:**	18.07.2016
Priority:	Normal	**Due date:**	14.09.2016
Assignee:	Matoušek Jindřich	**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:	RA3: Phonetically justified parameters for speech synthesis		

Description

Use formants from #3971 and replace MFCCs.

Compute both static and dynamic contours around a concatenation point (similarly as for F0).

Alternatively, formants estimated by *ESPS tool* (used e.g. in *Wavesurfer*) could be used. For the voice Jan (AJ), the formants are stored in ARTIC/Projects/cz/spkr_AJ/data/non-mastered/zkracene-pauzy/param/formants/formants_f10_o12_w25_i05. More details about the format of text files with formant values are given in ARTIC/Projects/cz/spkr_AJ/data/non-mastered/zkracene-pauzy/param/formants/README.txt

Related issues:

Blocked by HQSYN16 - Task #3999: Compute the formats	**Closed**	**09.08.2016**	**14.08.2016**
Follows HQSYN16 - Task #3971: Praat script to compute formants	**Closed**	**04.07.2016**	**17.07.2016**

History

#1 - 04.07.2016 23:04 - Matoušek Jindřich

- Follows Task #3971: Praat script to compute formants added

#2 - 25.07.2016 13:44 - Matoušek Jindřich

Tomáš Bořil added Praat scripts to compute formants - see #3971

#3 - 09.08.2016 14:22 - Tihelka Dan

- Blocked by Task #3999: Compute the formats added

#4 - 12.08.2016 17:03 - Tihelka Dan

I have hacked up the use of formats instead of MFCC in the concatenation cost (in addition to F0 and energy, which are computed as before).

Distances

For further explanation, expect *F1{t}, _F2{t}, F3{t}* and *F4{t}* being (z-score normalized) values of formants at time *t*. When *t = eL*, than it describes the time *nearest to* the end of the left concatenated diphone, and *t = bR* describes the time *nearest to* the beginning of the concatenated right diphone; i.e. we always examine the difference of *eL* to *bR* features (being taken from a phone center). When the concatenated diphones neighbored in the corpus, then it is ensured that *eL = bR*.

Now it is possible to experiment with 3 computation schema:

- **absolute difference of formants and their slopes:**

  cost = (abs(*F1{eL} - F1{bR}*) * *W1* + abs(*F2{eL} - F2{bR}*) * *W2* + ... + abs(*F4{eL} - F4{bR}*) * *W4* + abs(*S1{eL} - S1{bR}*) * *W1* + abs(*S2{eL} - S2{bR}*) * *W2* + ... + abs(*S4{eL} - S4{bR}*) * *W4* + *F0-cost* + *energy-cost*) / (*W1* + *W2* + *W3* + *W4* + 1 + 1)

  where *Sn* is slope of the *n*-th format computed from sequence of [*Fn{t-4}, Fn{t-3}, Fn{t-2}, Fn{t-1}, Fn{t}, Fn{t+1}, Fn{t+2}, Fn{t+3}, Fn{t+4}*] formant values, *t = eL* or *t = bR*.

- **Euclidean distance of the formant contour:**

  cost = (euclid(*C1{eL}, C1{bR}*) * *W1* + euclid(*C2{eL}, C2{bR}*) * *W2* + ... + euclid(*C4{eL}, C4{bR}*) * *W4* + *F0-cost* + *energy-cost*) / (*W1* + *W2* + *W3* + *W4* + 1 + 1)

where Cn{t} = [*Fn{t-4}, Fn{t-3}, Fn{t-2}, Fn{t-1}, Fn{t}, Fn{t+1}, Fn{t+2}, Fn{t+3}, Fn{t+4}*] is the sequence of formant values

- **Mean absolute difference of the formant contour**, which is the same as the previous, but except the *euclid(Cn{eL}, Cn{bR})* distance we use *mean(abs(Cn{eL} - Cn{bR}))*

For all the experiments, the weights were set to: *W1* = 0.8, *W2* = 1.0, *W3* = 0.7, and *W4 = 0.4*. Also, **there is no bandwidth considered now!**

Formants

There are 2 versions of formant estimations: ESPS+PRAAT (which gives us 6 possible experiments)

What next

Now there are several questions to answer:

1. how to design the experiment
2. what text use to experiment
3. which distance computation scheme to use
4. which formants to use (I would vote for PRAAT)

Any thoughts?

#5 - 12.08.2016 17:04 - Tihelka Dan

- Status changed from New to Feedback

#6 - 05.09.2016 13:00 - Tihelka Dan

- Assignee changed from Tihelka Dan to Matoušek Jindřich

#7 - 27.09.2016 14:29 - Tihelka Dan

As the first experiment, I will use PRAAT-computed formants with all the distance computations (i.e. ABS, EUCL, SLP) to get the log of units usage. That can be compared to the units used for baseline system (without formants in CC) and the most differing unit sequences can be further analysed. This is the same approach as in #3941.

Note that the cost is computed through the window of 9 values, which is about 60msec of signal (for frame length 20msec and shift 5msec). **Is that enough to capture the interesting formant properties?** Or should the region be extended?

#8 - 02.06.2017 09:19 - Matoušek Jindřich

- Status changed from Feedback to Closed

Replacement of MFCCs with formant frequencies was not very successful, see #4176. Other (supplementary) measures (like spectral slope) will be searched for.