Pitch Tracking

Comparison of different algorithms for Pitch Tracking

Bente Reichardt, Oscar Julian Bedoya Zapata,
Khooshal Saurty, Erik Fließwasser

Speech Technology, Department of Informatics, University of Hamburg

Introduction

This lab experiment investigates on the quality of different pitch tracking algorithms for speech recognition. The pitch tracking is applied on the speech sound file database PTDB-TUG containing records of sentences spoken by female and male speakers. The estimated pitches are compared to the given reference data in order to come up with a comparison of the quality of the pitch tracking algorithms AMDF, ESPS, and Praat. The implementation of this lab uses different tools and programming languages. Each algorithm is processed by particular configurations (Wavesurfer/Snack, Praat Script) whereas the evaluation of the outcome is done in Python code.

Data source

The data source for which the pitch tracking algorithms will be applied is the Pitch Tracking Database from Graz University of Technology (**PTDB-TUG**). It contains 2342 recorded phonetically rich sentences each spoken by 10 female and 10 male English native speakers. For each sentence and speaker, there is a recording by a normal microphone, the recording by a laryngograph and a reference recording taken as the ground truth for pitch estimation. The ground truth data were extracted from the laryngograph data after pre-processing by a specific high pass filter. The pitch extraction was performed by the RAPT algorithm which possibly has to be considered in the later discussion of the results.

In PTDB-TUG three types of sentences are found, as illustrated in the table below.

Sentence Type	Sentences	Speakers	Total	Sentences/ Speaker
Dialect sentences	2	20	40	2
Phonetically-compact s.	450	2	900	45
Phonetically-diverse s.	1890	2	3780	189
Total	**2342**		**4720**	**236**

Algorithms

ESPS The Entropics Signal Processing System is a comprehensive set of speech analysis/processing tool, which pitch traces can be displayed. It does not worry about the noises or complicated performances, it is just a standard pitch tracking software provided by Wavesurfer.

AMDF Average Magnitude Difference Function is another pitch detection method based on autocorrelation function, which can give quite accurate results for highly periodic signals, but suffers from incorrect pitch detection in noisy conditions.

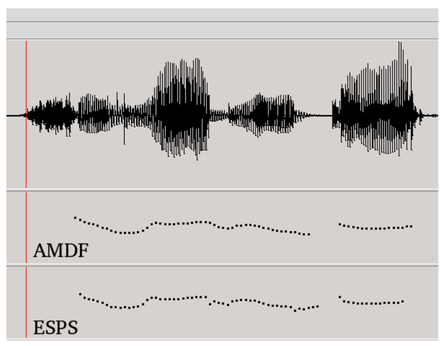

Figure 1: *View from the difference recorded female voice through a microphone by ESPS and AMDF in Wavesurfer.*

Praat is a tool to analyze and generate acoustic files. It is developed by the phonetic science department of the University of Amsterdam. Even if the Praat tool allows more than a pitch tracking algorithm, for our experiment we will mainly use this part and the visualization techniques.

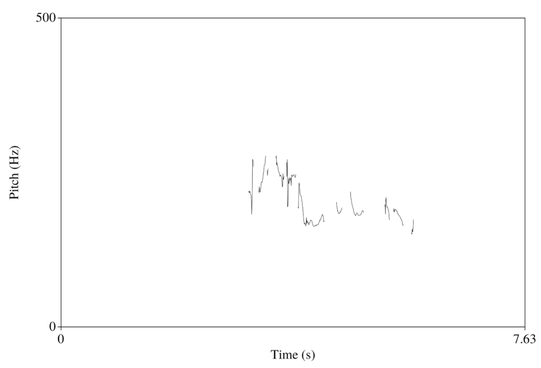

Figure 2: *View from the difference recorded female voice through a microphone by ESPS and AMDF in Wavesurfer.*

Comparison

For the comparison of the different algorithms, a quantitative measurement of the quality of each algorithm is performed. The pitch estimated by an algorithm is compared to the ground truth in three different ways: Based on the f-measure method **Accuracy** and **Recall** of the algorithms are measured as an indicator of the classification performance. Accuracy shows the number of correct predictions (pitch or no pitch) from all predictions made. Additionally, Recall gives a measure of the completeness of the algorithm. Each frame with rightly identified pitch is taken into account for the error measurement. The Mean Squared Error (**MSE**) method is chosen as the measure of the algorithms quality in terms of erroneousness. Since the squared error does not consider octave errors and the fact that higher frequencies (of female speakers) lead to bigger errors all the values are converted from Hertz to the logarithmic unit Cent and constrained to a single octave.

Results

In the first experiment the performance of the sentences recorded by microphone and laryngograph are analyzed. For microphone recordings the ESPS algorithm shows the best results in Accuracy and Recall. AMDF and PRAAT show slightly to considerable lower results in both catgeries. The differences in the results of laryngograph recordings are larger. ESPS achieves very good results in Recall and Accuracy, PRAAT and AMDF, in particular, show poor results for Recall and reasonable lower results for Accuracy.

	MIC (Prec./Acc.)	LAR (Prec./Acc.)
AMDF	0.87417 / 0.92144	0.47990 / 0.87497
ESPS	0.91209 / 0.94765	0.92449 / 0.97250
PRAAT	0.80851 / 0.93377	0.75180 / 0.93727

Table 1: *Recall and Accuracy results for recordings by microphone (MIC) and laryngograph (LAR) for each of the algorithms.*

The results for Mean Squared Error are very balanced between the three algorithms and the two types of recordings. For microphone recordings AMDF has a slightly lower error, whereas for laryngograph recordings ESPS shows a slightly better performance.

	MIC (MSE)	LAR (MSE)
AMDF	598.489	599.830
ESPS	600.140	597.489
PRAAT	598.789	599.0200

Table 2: *Mean Squared Error results for recordings by microphone (MIC) and laryngograph (LAR) for each of the algorithms.*

The second experiment investigates on differences for female and male speaker's recordings. For female speakers the ESPS achieves the best results for Recall and PRAAT for Accuracy. For male speakers, again, ESPS performs best in Recall and in Accuracy, as well. PRAAT provides very poor results in terms of Recall, whereas AMDF gives slightly lower results than ESPS.

	FEMALE (Prec./Acc.)	MALE (Prec./Acc.)
AMDF	0.89268 / 0.91731	0.85554 / 0.92568
ESPS	0.95109 / 0.94167	0.87284 / 0.95378
PRAAT	0.92084 / 0.95030	0.69617 / 0.91724

Table 3: *Recall and Accuracy results for recordings by microphone of female and male speakers for each of the algorithms.*

Mean Squared Error for female and male speakers shows only a slight differences between algorithms. PRAAT has the lowest results for both female and male speakers. The error for male speakers tend to be higher than for female speakers for all of the three algorithms

	FEMALE (MSE)	MALE (MSE)
AMDF	596.573	600.500
ESPS	599.205	601.166
PRAAT	597.967	599.610

Table 4: *Mean Squared Error results for recordings by microphone of female and male speakers for each of the algorithms.*

Conclusion

In this lab experiment, we have evaluated three different algorithms. Each method can be used to analyze speech data regarding the pitch contour. The results show that there are only slight differences in the measures taken into account between the algorithms. In total, the ESPS algorithm provided the best results. Especially for laryngograph recordings AMDF and PRAAT show poor results, as well as PRAAT for microphone recordings of male speakers. Concluding this lab experiment we can say that ESPS is a pitch tracking algorithm with a high quality in several scenarios, whereas AMDF and PRAAT have only some drawbacks in particular configurations.