Prosody in public speech: analyses of a news announcement and a political interview

Eva Strangert

Department of Philosophy and Linguistics
Umeå University, Sweden
eva.strangert@ling.umu.se

Abstract

The study concerns informative and argumentative functions of prosody in the public domain. Analyses were based on speech samples from a professional news announcer and a well-known politician, each with a reputation as a highly skilled speaker. Results show both speakers to use prosody in a very efficient way. In the news reading, prosodic variation mainly serves informative demands, primarily the focusing of important information. The interview allows for a greater variety of expressions and the skill of the speaker manifests itself in her use of a wide repertoire of argumentative and emotionally colored expressive acts conveyed by prosody. The goal of this research is to shed light on what makes “a good speaker” and also how speech can be optimally adjusted to the demands of the situation.

1. Introduction

In recent years, research on speaking style variation has been intense. There has also been a growing interest in emotional and attitudinal aspects of speech. In the paper presented here, these two directions of speech research are combined.

The focus in the present study is prosody in speech produced by skilled speakers and directed to an audience via public media. This area has the potential for being both empirically and theoretically interesting, as skilled speakers have the capacity to strengthen human communicative resources to the extreme; these speakers may at the same time be informative, emotional, argumentative and elaborative in their speech.

We will give a short overview of previous research in this area as well as other studies of relevance for understanding what characterizes speech in public. We further report on analyses of two samples of public speech, a professional news reading and a political interview, some aspects of which are also compared with non-public, non-professional speech data.

The attention of the listener is crucial in the public speech domain, whether the purpose is to be mainly informative (news reading) or argumentative (political speech). To attract the listeners’ attention, the speaker, in addition to having an interesting piece of information to deliver, preferably should wrap the information in a form that makes the message go through. That is, in addition to having an interesting content, the speaker may profit from being “a good speaker”. It is assumed that this means an optimal use of, primarily, prosody. Thus, the purpose of the present work is to shed light on public speech behavior and in particular how prosody may contribute to successful interaction with the audience. That is, through analyses of successful public speech we may eventually come up with answers to what makes a good, or charismatic, speaker.

2. Background

O’Connell and Kowal [1] give a survey of studies of American and German political speech in public media. They compare read-aloud and spontaneous speech in this setting and find differences of the same kind as for non-public speech, but they do not enlarge upon how public speech might differ from non-public. This issue, however, is approached by Touati [2] and Deuz [3] and recently by Braga and Marques [4]. This research [4] has its focus on how prosodic features contribute to the listeners’ attention and interpretation of the message in political debate. The listeners conception of a speaker as “convincing”, “powerful”, “committed”, etc is related to different (combinations of) prosodic features, or maximes. The study builds on the idea put forward by Gussenhoven [5] and developed further by Hirschberg [6] of universal codes for how prosodic information is produced by the speaker and perceived and interpreted by the listener. Wichmann [7] and Scherer [8] deal in a similar way with the relation between prosody and what can be described as “affective meaning”. These studies in which prosody is seen as the main contributor to expressing “the meaning” of the message have great relevance for the present work. In addition, prosody in itself – speech rate, intensity, fluency etc – is important for characterizing a speaker, as well as expressions of emotions [7].

However, the matching of subjective ratings of speech and prosodic features is far from simple. For example, an utterance may convey several states, attitudes or emotions simultaneously. Promising work to shed light on the matching of complex subjective ratings of emotions and prosody is given in [9].

3. Read-aloud and interview data

In the following, two kinds of Swedish speech data will be analyzed. One is a professional (TV) news reading with a total of 59 words, a single news item dealing with an expected high increase of salaries in Sweden. The other is a radio interview with a well-known politician on world politics after the US Trade Center attack. The interview is about 25 minutes long and includes 4102 words. Both the news announcer and the politician are women and each has a reputation as a very good representative of their respective professions. The news-reading analysis elaborates on a material previously presented in [10] and the interview has been analyzed as regards phrasing behavior in [11, 12].

3.1. Characterization of the two speech styles

The news reading is characterized by a rather fast overall tempo. The pauses are few and short and the listener gets the
impression of a very efficient style of speaking. Careful
listening however reveals portions of relatively slow speech,
predominantly in conjunction with prominent words.
Prominent words, in addition, are spoken with extreme
emphasis and heard to be articulated very distinctly. The
speech style can be described as very informative,
authoritative, and trustworthy.

In the interview, the speaker sounds very fluent, although
she varies her speech tempo and inserts pauses very
frequently. In addition to tempo variations, the speech is
dynamic with regard to tonal and loudness features, which are
used to stress and emphasize portions of speech. The style is
lively and energetic and the listener gets the impression of a
very committed and convincing speaker having a rich rhetoric
repertoire at her disposal, including both phonetic/prosodic
and other expressions. One of these is repetitions in form,
where prosody appears to play an important role in achieving
the expected effect on the listener. The assertiveness, power
and skill in handling the situation also are reflected in her
ability to oppose interruptions by the interviewer.

Summarizing, while the style of both speakers appears to
be both informative (stressing important portions of speech)
and argumentative, the more neutral, informative elements
dominate the news reading, while the political speech is
predominantly argumentative.

4. Acoustic-prosodic analysis

A number of prosodic features were selected for analysis,
considering them to have impact on how informative and
argumentative impressions are formed by the listener. The
selected features are similar, but not identical, to the maxims
of Braga and Marques [4].

4.1. Speech and silence

4.1.1. News reading

The perception of an efficient style of speech is supported by
acoustic measurements. The proportion of silence to speech is
9%. In a non-professional reading of a similar text of the
same length, this proportion was 28 %. These figures could
also be compared to the 46% ratio for a sample of (non-
professional) spontaneous speech [10]. The articulation rate
(syllables/second, excluding false starts and repetitions) is
5.8 for the news reading and for the non-professional reading
as well. (The corresponding figure for the spontaneous speech
is 4.2.) The very different silence-to-speech ratios in the two
read-aloud samples then should be attributed primarily to
differences in pause length. This is also supported by data;
while there are pauses (silent intervals) ranging from 40 to
600 msec in the professional reading, the range is much wider
in the non-professional read speech (245-2040 msec). A
similar range of pauses between read sentences were found in
a previous study of ten non-professional speakers [13].

4.1.2. Interview

The silence-to-speech ratio is the same as for the professional
news reading, that is, 9%. Considering that the interview was
produced without a prepared manuscript, this is an extremely
low figure (cf. the ratio of 46% for the non-professional
spontaneous speech above). That the pauses (silent intervals)
are generally short with a mean duration of 230 msec

contributes to the low silence-to-speech ratio. Only four
pauses out of 522 exceed 1 second in duration.

4.2. Phrasing and fluency

4.2.1. News reading

The following example shows typical phrasing in the news
reading. Perceived breaks (L) and focused words (in bold) as
annotated by one experienced transcriber appear here com-
bined with pause durations (within parentheses):

(1) Men landets # <eh> kommuner o(ch) lan(d)sting råknar
me(d) # (40) upptill # (90) sju - # - procentiga
lö(n)kningar(n) nästa år. ‘But our country’s communes and
county councils expect # up to # seven # percent increases
of salaries next year.’

The phrasing here is very evident. There are two very
short pauses, but several perceived breaks in addition. The
breaks and pauses cut up the sentence into smaller chunks,
which reflect a semantic rather than a syntactic structuring.
However, though this example may be considered to some
extent extreme, as all breaks are in non-syntactic positions, it
demonstrates a kind of “semantic pausing” which occurs
frequently in news reading. Another example is:

(2) Det(t) visar en # kartläggning som # …. ‘This is shown by
a survey that…’

The pauses of longer duration occur exclusively between
sentences. This, combined with the very short sentence-
internal silent intervals and frequent filled pauses
(exemplified here by <eh>), most reasonably contributes to
the impression of the speaker as very fluent.

4.2.2. Interview

The chunks (portions of speech between perceived breaks) are
similarly short in the interview. Chunks with 2-4 words are
the most frequent, but single-word chunks also occur. While
most of the perceived breaks occur in syntactically motivated
position, nevertheless as much as 20% do not. (Perceived
breaks (and perceived focused words) were determined on the
basis of majority votes by three experienced transcribers.)

In terms of the commit-and-restore model developed by
Clark and Wasow [14], breaks occurring within a syntactic
constituent “violate continuity”, as the model initially predicts
that “speakers prefer to produce constituents with a
continuous delivery”. That is, speakers are assumed to aim at
producing entire constituents without interrupting themselves.
When they fail, they behave in a non-random way. According
to the model, speakers make an initial commitment as to what
will follow by initiating the constituent without having
decided on all of it. By such commitments, the speaker
signals to the listener that he/she is going to continue
speaking.

In many of the 20% cases of violations of continuity, the
perceived break occurred after initial function words, most
frequently after prepositions, conjunctions and subjunctions.
An example is:

(3) Jag tycker att Sverige ska kräva # att # bombningarna
upphö. ‘I take the view that Sweden should demand # that #
the bombings are stopped.’

That is, we find a kind of pausing similar to that in the
news reading. However, the mechanisms behind the
seemingly identical pause behavior are only partly the same
(see 4.3). The interruptions after function words and in other syntactically unmotivated positions in the spontaneously produced interview depend to a great extent on the on-line processing; the speaker has almost no time to plan what to say next. Signaling the intention to continue by coming to a stop after the initial portion of a constituent, gives the time to come up with what will follow and also has the advantage of stopping others from breaking in. For a politician trying to avoid being interrupted, pausing in such positions can be advantageous.

However, producing breaks in non-syntactic positions can also be seen as a means to emphasize important parts of what is being said. This is a strategy shared by the news announcer and the politician and will be covered later in 4.3.

Filled pauses and/or a slowing down frequently occur instead of pause insertion after initial commitments. This is the main reason why the silence-to-speech ratio is comparable to what we find for the news-reading text (see 4.1.1). But also the fact that the speaker uses the slowing-down strategy more generally when speaking – and conceivably as a means to plan ahead – contributes to fluency. Details on this slowing-down/lengthening phenomenon and how computations were made can be found in [11].

4.3. Focus and emphasis

4.3.1. News reading

There are several examples in the news reading of non-syntactic phrasing (4.2.1). The speaker most reasonably makes such adjustments in order to stress and emphasize important speech portions. The examples (1) and (2) are just two instances. A pause or break before critical words occurs combined with other prosodic means, F0 peaks in particular. Figure 1 shows one portion of the sequence in (1): ... me(d) # (40) upptill # (90) sju - # - procentiga lön(e)ökninga(r) ...

Steep F0 rises can be seen in the grave accented two-peaked contour of the compound word sjuprocentiga. (An increase of salaries of as much as seven percent was considered extreme and a disaster for Swedish economy.) The F0 adjustments are here combined with a slowing-down and pauses in two positions before the focused word. In addition, there is a break within the word, making it standing out even more. Frequent reductions (indicated by parentheses) of surrounding non-focused words also contributes to the highlighting of the focused word. Thus, the speaker make extreme use of prosody to achieve what is expected of her – to give the very essence of news to the listener.

4.3.2. Interview

Pauses and other slowing-down phenomena abound in the vicinity of focused words. However, as they may as well be a consequence of the on-line planning process (word retrieval), we cannot be sure about to what extent they are intended to serve as an informative means. There is no question however, that F0 is used very efficiently for this purpose. One such passage is exemplified in (4), where the speaker exploits focusing not only to highlight what she is saying, but also, by repeating the same structure and partly the same words adds rhythm and extra liveliness to her message. Figure 2 shows the three focused words helst, each at the end of a chunk of a speech followed by a pause.

(4)... den gäller ja inte hur som helst # (200) hur länge som helst # (40) och med vilka medel som helst # (60) ... '... it does not apply whatever the circumstances # for whatever period of time # and by whatever means # ...'

Figure 2. Focused words in three consecutive chunks.

4.4. Dynamics of prosodic features

4.4.1. News reading

The speaker not only changes between speaking faster and slower and between a very clear articulation and more reduced speech, but is very dynamic also in other respects. F0 varies between 129 and 351 Hz across the speech sample with a mean of 200 Hz. There is an extensive use of F0; focus words are frequent and have well marked F0 peaks. Intensity is at a generally high level, except a drop at the very final end of each of the five sentences.

4.4.2. Interview

The F0 range is wider in the interview, varying between a minimum of 118 and 427 Hz (mean 207 Hz). However, if the frequently occurring laryngealizations are considered, the range is even wider. The speaker, similarly as the news reader, makes extensive use of F0 to put extra weight on important words. However, unlike the news reader, the politician has a greater repertoire. Words heard as focused are sometimes produced with an expanded, and sometimes with a compressed F0 peak. An example of a series of such compressed peaks is found in:

(5)... vi som inte ställer upp som en man bakom # dom vi är terrorist # (90) ... 'we , who do not follow as one man behind # them we are terrorists # ...'

This passage thus includes both expanded (on terrorist) and compressed peaks (distinct but low peaks on upp, en and
man) on the words perceived as focused. The speaker express indignation or rather, very sharp disapproval. She does so first by giving strong emphasis (through excessive lengthening and strong vocal effort) separately to upp, en and man, and then finishes the sequence by the expanded F0 peak on the last word, terrorist, see Figure 3.

Thus, the impression of focus is not always tied to a great excursion of F0. Even if we normally expect high F0 peaks on focused words, we should expect speakers to exploit also less regular patterns [6]. In particular, we should expect skilled speakers to do so. The skill of the present speaker manifests itself in her use of a wide repertoire of argumentative and emotionally colored expressive acts. She easily shifts between being agitating, condemning, appealing etc. and the listener gets the impression of a self-confident and powerful person with great capacity in political debate. To this capacity contributes her ability to counteract interruptions by others, which she deals with by increasing the volume of her voice.

5. Discussion

Both the news reading and the interview data indicate that public speech produced by professionals differ in many respects from speech by non-professionals. The news announcer and the politician both have a very efficient style of speaking, with short pauses (as compared with non-professional speakers) occurring in syntactically motivated positions and other positions as well. The non-syntactic pausing, however, has different explanations in the two speech styles. In the news reading pausing within syntactic constituents, before semantically heavy words, strengthens the importance of what will come up. In the spontaneous speech in the interview, silent and filled pauses and the frequent slowing-down rather serve to counteract problems caused by the on-line speech processing; they allow time to plan ahead. Though the news reader efficiently use variations of F0 and duration and changes between fully articulated and reduced speech to serve informative demands, the politician has an even greater repertoire, partly because she speaks spontaneously and partly because she can go beyond being neutral and informative. The speaker in the interview does so with great skill. She elaborates her speech exploiting repetitions and rhythm, she alternates between very different strategies for focusing words, and she is skilled in expressing attitudes such as indignation and irony.

These findings shed light on how speech behavior can be optimally shaped to the demands of the situation. They also give indications as to what makes a skilled speaker. By a “skilled”, or “good” speaker here we refer to a person having the capability to attract the interest of the listeners, and who does so not only because he/she may have an interesting message, but because he/she is good at expressing it. This should apply to speakers in the public domain, but no doubt can be generalized also to other contexts.

To be able to generalize these findings, future research will include analyses of a variety of speakers (more or less skilled) and the use of listener panels [cf. 9] to judge speaking behavior. Such judgments may also include synthetically manipulated speech to learn about how variations of individual prosodic features influence listeners.

6. Acknowledgements

This work was supported by a grant (421-2004-2338) from The Swedish Research Council.

7. References