Prosodic Correlates of Rhetorical Appeal

Voice Wave Analysis of Ethos, Pathos and Logos

Michel Nienhuis
m.d.nienhuis@student.utwente.nl

ABSTRACT
In this paper a study is described on the prosodic voice wave characteristics of different rhetorical styles. Analysed were pitch, intensity and duration features of segments annotated ethos, pathos and/or logos in the 2004 Democratic National Convention Keynote Address of Barack Obama (USA). It was discovered that ethos shows a lower mean pitch and a slower speech rate, possibly to confer more credibility, and that logos, representing rational arguments, is spoken in a faster manner than pathos and ethos. However, these results are not statistically significant.

Keywords
Rhetoric, persuasion, prosody, voice wave, praat.

1. INTRODUCTION AND RELEVANCE
Rhetoric and its purpose, persuasion, have been studied for more than 2500 years. Aristotle (384-322 BC) famously wrote “Rhetoric may be defined as the faculty of observing in any given case the available means of persuasion.” in his work “Rhetoric”, written in 350 BC [1]. Aristotle introduced three modes of persuasion to classify the speaker’s appeal to the audience: ethos, pathos and logos. Ethos is an appeal to credibility: it is meant to prove the speaker’s qualifications on the subject and thus establish plausibility. Pathos is an appeal to the audience’s emotions. Logos is a logical appeal to rationality. Certain parts of a speech might be formal parts that do not fall in any of the above modes.

As Braga and Marques put it, “Assertiveness and determination in speech are fundamental qualities required to argumentation and persuasion in the political debate”, and at the same time, “the speaker also uses prosodic strategies in natural speech by manipulating tone, rhythm, duration, accent and energy in a way that can be decisive in conveying an opinion in a political debate. These prosodic strategies are intentionally chosen by the speaker in order to reinforce their discourse construction.” [2]. Based on this notion that one can or should vary prosody using a strategy, and also based on common sense, it can be assumed that there are prosodic differences between these three Aristotelian modes of persuasion. For instance, when trying to appeal to the audience’s emotions, one might try to speak louder, in a higher pitch or in a faster manner, for that is what many people do in discussions when they get emotional themselves. Showing emotions with your voice in that manner might trigger emotions in the audience. In the same manner, one who tries to appeal to credibility might speak with a lower pitch and in a slower manner to confer authority.

Also, as we will see in the literature research, both pitch and duration features are found to correlate with perceived speaker skill and charisma, and one might hypothesize that an appeal to emotion would need those perceptions more than an appeal to rationality.

No studies have been found that tried to find prosodic differences between these Aristotelian modes of persuasion, or any mode of persuasion for that matter. Linking the field of rhetoric and persuasion to the field of prosody does not seem to have triggered much research.

In this paper, a study into the prosodic differences between the three modes of persuasion is described. This was done by analysing the acoustic sound wave, annotated for ethos, pathos and logos or any combination thereof, of one of the most famous speeches of Barack Obama for differences in pitch, intensity and duration features. Barack Obama is generally considered one of the most eloquent speakers of recent history.

On a scientific level, my main motivation is that it will be interesting to learn if there are shared prosodic characteristics in the rhetoric modes of persuasion. The results from my research can be used in four ways. Firstly, computer-synthesized speech can be made more persuasive, for instance in an advertising context. Secondly, they might help for political analyses. Thirdly, they might be used to develop applications that teach the user how to use his voice depending on the rhetorical modes of persuasion used. Finally, in earlier research, it was shown that computer-synthesized voice is just as persuasive and conveys the message just as well as human voice, although speaker and voice are perceived less favourable [3, 4]. It might be possible to make the computer-synthesized voice more natural or favourable with the results of my research.

2. LITERATURE RESEARCH
Rhetoric, as defined by Aristotle, is part of persuasion. Persuasion can consist of many things both verbal and non-verbal, but what we are interested in is persuasion by means of the spoken word.

Sherblom and Reinsch stated in 1981 that much of the research on persuasive speech until that date “had been focussed on internalized phonetic shape at the linguistic level rather than properties of the acoustic sound wave” [5]. Since then, the prosody, which is part of the properties of the acoustic sound wave, has received somewhat more attention. But no studies were found that studied such prosodic features within any rhetorical framework. This means that the literature research for this paper has to build more on prosodic research in neighbouring fields, like advertising and speaker eloquence.
Within the advertising context, Chebat, Gelinas-Chebat et al. have carried out research of prosodic elements influential on the persuasiveness and credibility of speakers. They discovered that voice intonation and voice intensity shifted the attitudes of students positively when receiving an advertising message, but only under low-involvement conditions. The most efficient voice for advertisement combined low-intonation and low-intensity features. This combination bypasses the respondent’s defense mechanisms [6]. Whether this also applies to politics has to be seen, for political speech might not be considered equal to advertisement by the brains. In another study, voice intonation affects source credibility significantly more under high involvement as opposed to low involvement, while voice intensity affects source credibility significantly more under low involvement instead of high. [7]. It was also found that a moderate intensity, an unmarked intonation and a fast speech rate amounted to the highest credibility of the source [8]. In both [7, 8] it was found that voice characteristics significantly affected attitudes towards the advertised product and the intent to buy.

Within the field of prosody, research carried out by Rosenberg and Hirschberg on acoustic and prosodic correlates of charismatic speech, conducted with American political speech, showed that there is a significant agreement across subjects on which speech is charismatic and which also showed that enthusiastic, charming, persuasive, passionate and convincing speech is positively correlated with charisma. The same study showed that higher values of mean F0, standard deviation of F0 and the maximum F0 correlated with higher charisma ratings, and that higher intensity and higher speaking rate also correlated with higher charisma ratings [9].

There has also been related research on the topic of speaker skill in correlation with prosodic features. Strangert [10] found F0 and duration features to be significantly different with speakers rated high or low on speaker skill using samples of speech produced in Swedish parliament debates. The highest rated speaker had a higher mean and a more varying F0 compared to the lowest rated speaker, as well as a faster speech rate. Strangert also states that rhetoric aspects might be interesting to research, “as the choice of linguistic form no doubt can influence speaker ratings”. Strangert and Gustafson used these findings to test synthesized speech, which showed increased F0 dynamics to have the greatest effect on speaker skill, but elimination of disfluencies and hesitation pauses, plus increased speech rate also played a role [11].

3. PROBLEM STATEMENT AND RESEARCH QUESTION

No studies have been found on the prosodic correlates of pitch, intensity and duration within the Aristotelian framework of ethos, pathos and logos. In this study, I tried to discover whether there are any correlates using a case study.

The research question derived from the problem statement is:

are there correlates between the Aristotelian modes of persuasion ethos, pathos and logos on the one hand and the prosodic features of pitch, intensity and duration on the other?

In other words: are there prosodic features or combinations of features that indicate (define) the three modes of persuasion?

4. RESEARCH METHOD

4.1 Speech selection

To analyse the prosodic differences in rhetorical modes of persuasion, the 2004 Democratic National Convention Keynote Address of then Junior Illinois State Senator Barack Obama of the United States of America, delivered 27 July 2004 at the Fleet Center, Boston, MA, USA, was selected. This selection was done on the basis of the general opinion of the public and the media that Barack Obama is a very eloquent speaker, possibly one of the best speakers of recent history. It was assumed that differences in prosody may be bigger with eloquent speakers, as they may use the rhetorical modes of persuasion in a more efficient manner. Furthermore, pundits argue that this particular speech is among his best speeches, or perhaps even his best, which might also have worked towards a higher chance of finding correlations. The MP3 version of the speech, provided by AmericanRhetoric.com, was used [12].

The speech has a duration of about 15 minutes.

4.2 Annotation

For the written transcription of the speech the PDF version from AmericanRhetoric.com was used [12].

The method of annotation was marking every sentence with either ethos, pathos or logos, or any combination of them. Research on paragraph level prosody was not performed. The annotation was done by both myself and a graduate student Master of Business Information Technology, who was instructed about the three rhetorical modes of persuasion. After both annotations, consensus on the markings we did not agree on was reached through discussion. Consensus was found on all sentences.

For example, the two sentences “It is that fundamental belief: I am my brother’s keeper, I am my sister’s keeper that makes this country work. It’s what allows us to pursue our individual dreams and yet still come together as one American family.” were marked as pathos, because we think they are meant to stir the audience’s emotions. In the same manner, the sentences “In this election, we offer that choice. Our Party has chosen a man to lead us who embodies the best this country has to offer. And that man is John Kerry.” were marked as ethos, for we believe that he is trying to, in this case, make John Kerry more credible. After all, this speech is designed to rally the people for the election of John Kerry as President of the United States.

But, for instance, the sentence “From his heroic service to Vietnam, to his years as a prosecutor and lieutenant governor, through two decades in the United States Senate, he’s devoted himself to this country.” was marked as both ethos and logos, because Obama tries to make John Kerry more credible, but at the same time gives a rational argument as to why. As there are a lot of such sentences in the text, a lot of mixed labels were used. This made for more types of labels, with fewer instances of each label.

Praat [13] was used for analysing the voice wave as well as for annotating the speech. Silences between sentences, as well as applause from the audience, was not included when possible (marked “G” within Praat, which were then not included for analysis).
Intensity features

The loudness (energy) of the voice, measured in decibels. Measures are mean intensity, intensity standard deviation, intensity maximum, intensity minimum and intensity range (maximum – minimum). These intensity measures were performed using the same time steps of 1 millisecond and the same pitch floor of 75 Hertz.

4.3.3 Duration features

Duration features measured were the amount of speech time versus silence within each sentence (as a percentage of the total time) and the amount of words per minute within each sentence (speech rate).

The amount of speech time versus silence was calculated by dividing the amount of time with pitch (eg. where voice is present) by the total time that it took to utter the sentence.

The speech rate (amount of words per minute) was calculated manually, by counting the number of words and dividing that amount by the amount of seconds it took to utter the sentence, multiplied by 60 to get the amount of words per minute.

4.4 Hypothesis

As explained in the Introduction, it can be assumed that there are prosodic differences between the three Aristotelian modes of persuasion.

From the literature research, it shows that intonation, intensity and duration features significantly differ with different speaker skill, charisma rating, and perceived persuasiveness and credibility. Because these features show to have such a big influence on the perception of voice, it is hypothesized that, since prosodic differences are expected, these features will vary between ethos, pathos and logos.

Because of time constraints disallowing statistical tests, this hypothesis cannot statistically be accepted or rejected, but we will only elaborate on the data found and comment on the probability of acceptance of the hypothesis.

5. RESULTS

The results are also separated into the same three categories: pitch features, intensity features and duration features, after which aggregated results will be presented.

5.1 Pitch features

Table 1 shows the results of the pitch features, measured in Hertz. Results are averaged over all instances of each label (avg), and the corresponding standard deviation is shown (sd).

These results show that the average mean pitch varies by some 20 Hz across the labels. L seems to be associated with higher mean pitch values than E, and P is right up there with L, although LP combined shows a mean pitch more towards E. EL seems consistent with this: the L seems to elevate the E, as does P in EP, but to a lesser extent.

Standard pitch deviations do not seem to differ much within the sentences, they are all around 20 Hz. Maximum pitch and pitch range show the same as mean pitch: L and P have higher values, as does LP combined, whereas sentences with E in them show to have lower maximum values, although they deviate less. Minimum pitch shows the same as maximum pitch.

The pitch slope also seems to be higher with L, P and LP, but also with EL. E, ELP and EP seem to have less slope. Here it also shows that higher averages are accompanied by higher standard deviations. Figure 2 shows the mean and maximum pitch results in a bar chart.
Table 1. Pitch features

<table>
<thead>
<tr>
<th>Pitch features</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
<th>Slope</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (avg)</td>
<td>201.67</td>
<td>19.16</td>
<td>226.17</td>
<td>157.45</td>
<td>413.41</td>
<td>68.72</td>
</tr>
<tr>
<td>E (sd)</td>
<td>4.39</td>
<td>2.60</td>
<td>8.20</td>
<td>10.20</td>
<td>88.61</td>
<td>5.43</td>
</tr>
<tr>
<td>EL (avg)</td>
<td>211.51</td>
<td>19.53</td>
<td>237.97</td>
<td>161.62</td>
<td>450.64</td>
<td>76.36</td>
</tr>
<tr>
<td>EL (sd)</td>
<td>9.58</td>
<td>5.68</td>
<td>9.56</td>
<td>20.01</td>
<td>83.83</td>
<td>19.56</td>
</tr>
<tr>
<td>ELP (avg)</td>
<td>205.91</td>
<td>19.87</td>
<td>231.81</td>
<td>155.99</td>
<td>416.25</td>
<td>75.82</td>
</tr>
<tr>
<td>ELP (sd)</td>
<td>23.58</td>
<td>5.54</td>
<td>27.50</td>
<td>28.91</td>
<td>105.18</td>
<td>22.43</td>
</tr>
<tr>
<td>EP (avg)</td>
<td>204.32</td>
<td>19.56</td>
<td>231.18</td>
<td>156.21</td>
<td>386.31</td>
<td>74.97</td>
</tr>
<tr>
<td>EP (sd)</td>
<td>10.64</td>
<td>4.69</td>
<td>10.84</td>
<td>15.36</td>
<td>92.56</td>
<td>15.36</td>
</tr>
<tr>
<td>L (avg)</td>
<td>221.86</td>
<td>23.21</td>
<td>255.06</td>
<td>167.96</td>
<td>508.28</td>
<td>87.10</td>
</tr>
<tr>
<td>L (sd)</td>
<td>16.24</td>
<td>6.27</td>
<td>22.36</td>
<td>16.25</td>
<td>144.91</td>
<td>28.45</td>
</tr>
<tr>
<td>LP (avg)</td>
<td>209.09</td>
<td>23.10</td>
<td>242.19</td>
<td>159.16</td>
<td>477.95</td>
<td>83.02</td>
</tr>
<tr>
<td>LP (sd)</td>
<td>16.56</td>
<td>7.68</td>
<td>22.99</td>
<td>20.45</td>
<td>178.34</td>
<td>23.94</td>
</tr>
<tr>
<td>P (avg)</td>
<td>222.40</td>
<td>23.93</td>
<td>254.62</td>
<td>169.27</td>
<td>459.43</td>
<td>85.35</td>
</tr>
<tr>
<td>P (sd)</td>
<td>19.33</td>
<td>10.27</td>
<td>21.64</td>
<td>32.24</td>
<td>177.56</td>
<td>31.19</td>
</tr>
</tbody>
</table>

5.2 Intensity features

Table 2 shows the results of the intensity features in dB.

These results show L to have the highest mean intensity, followed by P and EL. EP and ELP have the lowest. The standard deviation within the sentences seems to be fairly constant overall. Maximum intensity seems to be lower in EP, but not in E and P separately. EP also seems to have the lowest minimum intensity, whereas L and P have the highest minimum, yet the lowest ranges. No clear pattern seems to emerge.

5.3 Duration features

Table 3 shows the results of the duration analyses. Figures 3 and 4 show the amount of speech and the speech rate in bar charts.

The amount of speech (where 0.70 equals 70%) is the highest with E and EL, whereas it is lowest with ELP. The difference is 16%. It seems that when P is introduced, words are pronounced in a shorter manner, hence a lesser amount of speech. The speech rate, measured in words per minute, tops with L and LP, while being lowest in P, ELP and EP.
5.4 Aggregating ethos, pathos and logos

It might be interesting to aggregate the results, so that only ethos, pathos and logos remain. This means that, for ethos, E, EL, ELP and EP are aggregated and averaged, and so forth. It also means that sentences with more than one label will be used in multiple figures. These aggregate results might shed a different light on the prosodic features of ethos, pathos and logos.

5.5 Aggregated pitch features

Table 4 shows the aggregated pitch features. Figure 5 shows the aggregated pitch mean and pitch maximum in a bar chart.

Table 4. Aggregated pitch features

<table>
<thead>
<tr>
<th>Pitch features</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
<th>Slope</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (avg)</td>
<td>205.63</td>
<td>19.61</td>
<td>231.99</td>
<td>157.04</td>
<td>406.16</td>
<td>74.95</td>
</tr>
<tr>
<td>E (sd)</td>
<td>14.93</td>
<td>4.85</td>
<td>16.78</td>
<td>19.98</td>
<td>95.03</td>
<td>17.44</td>
</tr>
<tr>
<td>L (avg)</td>
<td>210.44</td>
<td>21.91</td>
<td>241.04</td>
<td>159.96</td>
<td>464.34</td>
<td>81.08</td>
</tr>
<tr>
<td>L (sd)</td>
<td>18.08</td>
<td>6.86</td>
<td>23.49</td>
<td>21.98</td>
<td>150.00</td>
<td>23.57</td>
</tr>
<tr>
<td>P (avg)</td>
<td>210.67</td>
<td>21.86</td>
<td>240.75</td>
<td>160.43</td>
<td>438.93</td>
<td>80.32</td>
</tr>
<tr>
<td>P (sd)</td>
<td>18.39</td>
<td>7.64</td>
<td>22.54</td>
<td>24.44</td>
<td>151.93</td>
<td>23.96</td>
</tr>
</tbody>
</table>

These aggregates show values that are very close to each other, with no clear pattern emerging. These figures are too close to call.

5.6 Aggregated intensity features

Table 5 shows the aggregated intensity features.

Table 5. Aggregated intensity features

<table>
<thead>
<tr>
<th>Intensity features</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (avg)</td>
<td>71.64</td>
<td>5.56</td>
<td>76.19</td>
<td>56.77</td>
<td>19.42</td>
</tr>
<tr>
<td>E (sd)</td>
<td>2.15</td>
<td>0.99</td>
<td>2.19</td>
<td>4.07</td>
<td>3.00</td>
</tr>
<tr>
<td>L (avg)</td>
<td>72.40</td>
<td>5.42</td>
<td>76.90</td>
<td>58.06</td>
<td>18.85</td>
</tr>
<tr>
<td>L (sd)</td>
<td>2.57</td>
<td>1.13</td>
<td>2.19</td>
<td>4.53</td>
<td>3.46</td>
</tr>
<tr>
<td>P (avg)</td>
<td>72.13</td>
<td>5.33</td>
<td>76.55</td>
<td>57.94</td>
<td>18.61</td>
</tr>
<tr>
<td>P (sd)</td>
<td>2.60</td>
<td>1.27</td>
<td>2.12</td>
<td>5.32</td>
<td>4.00</td>
</tr>
</tbody>
</table>

The amounts of speech seem to be very close to each other. The speech rate however, shows the highest rate for the logos, followed by pathos, and ethos has the lowest rate, but also with the lowest standard deviation.
It can be concluded that the hypothesis can be accepted, but not confers greater credibility.

The slowest speech rate, possibly indicating that slower speech is spoken in a faster manner than, representing rational arguments, can be concluded that although not significantly, and also has a faster speech rate. It emerge from the intensity figures.

Ethos also has a high intensity, but no clear pattern seems to emerge from the intensity figures.

Logos also has a high intensity, but no clear pattern seems to emerge from the intensity figures.

Duration features hold more promise in that regard. It seems that logos has a longer pronunciation than ethos and pathos, although not significantly, and also has a faster speech rate. It can be concluded that logos, representing rational arguments, is spoken in a faster manner than pathos and ethos. Ethos has the slowest speech rate, possibly indicating that slower speech confers greater credibility.

It can be concluded that the hypothesis can be accepted, but not in any statistically significant manner.

6. CONCLUSIONS

On the pitch side, logos and pathos seem to be associated with higher pitch values than ethos. It could be that ethos needs a lower mean pitch to confer more credibility. At the same time, logos has the highest pitch slope and the highest maximum pitch.

Logos also has a high intensity, but no clear pattern seems to emerge from the intensity figures.

Duration features hold more promise in that regard. It seems that logos has a longer pronunciation than ethos and pathos, although not significantly, and also has a faster speech rate. It can be concluded that logos, representing rational arguments, is spoken in a faster manner than pathos and ethos. Ethos has the slowest speech rate, possibly indicating that slower speech confers greater credibility.

It can be concluded that the hypothesis can be accepted, but not in any statistically significant manner.

7. DISCUSSION & FUTURE RESEARCH

First off, the conclusion that ethos shows a lower pitch and a slower speech rate is in disagreement with [8], where moderate intonation and a faster speech rate are found to be the most credible, at least in an advertising context. My personal expectation that the use of pathos results in higher pitch or intensity was not shown.

It is hard to draw conclusions from this dataset, as it is small. Furthermore, since it was not possible to apply statistical tools, no significant relations were found. The conclusions drawn do have some value, however.

The research suffered from time constraints, for originally it was planned to also study the Acceptance Speech of Barack Obama on the 2008 Democratic National Convention, accepting his presidential candidacy in a 45-minute speech, as well as compare the data found to an interview with Barack Obama to see if there are any differences. If significant correlations were found, it was also planned to field test these results in a small-scale computer-synthesized voice test. But it soon became clear that all those things could not be realised within the time span of this project.

The result is a case study of one speech, with several notes to place. First off, the annotation was based on consensus between a graduate student and myself, both of whom are not experts on rhetoric. It was also clear that a lot of sentences cannot be classified as only ethos, pathos or logos, but are a combination of them. The result was more labels with fewer instances of each label, undermining statistical analysis.

When we look at the actual analysis, we also see some problems. The speech was delivered to a large crowd, so that the voice of Obama is not the only one being heard. Praat did manage to filter out most of the audience noise, but we cannot be sure that it is only Obama that generated pitch. Furthermore, the audience applauds a lot, which makes the abovementioned problem bigger. Apart from the audience, an echo of Obama’s speech can be heard, further polluting the pitch figures.

Since this study covered only one speech of 15 minutes, results found can hardly be generalized. The conclusions drawn are therefore only valid for this particular speech, but possibly give indications for future research.

A lot of future research in this field is possible. As said before, the link between rhetoric and prosody does not seem to have gotten much, if any, attention in the past. It will be interesting to build a large corpus of non-polluted speeches and investigate whether there are significant correlations between rhetorical modes and pitch, intensity or duration features. The same goes for other rhetorical devices and styles and their relation to prosody.

It might also be interesting to see whether people rated high on speaker skill use different prosody within their rhetorical modes than speakers rated low on speaker skill.

Furthermore, on the rhetorical side, it can be interesting to research the perception of speech in relation to the amounts of Aristotelian modes of persuasion used. Does more pathos make for a better speech?

ACKNOWLEDGMENTS

I would like to thank my guiding teacher, Dirk Heylen, for his knowledge, help and support from start to finish. I would also like to thank Ivo Nouwens, a graduate student of the Master Computer Science (track Human Media Interaction) for his help with the Praat programme and the analysis and Sybren Wijnia, the graduate student who helped with the annotation.

REFERENCES


