

Prosodic Elements of a Political Speech and its Effects on Listeners

Benjamin Weiss

Department of German Language and Linguistics
Humboldt University Berlin, Germany

benjamin.weiss@rz.hu-berlin.de

Abstract

Five longer utterances were analyzed. They were taken from a speech of Angela Merkel held on a party congress. Their impact on subjects was evaluated by means of a listening experiment. It showed, that the stimuli belong to two different groups. One of the groups was judged significantly better than the other independent of political bias and pre-knowledge of listeners.

An acoustical analysis of prosodic features (speaking rate, pausing, average F_0 and micro-prosodic contour of F_0 in stressed syllables) supported the differences in judgment between the two groups of stimuli.

A re-synthesis experiment verified the influence of the measured parameters on the judgments.

The two groups of stimuli are interpreted as different modes of speaking. Those modes correspond to content and function of an statement: The speaker used a calm variant to “inform” and “argue” and an excited variant to “appeal” or “be indignant over a topic”. The calm one was judged more likable. Reasons for the differences in effect are provided in the discussion.

1. Introduction

Prosodic features are known to carry linguistic information like stress as well as para-linguistic, such as age, gender and emotional state [1]. For a speaker like a politician, who wants to affect listener audience, the sound of the voice and way of speaking need not only to be likable, but also plausible. Their articulation needs to fits to the content. For example, a lower pitch in general comes across as more confident and pleasing [2], but it is judged not appropriate to a part of a speech that intends to carry away listener [3].

The specific style of speaking affects the listener’s impression on the speaker. So, a good speech is not only dependent on the content, structure or even diction, but strongly on the presentation [4].

There are few studies of political speeches regarding prosodic elements, as mostly content and rhetorical strategies have been analyzed. Even then the main focus is on structuring due to pauses (i.e. [7]).

The main aim in this paper is to present identified acoustical parameters of one speaker that invoke an emotional effect on listeners.

2. Material

The analyzed speech was held at June 7th 2001 at the party congress of the German conservative party CDU. The politician, Angela Merkel, had already been chairwoman of this party in opposition at this point. The speech is remarkable, as it marks an important moment in the beginning of the campaign for the parliamentary election and deals with the main topics for this

coming year of campaign. At that time, Merkel was in discussion as the conservative candidate for the chancellor, the first time a female in the history of Federal Germany.

The speech itself lasted about 55 minutes. In a preprocessing step, the structure of the whole speech was noted. It contains textual parts of about 10 minutes each with utterances from mostly 20 to 40 seconds. These utterances form the basic parts of the different propositions. Five of them were chosen as typical statements of her. They included no slips, which occurred very frequently.

These five chosen utterances build up one concluded statement each (exact wording, see Appendix). They are long enough to invoke an emotional effect by listeners and encompass some intonational phrases. Therefore they provide enough material to analyze the manner of pausing for each of them.

3. Experiment one

In the first experiment the effect of the selected utterances on listeners was quantified. 25 subjects participated (22–36 years of age), 11 male, 14 female. 19 of them were students, the other 6 were employees in non-university fields. Nobody was paid for the participation.

The five stimuli were presented randomized. If required, each stimulus would be repeated. Mostly, the repetition was done 2–3 times.

The questionnaire provided a semantic differential [5]. The subjects were asked to rate their associations with each stimulus along bipolar scales of a series of contrasting pairs of adjectives. Each scale provided 7 grades (3 to 0 to 3). The specific (positive) poles are “likable”, “comprehensible”, “natural”, “confident”, “passionate”, “dedicated”, “insistent”, “active”, “sensitive”, “pleasing”, “diversified”, “expressive”, “honest”, “declaratory”, “reliable” and “factual”. The scale “pleasing – not pleasing” was set as the control pair to “likable”, while “declaratory” controls “factual” and “active” “diversified”. The alignment of a pole and its complement was varied to prevent all poles with positive association to appear on one side of the questionnaire.

Additionally, each subject was interviewed on the basis of 8 questions:

1. What in the different utterances attracted your attention?
2. Which parts were especially pleasing? Why?
3. Which parts were especially annoying? Why?
4. What, in the politician’s way of speaking, stands out?
5. Did you recognize the speaker?
6. What is your general attitude towards the speaker?
7. What is your general attitude towards the party?

8. Additional opinions, i.e regarding the questionnaire or stimuli?

The procedure including the interview lasted 20–30 minutes for each subject.

3.1. Results

Whereas 17 subjects claimed to have a neutral attitude towards the party, 5 told to have a positive, 3 a negative attitude. Everyone recognized the speaker, while 8 of the subjects specified to have a positive attitude towards her, 5 a negative. The others claimed to be neutral.

Friedman’s and Wilcoxon’s non-parametric tests were done on the data of the questionnaires. The means but not the ranks differed for the gender and the political alignment. The overall means were lower for males and the subjects that reported to be not-aligned to the party or person.

10 of 16 polys showed significant differences ($p < 0.05$) in the answers (“likable”, “comprehensible”, “confident”, “insistent”, “active”, “pleasing”, “diversified”, “honest”, “reliable” and “factual”). Stimulus A and B differed not, while C was more “diversified” than A and more “diversified” and less “factual” than B.

Stimulus D and E were judged significantly worse and more active (more “insistent” and “active”, less “factual”) than the others. While E differed in 8–9 pairs to A, B, C respectively, D differed in 6 to A, in 4 to B and 1 to C. Stimulus E was also less “likable” and “insistent” than D.¹

3.2. Discussion

Altogether, the stimuli can be ordered in the level of activity, from A being judged “calmest” to E that was ranked most “active”. The more “active” a stimulus was rated the less positive (i.e. “likable”, “natural”, “confident” but also “diversified”) it was judged.

In number and distances between significant differences there are two groups to be found, A, B and D, E. For stimulus C the results are not as clear, but it could be grouped together with A and B, as it differs in average in maximum 0.5 grades in the significant scales to A, B but 1 to 2 full grades to D, E.

These findings are supported by the qualitative part of the interviews. The subjects criticized D and E because of the high and uncomfortable pitch and “annoying” accentuation.

13 subjects mentioned their difficulty to judge C, because the last part seemed to be different from the main one. Therefore the outstanding intonational phrase of C was not included during the acoustic measures. This edited stimulus will be referred to as C1.

No subject could reproduce the contents of either statement, some could remember catchwords like *unemployment* and *Mal-lorca*.

4. Acoustical measurements

The speech was recorded on video from television broadcast. Because of the strong echo and the possible use of compression, no intensity or spectral analysis was carried out.

4.1. Global fundamental frequency

The F_0 was measured manually, excluding laryngalization, echo and the few parts with strongly interrupting noise or speech.

¹E was also more “active” but less “diversified”. As these scales resemble each other semantically, they were not interpreted.

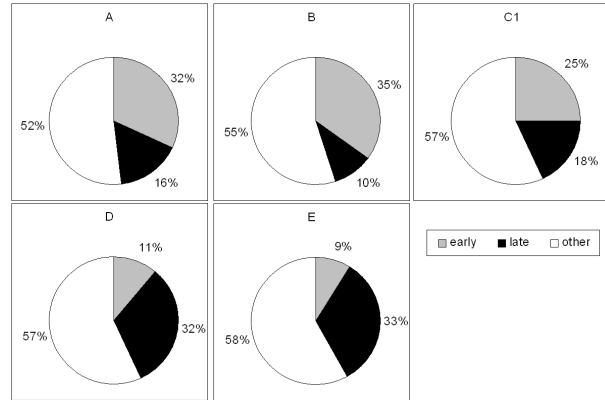


Figure 1: Time of change of pitch pattern in stressed syllables (in %)

The medians of the different stimuli are significantly different from each other. A, B and C1 differ from D and E. The first group has lower global F_0 and more variation (F_0 -range) than the other group (see Table 1).

Table 1: Global F_0 information (ERB-scale)

stimulus	F_0 median	F_0 -range
A	7.90	5.22
B	7.37	5.95
C1	7.53	5.41
D	8.08	4.84
E	8.03	4.70

4.2. F_0 -contour in stressed syllables

To analyze the strategy of accentuation, the F_0 -contours of stressed syllables (main word-stress) were examined. To classify the contours, the data was stylized by straight lines [6]. Two features have been observed, the manner of the contours (rising, falling, level, a combination of them) and the time of change (between the onset and offset of the vowel, before that and after). Early or late changes include only the extreme values of raising-falling and falling-raising contours and the moment of gradient change from constant to not constant lines (mostly constant-raising).

D and E have more raising contours (59%, 55%) than A, B, C1 (42%, 37%, 51%). Whereas medial moments of contour changes or constant gradients of the lines make up about 55% of each stimulus, D and E show a higher late-to-early ratio (2.9, 3.7) than A, B, C1 (0.5, 0.3, 0.7, see Figure 2).

4.3. Tempo

The utterances differed not in their global syllable rate. However, there are differences in ratio of the duration of stressed vs. unstressed syllables. They are 1.7 for A, B, C1 and 1.4 for D, E. Also, the deviation is higher for A (0.09 s), B (0.08 s) than the other (C1 0.07 s, D 0.06 s, E 0.06 s).

4.4. Pausing

The ratio of articulated time to pause duration and the ratio of articulation time by number of pauses is shown in Figure 2):

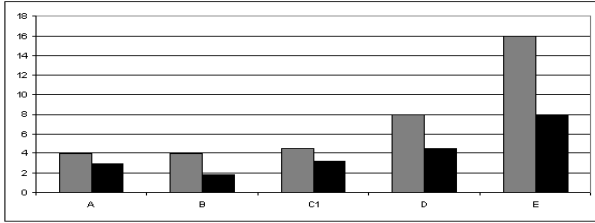


Figure 2: *Duration of articulation time by pausing time (grey) and articulation time by number of pauses (in s) (black)*

As we can see, the more “active” stimuli have a higher articulation duration compared to the duration of pauses and a higher number of pauses relative to the stimulus length.

Pausing is important to structure spoken speech for the listener. Not only their number and duration, but also their syntactical position is important to note. The number of phrase or clause boundaries marked by pauses might be a factor for a “comprehensible” judgment by the subjects. In stimulus A 80% of the clause boundaries are marked with a pause (B 75%, C1 67%, D 23%, E 45%).

4.5. Summary

The measurements of several prosodic features made it either possible to order the stimuli in a way similar to the results of the perception experiment I, or even group them: There is a difference between stimuli A, B, C on the one hand and D, E on the other. The first group was described as calm, the other as excited. The calm variants occur with declarative content the other with persuasive content.

The more active articulated utterances can be characterized by higher F_0 , less F_0 -variability and more and late rising F_0 -contour in stressed syllables. The stressed syllables are less longer than unstressed and less variable in duration. There are also fewer pauses.

These acoustic differences accompany a less good effect on listeners, independent of the context of the utterances. While more pauses might help listener in perception (thus stronger “comprehensible” judgments), one might ask if the pitch or the differences in stress affect the contrasts in positive ratings such as “natural” or “pleasing”. The interviews indicates such hypotheses.

To investigate, if these prosodic features are actual responsible for the effect on listeners, a second perception test was done with re-synthesized stimuli.

5. Experiment two

To verify the effect of the acoustic parameters on the listening experiment, an additional experiment was set up. Stimulus E was judged worst in experiment I and so modified with PSOLA. The hypothesis was that changing the described acoustical features to the values of A or B, subjects would describe their impression on the manipulated stimuli as more “likeable”, less “active” and so on.

The features were changed in a systematic way: 7 additional pauses were included at clause boundaries. F_0 was lowered. In stressed syllables, F_0 -contour was made with earlier slope to gain the same maximum.

While building up the stimuli it turned out, that including pauses combined with modified F_0 -contour resulted in exagger-

ated stimuli: Therefore two times the F_0 -contour was eased in favor of a neighbored one for the two stimuli with and without lowered F_0 .

The manipulation was done only slightly, as the echo produced artifacts in the stimuli. It was impossible to produce more than one value for each feature that would be acceptable and audibly different.

The proceeding was similar to experiment one. 20 subjects participated (age: 21–26, 12 females, 8 males).

5.1. Results

The manipulation showed several significant results. Best stimulus was the one with only lowered F_0 (more “likeable”, “comprehensible”, “natural”, “pleasing”, “reliable” than the original E). It was the only one, that was not immediately recognized as manipulated.

Adding only pauses leads to a more “comprehensible”, less “expressive”, more “declaratory” judgment.

The alternative stress contour resulted in a more “insistent” and “active” feedback.

The combination of adding pauses and lowering F_0 resulted in a less “likable”, less “active”, more “declaratory” and “factual” “judgment”. The stimulus with manipulated pauses and stress produced less “passionate”, less “active”, less “insistent” and more “factual” results.

The stimulus with manipulated F_0 global and local was judged less “passionate”, “expressive” and more “factual”.

The one with all three parameters changed was more “likable”, less “natural”, more “declaratory” and more “factual”.

5.2. Discussion

All manipulated parameters and their combination resulted in judgments similar to the target utterances A and B. All in all they were described as less active (less “active”, “insistent”, “passionate”, “expressive”, more “factual”, “declaratory”) with the only exception of the one with a different stress intonation. This one was more “insistent” and more “active”. This effect ceases in combination with other parameters.

The stimuli with lowered F_0 and added pauses was more “comprehensible”. More “reliable” was the one with lowered F_0 .

However, the combinations did not add up the judgments. Changing more than one parameter leads to less “likable” (pauses & F_0) and less “natural” utterances (all parameters). The interviews give a possible reason for this. The sound quality was stated as bad and some subjects told they used “natural” not for describing the speaking style, but the sound quality of the stimuli.

6. Conclusions

Two different speaking styles were found for the politician. One can be described as calm and factual, the other as active and expressive. The calm one was judged better. An acoustic analysis showed, that the calm variant is characterized by lower and more variable F_0 , more clause boundaries marked with pauses and a different stress pattern (more diversity in syllable duration, higher overall duration in comparison to unstressed syllables and less raising F_0 -contours and less late but more early activity of the contours).

The re-synthesis experiment showed, that the global F_0 and added pauses are responsible for the judgments concerning the

activity. A earlier and faster slope resulted in a more “insistent” and more “active” judgment.

There is a correlation between “activity” and “unpleasing” judgments, that were not solved by experiment two. It is plausible, to conclude from the results that the politician’s insistent way to speak is indeed less natural and pleasing than her calm parts in the speech. Compared to other acoustical measurements of politicians speech, she makes less pauses in the active stimuli [7], while other politicians increase the number of pauses, when excited [3]. This seem to cause the less “comprehensible” and less “confident” judgment. The different style of F_0 -contour in the more active stimuli is inappropriate, as this late raising was shown to act less “insistent”. At last, the speaker’s pitch might just be too high. While any emotional expression (anger or joy) generally leads to a higher and therefore fitting and natural global F_0 , it does not become less diversified [1] which is true in this utterances.

In summary the influence of content of the stimuli and political alignment of the subjects was controlled. The prosodic features analyzed have proven responsible for different impacts on listeners. Especially the different F_0 -pattern in stressed syllables are a promising parameter for further studies.

7. Acknowledgements

The author wants to thank W. Sendlmeier for his support and helpful suggestions.

8. References

- [1] Murray, I. R. and Arnott, J. L., “Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion”, *JASA*, Vol. 93(2), 1992, p 1097–1193.
- [2] Thiel, S., “Die mittlere Sprechstimmlage als Wirkungsfaktor”, in *Ergebnisse der Sprechwirkungsforschung*, E.-M. Krech, J. Suttner, E. Stock, Ed., Halle-Wittenberg: Abt. Wissenschaftspublizistik d. Univ., 1987, p 202–208.
- [3] Paeschke, A. and Sendlmeier, W., “Die Reden von Rudolf Scharping und Oskar Lafontaine auf dem Parteitag der SPD im November 1995 in Mannheim. Ein sprechwissenschaftlicher und phonetischer Vergleich von Vortragstilen”, *ZFAL*, Vol. 27, 1997, p 5–39.
- [4] Quintilianus, M. F., “Institutio oratoria”, Vol. 2, H. Rahn, Ed., Darmstadt: Wiss. Buchges., 1995.
- [5] Osgood, C. E., “The Nature and Measurement of Meaning”, *Psychological Bulletin*, Vol. 49, 1952, p 197–237.
- [6] d’Alessandro and C., Mertens, P., “Automatic pitch contour stylization using a model of tonal perception”, *Computer Speech and Language*, Vol. 9, 1995, p 257–288.
- [7] Kowal, S., “Über die zeitliche Organisation des Sprechens in der Öffentlichkeit”, Bern: Huber, 1991.

9. Appendix

Stimulus A: (38 s) Und damit kein Missverständnis entsteht: Es geht dabei wirklich nicht vordergründig um Disziplin, sondern, liebe Freunde, ich kann es gar nicht ernsthaft genug sagen, es geht um das Wohl, es geht um das Schicksal dieser Union . . . Es geht um das Schicksal dieser Union nach einer Wahlniederlage 1998 und vor der entscheidenden Aufgabe im Jahre 2002, nicht im Jahre 2006, nicht im Jahre 2010, sondern im Jahre 2002 wieder die Regierung der Bundesrepublik Deutschland zu stellen.

And to avoid any misunderstanding: In the first place I’m not concerned with discipline, dear friends, I can’t emphasize too much, I’m concerned with the well-being, the fate of this Union . . . the fate of this Union after defeat in the election of 1998 and facing the decisive task for the year 2002, not 2006 and not 2010, but in the year 2002, to take up government of the Federal Republik of Germany.

Stimulus B: (23 s) Wenn wir uns auf eine solche Diskussion einlassen, dann machten wir, davon bin ich zutiefst überzeugt, einen entscheidenden Fehler. Wir machten den zweiten Schritt vor dem ersten, und wir wüssten genau, dass niemand an diesem Fehler mehr Interesse hätte als die Sozialdemokraten und die rot-grüne Koalition, und deshalb werden wir diesen Fehler nicht machen, liebe Freunde.

If we allow for such a discussion, then we would make, I’m deeply convinced, a decisive mistake. We would do the second step before we had made the first, and we would know for sure that no one would have a deeper interest in us making this mistake than the Social Democrats and the red-green coalition, and that’s why we won’t make this mistake, dear friends.

Stimulus C: (20 s) Und genau deshalb glaube ich auch, dass der Weg, den wir gemeinsam beschritten haben . . . Ein Weg, der die Konsequenzen aus der Niederlage 1998 zieht und sagt: Jawohl, wir haben verstanden, wir haben nicht auf jede Frage die zufrieden stellende Antwort gehabt, aber wir werden sie suchen und zwar suchen unter dem Motto: Mitten im Leben bei den Menschen stehen zu wollen.

And exactly that’s why I believe that the path we’ve chosen in union . . . A path that draws conclusions from defeat in 1998 and says: Yes for sure, we understood, and we didn’t have a satisfying answer for every question, but we will look for it, meaning we will look for it according to our motto: In the middle of Life, standing close to people.

Stimulus D: (27 s) Aber meine Damen und Herren, nun ausgesprochen bei der Frage, wenn es um den Beginn des menschlichen Lebens geht; wenn es um die Frage geht, wie können wir die Gentechnologie auch einsetzen, um Krankheiten zu heilen, aber wie müssen wir gleichzeitig der Würde des Menschen gerecht werden. Bei dieser Frage lassen wir uns nicht vom Bundeskanzler sagen, legt eure ideologischen Scheuklappen ab. Das ist der denkbar falscheste Ort für eine solche Aussage.

But my dear Sirs and Misses, especially in the question of the starting point of human life; in the question of application of gene technology to heal illnesses, but how do we at the same time be appropriate to human decency. In this question we don’t admit the Chancellor to say, pass your ideological blinkers. That’s the most wrong context to make such a demand.

Stimulus E: (25 s) Warum müssen wir eine private Altersvorsorge organisieren, bei der durch jeden Paragraphen hindurch schimmert, dass wir immer der Meinung sind, alle wollen nur betrügen. Warum müssen wir die Menschen zwingen, sozusagen bei dem Anlegen eines privaten Wohneigentums erstmal einzusparen, dann dürfen sie was rausnehmen, zum Schluss müssen sie’s wieder zurückzahlen, und wenn sie im Alter auf Mallorca sind, müssen sie dem Staat wiedergeben: So überzeugt man doch Menschen nicht davon, für ihr eigenes Lebensrisiko eine Initiative zu ergreifen.

Why do we have to organize a private old age insurance, in which every paragraph seems to cry out that we feel everybody always only wants to cheat on it. Why do we have to force people, so to say to save for their fund for building private living, then they may take out something, and in the end the have to give it back again, and when they spent their old age on Mallorca, they have to give pay back to the state: With that method you don’t convince people to take initiative for their own life risks.