Variability in articulation and timing in connected speech of different style

Lioba Faust
Institut für Kommunikationsforschung und Phonetik (IKP)
Poppelsdorfer Allee 47, 53115 Bonn, Germany
Phone: +49-228-735641 E-Mail: lfa@ikp.uni-bonn.de

Abstract

In this study, variation in speech utterances of the same speakers in different communication situations was examined. A corpus of seven dialogues was recorded each in four different types, produced by fourteen speakers. From the corpus, utterances by two speakers were selected for a listening experiment in order to find out to what extent listeners are able to identify spontaneous and read utterances. Variability in speech was analysed with regard to the speakers and to the different dialogue types. The global structure of speech production was observed for timing processes: speech rate, articulation rate and pauses were measured. Speech production was analysed as deletions, substitutions and insertions of sounds. The results revealed that variation is strongly speaker-dependent. However, casual speech tends to be less carefully articulated than read speech. The two speakers selected for the listening experiment showed contrasting articulatory patterns with corresponding classifications by the listeners. The results illustrate that speaking styles have to be regarded as productions of different communication situations with each one showing its own speaker-dependent articulatory pattern.

1. Introduction

It is well known that speech varies to a large extent depending on different speakers. It is also a fact that different communication situations have an influence on speech utterances. Research in phonetics and speech engineering has been focusing on the variability in connected speech collected not only under laboratory conditions but also in real-life situations. Up to now, speech engineering - and especially speech recognition - for the most part analysed read speech recorded under lab conditions. As in everyday use read speech can differ largely from speech uttered spontaneously, it is no longer sufficient to be restricted to read speech. The more research is performed, the less satisfying the rough distinction between read and spontaneous speech becomes. Considering speech utterances from different communication situations, "speaking style" was defined as a promising term. In order to collect real-life speech utterances from different speaking styles, a couple of techniques were developed, e.g. the Wizard-of-Oz-experiment, interview talks etc. Generally, the term "speaking styles" was used to characterize casual or informal speech, careful, formal or clear speech and read speech (Eskénazi 1993, Blaauw 1994, Beckman 1995).

Of course, the collection of speech data for acoustic-phonetic analysis is rather difficult and, to some extent, has to be performed under lab conditions. However, it is clear that the distinction between spontaneous and read cannot be sufficient to characterize speech utterances. Before determining utterances as belonging to a certain speaking style or type of speech, it is interesting to analyse their phonetic variability.

As a starting point, it is important to look at the speech we are used to in real life. Indeed, we find a lot of examples for speech uttered spontaneously, and less examples for speech read out aloud. In table 1 some examples are presented.

Table 1: Examples of communication situations

Communication situations and "speaking styles" of everyday conversation	
➤ chatting ➤ talking with friends ➤ speaking to the boss ➤ explaining something to foreigners ➤ telling jokes ➤ interview talks ➤ talking to a child ➤ talking to hearing-impaired persons ➤ discussing	➤ professional news-reading ➤ reading fairy tales to a child ➤ reading a letter/an article in a newspaper ➤ reading to a blind person ➤ professional reading of literature
phrased by the speaker	composed by another person
"spontaneous speech"	"read speech"

From the examples we learn that different "speaking styles" are found almost as often as there are different communication situations. The distinction between spontaneous and read speech is only related to the fact that spontaneous speech is phrased by the speaker himself or herself in the moment of speaking and that read speech is written down and, at least ordinarily, composed by another speaker. We cannot yet conclude that there are great phonetic differences between speech uttered spontaneously or read out, but we may perceptually observe variation in different communication situations and for the same speakers.

The present study attempts to give an answer to the following questions:

1. What about the *range of variability* within and between different speech styles?
2. Are there *crucial differences* between read speech and speech uttered spontaneously?
3. What happens to the *production of speech* when a speaker changes to another speaking style?

2. Description of the experiment

2.1 Design and performance of the speech corpus

In order to get rather natural utterances in real-life conversation, dialogues were recorded in four different types: fourteen speakers (eleven female, three male) between 21 and 27 years of age participated in seven dialogue recordings of German with each one being performed in the following different speech types:

1. Recording of a *totally free conversation* without any instructions. The speakers did not know that they were recorded and the communication situation was relaxed. The resulting speech was casual.

2. Recording of a *time-scheduling negotiation dialogue* using a formal mode of address. The resulting speech was supposed to be less casual than in the first dialogue.

3. The task for the speakers was the same as in 2., but the recording conditions were controlled, i.e. the speakers had to press a button before they started speaking, and they could not speak simultaneously.

4. The last recording was a re-reading of the transcribed utterances of the second dialogue type. Hesitations, word repetitions and repairs that had been produced and transcribed were generally dropped for the copy to be re-read. The most important issue for the re-reading copy was to preserve the dialogue structure and, in general, the grammatical structures of the utterances.

The speakers were sitting in the institute's speech laboratory. The communication was performed using headset microphones (Sennheiser HMD 414-6) for the dialogue types 2 and 3 and room microphones (Neumann KM 140) for the casual and the read conversation. Speech was digitally recorded on two separate channels.

2.2 The method of experimentation

2.2.1 Listening experiment

Ten phonetically educated and twelve naïve listeners (beginning students) were asked to classify the selected utterances as "spontaneous" or "read" in a forced choice task. Furthermore, the listeners had to rate the degree of reliability for their decision on a five-point scale from "very safe" to "very unsafe". Moreover, the essential linguistic or phonetic features underlying the listeners' decisions had to be specified. The given features were: syntactic structure, speech fluency, repairs, articulation, intonation, and speech rate.

For the listening experiment the selection of utterances out of the corpus was done on the following conditions: One female speaker was selected whose spontaneous and read speech patterns appeared to differ very much. A second female speaker from another dialogue was selected whose speech patterns seemed to be quite similar in both dialogue types. From each speaker, three utterances were selected from the dialogue types 1, 2 and 4. The utterances were of phrasal length, grammatically sentence-like, with different intonational patterns, and they contained a certain amount of pauses, hesitations and repairs. From the careful spontaneous and the read versions, identical utterances were collected. The utterances were presented in random order, and each utterance was played twice.

2.2.2 Proceeding for segmental analyses

All dialogue utterances were transcribed literally, segmented and labelled manually and marked with phrasal accents. The labelling was performed by the author and two other trained phoneticians.

It was decided not to be restricted to segmental analyses on the sound level, but to examine as well the global structure of speech production, i.e. **timing structure**. Timing was measured as

1. *speech rate* including any kind of pauses, hesitations or lengthenings,
2. *articulation rate*, i.e. fluently produced utterances, and
3. *"interruptions"*, i.e. pauses and sound-lengthenings, which were also measured according to their phrasal position.

Timing structure was decided to be measured as number of sounds/s (and, concerning speech rate, labels/s). The usual manner in measuring timing structure as the number of syllables/s (Kohler et al. 1981) was rejected for two reasons: 1. Especially in the first dialogue type, the number of syllables was very difficult to determine as non-accented syllables were, at least for some speakers, to a large extent contracted and, thus, deleted. Therefore, it was more obvious to examine sounds instead of syllables. 2. The analysis of sounds had to be performed anyway for the measuring of articulation.

Speech rate was measured as the number of labels/s, articulation rate as the number of fluently produced sounds/s where "interruptions" had been excluded. "Interruptions" were classified into breathing, silent and filled pauses, and sound-lengthenings. Each kind of interruption was labelled by a certain label symbol in order to make it easy to exclude them from the measuring of articulation rate and to examine them separately. Sound-lengthenings were also observed in different positions: 1. utterance-final, 2. phrase-final followed by a pause, 3. phrase-final, 4. interrupting (i.e. within a phrase) followed by a pause and 5. interrupting. The amount of pauses was examined as occurring non-interrupting (i.e. corresponding to a syntactic boundary) or interrupting.

Sounds were marked as lengthened using the following method: The average sound length of each sound category was measured for each speaker and each dialogue type, then calculated for all speakers. From the sounds that had been manually labelled as lengthened the lowest value of a sound category was noted and divided by the calculated average for each sound category. Thus, a coefficient resulted of which the average for all sound categories was calculated, and, afterwards, for all dialogue types. The resulting coefficient was 2.5. Multiplying the sound length of all labelled sounds by 2.5, the resulting values were compared to the real produced sound length by a computer program. Longer sounds were automatically labelled as lengthened, shorter sounds were not.

Measuring **articulation**, reductions were examined as the number of sound deletions, substitutions and insertions compared to the citation form of the corresponding word which was taken as reference. Furthermore, segmental reductions were measured according to sound categories (plosives, fricatives, nasals, vowels, glides, liquids).

The first step for the examination of speech production was to carry out the reference phonetic transcription using a computer program that converted the literal transcription into the citation form (P-TRA, created at IKP, Stock 1992).

The citation form of the words then was taken as the reference for the measurement of reductions. The comparison of the symbol strings was performed automatically by applying a window that included a variable number of symbols to be compared. Finally, the total amount of sounds produced by a speaker in a dialogue type corresponded to the labelling string. This labelling string was compared to the reference string. Thus, missing sounds in relation to the citation form were defined as deletions, sounds that were different from the citation form were defined as substitutions. Sound substitutions according to sound categories were analysed manually. Sounds left in the labelling string compared to the citation form were examined as insertions.

3. Results

A summary of results will be presented concerning the correlation between the different speech styles and dialogue types, respectively. Concerning the listening experiment, the results for the two selected speakers will be presented separately.

3.1 Speaker-related results

3.1.1 Listening experiment

For the listening experiment, identical utterances from dialogue type 2 (careful spontaneous speech) and 4 (read) were selected, and, in order to examine to what extent casual utterances could be classified as spontaneous, utterances from dialogue type 1 were presented to the listeners as well.

Table 2: Correct classification of utterances (percentage for 22 listeners)

Speaker	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 4 (read)
A	97.0%	93.9%	62.1%
B	90.6%	62.1%	43.9%

Table 2 illustrates the correct classification of utterances, i.e. utterances of dialogue type 1 and 2 were identified as spontaneous, and those of type 4 as read. For both speakers, the casual utterances were correctly identified in almost all cases. This by far clear result for the casual style may be due to the restricted acoustic conditions and to the informal content of the dialogue utterances.

The careful style requires a closer look at the individual speakers: The utterances of speaker A are to a large extent still classified as spontaneous, whereas for speaker B there occur misclassifications, i.e. more than a third of the listeners classifies the utterances as read. For the read speech style it is interesting to note that the utterances of speaker A are mostly classified as read, but that for speaker B the majority of listeners classifies the read utterances as spontaneous. From these results it may be concluded that casual speech is identified as spontaneous, whereas the identification of careful and read speech seems to be dependent on the speaker.

As to the reliability of listeners' decisions, listeners are generally "very safe" or "rather safe" about reaching a correct decision.

Table 3: Correct classification due to *linguistic features* (Most frequent answers by trained listeners)

Speaker	dialogue type 1 ("casual")		dialogue type 2 ("careful")		dialogue type 4 (read)	
A	articulation speech rate	63% 63%	fluency	73%	intonation	67%
B	intonation fluency	48% 44%	syntax	59%	intonation	75%

For the correct decisions, the distribution of the phonetic or linguistic features was examined. The most frequent answers by trained listeners are illustrated in table 3. Concerning the casual style, all features are mentioned more or less frequently by the listeners. In the careful style, "fluency" is most often mentioned for speaker A, for speaker B it is "syntactic structure". For the read speech style "intonation" is the convincing feature for the listeners to classify the utterances of both speakers correctly.

The results from the listening experiment show that especially for casual speech there are different linguistic features that dominate the perceptual impression of the listeners leading them to a correct classification. From this we learn that it is necessary to have a look at the "patterns" in speech in order to understand the listeners' decisions.

3.1.2 Timing structure

Since the analysis of speech rate contains all "interrupting parts" of the utterances, speech rate does not reveal satisfying results. To get an impression of the timing structure in the utterances of the two selected speakers, it is most interesting to examine articulation rate for the three dialogue types that had been presented to the listeners. The values of speech rate and interruptions for the whole corpus will be shown in section 3.2.1.

Table 4: Articulation rate (number of sounds/s)

Speaker	casual (1)	careful (2)	read (4)	development
A	16.9	14.6	13.7	↘
B	13.0	16.3	16.3	↗

Table 4 illustrates articulation rate as number of sounds/s. Comparing the results for the two speakers, the contrast is immediately perceivable: Whereas for speaker A the highest articulation rate occurs in the casual speech type, for speaker B it occurs in the read dialogue type. For both speakers the development from the casual towards the read speech type is obviously contrasting, i.e. for speaker A it is decreasing, for speaker B it is increasing. From these results we can conclude that speaker A speaks slower when reading than when speaking spontaneously, and, in contrast to her, speaker B articulates faster in read than in casual speech. Considering the total corpus of speakers, another important item has to be mentioned: The articulation rate of speaker A in casual speech is the highest compared to all speakers, speaker B reaches in casual speech the lowest, in read speech the highest rate in the total corpus.

These contrasting results for articulation rate may already explain the decisions of the listeners in the listening experiment.

3.1.3 Speech production

The results for speech production concerning the two selected speakers will also be restricted to the three dialogue types that were presented in the listening experiment.

Table 5: Deletions (missing sounds related to the total amount of sounds produced; speaker-specific)

Speaker	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 4 (read)	development
A	17.6%	8.5%	4.5%	↘
B	16.4%	8.0%	10.4%	↘↗

Table 6: Substitutions (related to the total amount of sounds produced; speaker-specific)

Speaker	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 4 (read)	development
A	13.2%	6.7%	4.9%	↘
B	10.2%	6.8%	9.9%	↘↗

For both speakers, the examination of reductions yields very clear results. Whereas the amount of reductions decreases for speaker A from casual speech towards read speech, the development for speaker B is vice versa. Table 5 illustrates the amount of deleted sounds compared to the total amount of sounds produced in this speech type.

In casual speech it is interesting to note that the amount of deletions for both speakers is similar, i.e. they do not articulate very carefully. In dialogue type 2 both speakers still reach a similar amount of reductions, and much less than in casual speech. When reading, speaker B increases the amount of deletions, whereas speaker A decreases even further.

For the articulation rate, the results are corresponding: They illustrate that speaker A articulates more slowly and more carefully the less "spontaneous" the situation, whereas for speaker B reading is not very different from speaking spontaneously as she articulates faster and just as little careful as in casual speech.

We can conclude that the manner of articulation is different for the two speakers according to the communication situation.

The results for substituted sounds (Table 6) show a similar pattern as for deletions and, thus, underline the development of speech production for the two speakers.

The conclusion that can be drawn from the experiments is as follows: For each speaker, the results for speech production, articulation rate and perceptual classification correspond. The less spontaneous the communication situation, the more careful the articulatory pattern of speaker A and the slower the articulation. Her utterances generally are correctly classified. The articulatory pattern of speaker B when reading is faster, less careful and her utterances are correctly classified only by a small number of listeners. Thus, it may be concluded that listeners need contrasting patterns to identify utterances as spontaneous or read.

3.2 Results for the whole corpus

Assuming a Gaussian distribution, the speakers A and B would be situated in the extreme positions on both ends of the scale. To look for speaker-independent results and in order to find characteristics in the different speech types the examination of all speakers is indispensable. Therefore, the results for all four dialogue types will be illustrated.

3.2.1 Timing structure

Table 7 illustrates the results for articulation rate. As pauses and interrupting parts of the utterances were excluded for the measurement, articulation rate was expected to yield promising results. Comparing the values for the average as well as for median or standard deviation it is evident that there are no remarkable differences between the 4 styles, i.e. in all dialogue types almost 15 sounds/s are produced. From this it must be concluded that the results are speaker-specific.

It is not very surprising that the results for speech rate (table 8) are unsatisfactory since the examination included also non-fluent utterance parts which vary widely in frequency and length for the different speakers.

In order to shed more light on timing structure it is indispensable to have a look at the distribution of the "interrupting parts", i.e. pauses and sound-lengthenings.

Table 7: *Articulation rate* (number of sounds/s for all speakers; style-related)

Dialogue types	average	standard deviation	median	minimum	maximum	number of speakers
dialogue type 1	14.59	1.05	14.55	13.02	16.87	14
dialogue type 2	14.86	1.05	14.43	13.39	16.59	14
dialogue type 3	14.55	0.96	14.46	12.75	16.13	14
dialogue type 4	14.84	1.49	14.92	12.62	16.72	14

Table 8: *Speech rate* (number of labels/s for all speakers; style-related)

Dialogue types	average	standard deviation	median	minimum	maximum	number of speakers
dialogue type 1	11.52	1.01	11.40	10.24	13.98	14
dialogue type 2	11.27	1.58	11.37	8.96	13.92	14
dialogue type 3	11.43	1.58	11.62	7.23	13.42	14
dialogue type 4	12.53	1.49	12.35	9.34	14.63	14

3.2.1.1 Distribution of pauses

Pauses were divided into breathing and silent and filled pauses (hesitations). The examination revealed that global results for the speech corpus can be drawn, but they must not obscure the fact that the distribution of pauses is speaker-specific.

Comparing the three pause types, a hierarchy can be observed: In all dialogue types, breathing pauses are most frequent, followed by silent pauses, followed by filled pauses (table 9).

Table 9: Length of pauses related to the total time of speaking for all speakers; style-related

Pause type average in %	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 3 ("careful")	dialogue type 4 (read)
breathing	5.08	7.11	6.67	6.14
silent pauses	6.46	4.87	4.95	2.21
filled pauses	2.30	3.25	2.52	1.15

Table 10: Pauses in non-interrupting position related to the total amount of each pause type; style-related

Pause type average in %	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 3 ("careful")	dialogue type 4 (read)
breathing	68.47	79.06	80.75	84.88
silent pauses	59.52	58.65	69.79	86.31
filled pauses	32.56	47.54	59.68	75.97

Breathing occurs in all dialogue types mostly *between* fluent utterance parts, i.e. non-interrupting (cf. table 10). This is found increasingly the less spontaneous the communication situation. It is clear that the structural planning in read speech is much lower than in spontaneous speech and, thus, in read speech speakers obviously breathe *before* a new phrase is uttered (85% non-interrupting). In the casual speech type the amount of non-interrupting breathing pauses is lower (68%, cf. table 10) as the planning may require more attention.

The examination of **silent pauses** yields interesting results as we find a strong decrease of occurence from casual over the two careful dialogue types towards read speech (table 9). The lack of silent pauses in read speech may be explained by reading skills of the speakers. Normally, reading should also contain a certain amount of silent pauses. In non-professional reading we often find a lack of pauses. Concerning the position of silent pauses, they are largely non-interrupting in read speech (86%) which is due to the same reason as for breathing in this case. In the casual dialogue type a smaller amount of silent pauses is non-interrupting (60%). Maybe the structural planning of the utterances provokes the pauses also *within* phrases.

The least frequent pause type in all dialogue types are **filled pauses** (table 9). Comparing the dialogue types, they are most frequent in the careful speech style (dialogue type 2). This may be due to the subject of the dialogue which was new and uncommon for the speakers. Filled pauses occur in all speech styles apart from casual speech largely non-interrupting. The results reveal that the amount of interrupting hesitations is higher than for silent and breathing pauses. For casual speech it is interesting to observe that filled pauses occur mostly *within* phrases (only 33% non-interrupting, cf. table 10). This may be due to the fact that the communication situation was absolutely spontaneous without any tasks for the speakers, such that a role change was possible as often as the speakers intended to and, therefore, phrasing was interrupted.

3.2.1.2 Sound-lengthenings

Sound-lengthenings are found to occur to some extent in all speech styles and the amount is similar: about 2-3% of sounds are lengthened compared to the total number of sounds produced in a single dialogue type.

The average in the four dialogue types is 1.9% (1) - 3.0% (2) - 2.4% (3) - 2.2% (4). It is difficult to give a reliable interpretation. An analysis of the results for the single speakers reveals that the frequency of sound-lengthenings does not vary as strongly as for other parameters. As the lowest value in the total corpus is found in casual speech it may be concluded that the most natural communication situation provokes the least lengthenings. Generally, the occurence of lengthenings seems to be largely independent from speech styles and rather due to speech rate and fluency of an individual speaker. With regard to these very close results in all dialogue types it is necessary to have a look at the position of occurence.

Generally, lengthenings occur more and more before syntactic boundaries the less spontaneous the speech style. Whereas in casual speech the amount of non-interrupting lengthenings is about 45%, in read speech non-interrupting lengthenings take the most part (75%). In casual speech the reason for the high amount of interrupting lengthenings may be the structural planning of the utterances. However, it must not be neglected that 22% of lengthenings occur utterance-finally. It is interesting to note that in dialogue type 3, where the recording conditions had been strongly controlled, the amount of utterance-final lengthenings decreased (only 8%) compared to the dialogue types 1 and 2. The task for the speakers provoked different structural planning and led to longer utterances than in the other dialogues. It might be for the same reason that phrase-final utterances followed by a pause take the most part in all dialogue types. It is evident that the lack of structural planning in read speech yields an only small amount of interrupting lengthenings.

The reasons for sound-lengthenings vary with respect to articulation rate, the subject of the dialogue, the relation between the speakers, speech habits and the communication situation. The results for sound-lengthenings once more reveal that each communication situation has to be seen on its own.

3.2.2 Speech production

Reductions were measured as deletions, substitutions and insertions of sounds. The results for all three reduction types will be discussed.

Insertions cannot be classified as reductions, but they were measured as modifications of speech production in the four dialogue types. The examination revealed that in all dialogue types to some extent sounds are inserted. It can be interpreted that insertions occur rather due to the phoneme context and to the individual speaker than due to the communication situation.

In table 11 the average of deletions and substitutions related to the total number of sounds produced by all speakers is illustrated. Both reductions occur for the most part in casual speech and least in read speech. At first sight this result looks very clear, so that casual speech is to be judged as less carefully articulated than read speech. As we have learned from the listening experiment, articulation was not the only feature for the listeners to classify an utterance as spontaneous or read. Moreover, timing structure yielded different results for the single speakers. Furthermore, reading skills and individual speaking styles must not be neglected as influencing features on speech production.

Table 11: *Deletions* and *substitutions* related to the total number of sounds produced by all speakers

Reductions (average in %)	dialogue type 1 ("casual")	dialogue type 2 ("careful")	dialogue type 3 ("careful")	dialogue type 4 (read)
deletions	14.46	8.18	8.34	6.87
substitutions	10.70	6.75	6.95	6.28

Comparing the four dialogue types it can be concluded that to some extent the tasks of speaking were responsible for the occurence of reductions. It is important to note that the amount of deletions, and also substitutions, is definitely higher in casual speech than in the careful spontaneous speech styles and, of course, in read speech as well. This may be due to the very natural communication in the first dialogue type that in some cases led the speakers to a rather informal situation.

The close results in the dialogue types 2 and 3 reflect the similar tasks of the dialogue subject in both situations. For read speech it can be interpreted that the concentration on the printed text and the engagement in a rather careful reading yielded less reductions.

To explain reductions in more detail it is necessary to examine their distribution according to sound categories. Concerning deletions, in all dialogue types mostly plosives are deleted which often concern word-final plosives. After plosives, the next most frequently deleted sounds are vowels. This is due to the fact that syllables are contracted such that the vowel is no longer present. This manner of reduction often occured across word-boundaries, e.g. in the sequences "ja aber" (yes, but) or "aber ich" (but I). Especially in less formal communication situations (casual style, but also the dialogue types 2 and 3) vowels often are deleted when they occur as inflectional endings of verbs, e.g. "denk' ich" instead of "denke ich" (I think), "wär's" instead of "wäre es" (it would be). Substituted sounds are for the most part vowels. This can be explained by taking into account that vowels are influenced by the phoneme context and vowel context, respectively, i.e. in most cases vowels are assimilations of the surrounding syllables.

The examination of speaker-specific results in speech production reveals that deletions and substitutions are very speaker-dependent although the global result for the total corpus gives more evidence than did the other parameters that were measured.

4. Conclusions and discussion

The experiment revealed interesting results concerning the variability of speech production in different speech styles. A particular claim of the study was to see whether there are crucial differences between read speech and speech uttered spontaneously. The listening experiment yielded contrasting results for two selected speakers. Of course, apart from the selection of

utterances and the acoustical conditions, these results might be influenced by reading skills. If reading skills are to be defined as the ability of perfectly transferring visual print patterns into acoustic patterns, then the utterances of speaker A should be considered as a perfect reading. The utterances of speaker B, however, are much more fluent and more natural-sounding than those of speaker A, although (or because) they contain more reductions. As the speech patterns of the two speakers are the most contrasting in the total corpus, the range of variability independent from the dialogue types is extensive. Thus, generally, read and spontaneous utterances are difficult to distinguish.

The listening experiment also reveals the way listeners expect speech to be, i.e. read speech as rather slow, carefully articulated and very intelligible, and spontaneous speech as rather slurred and not quite well articulated and, moreover, containing pauses and interruptions. The conclusion drawn from the listening experiment is, thus, that listeners are able to identify spontaneous and read utterances only if the differences are evident.

The answer to the question what about the range of variablity within a style is to be seen as speaker-specific. There are speakers whose speech patterns are even contrasting within the same style.

Concerning speech production, the patterns both in articulation and timing structure changing from one style to another are not corresponding for all speakers. Whereas articulation rate is very speaker-dependent, speech production reveals a tendency of "reduction patterns": more reductions occur in spontaneous speech than in read speech, but, keeping in mind reading skills and the claim to articulate carefully when reading.

The question whether there are crucial differences in read and spontaneous speech is to be answered as follows: The distinction between spontaneous and read is superficial, and is not a classification of articulatory variability. Furthermore, speech utterances are adapted to a given particular communication situation.

Therefore, speech should be seen as a spectrum of styles depending on situational facts and on speaker-specific facts, i.e. there are as many speech styles as there are communication situations.

Acknowledgement

This research was partly supported by the German Federal Ministry of Education, Science, Research and Technology (BMBF).

References

Beckman, M.E. (1995): A typology of spontaneous speech. In: Proc. ATR International Workshop on computational modeling of prosody for spontaneous speech processing, Kyoto, 2-23 - 2-34.

Eskénazi, M. (1993): Trends in Speaking Styles Research. In: Proc. ESCA Eurospeech, Vol. 1, 501 - 508.

Stock, D. (1992): P-TRA - Eine Programmiersprache zur phonetischen Transkription. In: Beiträge zur angewandten und experimentellen Phonetik. Steiner Verlag, Stuttgart.