RELATIONSHIP BETWEEN DISCOURSE STRUCTURE AND DYNAMIC SPEECH RATE

Florien J. Koopmans-van Beinum and Monique E. van Donzel

Institute of Phonetic Sciences/IFOTT, University of Amsterdam,
Herengracht 338, 1016 CG Amsterdam, The Netherlands
e-mail: florienk@fon.let.uva.nl

ABSTRACT

This paper regards one specific element of a larger research project on the acoustic determinants of information structure in spontaneous and read discourse in Dutch. From a previous experiment within that project it turned out that listeners used two main cues (viz. speaking rate and intonation) to differentiate between spontaneous and read speech [7]. The aim of the present experiment is to investigate the role of one of these prosodic cues, i.e., the local variability in speaking rate, and to study the relationship between the information structure of a spoken discourse at the one hand, and dynamic speaking rate measurements of that discourse at the other hand. Results show that there is a large variability in average syllable duration over the various interpausal speech runs for each of the eight speakers. No straightforward relation is found between the number of syllables within a run and the average syllable duration. We hypothesize that, at least in spontaneous speech, variations in speaking rate are related to the (global and/or local) information structures in the discourse. Global analysis of the discourse structure in paragraphs and clauses reveals that for each of the speakers the average syllable duration of the first run of a paragraph is longer than the overall mean value per speaker in more than 60 % of the cases. Inspection of the quartiles of runs with highest ASD-values and those with lowest ASD-values for each of the speakers shows quite different structures, which can be explained on the basis of partly local and partly global discourse characteristics.

1. INTRODUCTION

When listening carefully to somebody telling a story, one of the most striking aspects is the fact that there are many irregularities with respect to the fluency of the speech: the speaker is alternately speeding up and slowing down his/her speech production, using pauses and using variations in speech tempo. For the larger part this will be the result of planning the discourse: the time necessary to adeqately formulate what has to be told. Other than by adaptations for speaking styles and situational circumstances (number of listeners, reverberation, etc.), the speaker may create the possibility to plan the discourse and to reorganize this discourse planning, if necessary, by means of a specific pausing strategy [8].

However, apart from pauses, also a large variability within the speaking rate of every speaker is quite obvious. So far research on speaking rate in Dutch has mainly concentrated on overall measures over a whole discourse or over paragraphs [2], [3]. In the present project we try to relate more local variations in speaking rate to the structure of the discourse. So questions to be answered here are: do speakers use the possibility to variate their speaking rate in a systematic way, and as to how far is this variability related to the specific global or local structure of their discourse.

With respect to read discourses Crystal and House [1] studied durational characteristics of supraphonic units like syllables, stress groups, and interpausal runs. As a unit of analysis they used the duration of interpausal runs expressed in average syllable duration (ASD). Their results indicate that "stress characteristics are basic to the ASD variability seen in connected speech" (p.107) and "that the variability is not random, nor talker idiosyncratic, but is a function of the syllabic and stress characteristics of the materials" (p. 108). These results made us decide to use the same unit of analysis, i.e., the duration of interpausal runs for our connected speech materials, and to also express durational characteristics in average syllable duration (ASD). However, it is quite obvious that speakers 'spontaneously' telling a story may differ in their speech durational behaviour from speakers reading a 'prepared' story, as in the case of Crystal and House [1]. Using the same tools will give us a possibility to compare results. Results from our previous studies suggest that, at least in spontaneous speech, those parts or utterances that contain highly important information (indicated by textual analysis and by perceptual judgement), are produced at a slower speaking rate than parts expressing information that can be considered as being of less importance to the listener [3], [4]. Results from the present, more extended study will enable us to determine the relationship between the information structure of spoken discourse and the local variations in speaking rate in various speaking styles, and to determine whether speakers agree in the way they apply speaking rate variations in discourse. In a later stage results will be combined with data on pausing strategies, with data of intonation analyses, and with those on perceived prominence.

Our first assumption to be tested is that ASD is an adequate measure, reflecting variations in speech duration in general, when a discourse is divided in interpausal runs. This means that ASD should not have a constant value for each run per speaker. Another possibility is that a close negative relationship exists between the number of syllables in a run and the ASD over each run. This would mean that discourse planning might happen per interpausal run and that within a run speech production may behave like in multi-syllabic words: long duration for mono- or two-syllabic words and short duration for multi-syllabic words (e.g., [5]). A third possibility is that speakers display variable ASD-values per run, that cannot be accounted for by means of syllable numbers. In that case the explanation for the variability in speaking rate has to be found in the global and/or local structure of the discourse, probably in relation to the distribution of stressed syllables.

2. METHODS

2.1. Speakers and speech material

Since in the main project it is our aim to compare the results from spontaneous speech with those from read speech for a number of speakers, it was necessary to collect well comparable speech

materials for both speaking styles. For that purpose we asked eight speakers of standard Dutch (4 male and 4 female) to read aloud a short story in Dutch. Then the same speakers retold the story in their own words. Next we made verbatim transcriptions of the retold ('spontaneous') versions, and then each speaker was asked to read aloud the transcribed version of his/her narration ('re-read'). They were allowed to prepare themselves carefully and they had to indicate their own punctuations and clause structures. The various versions were stored as digitized audio files (low pass filtered 24 kHz, sample rate 48 kHz, 16-bit precision).

Subsequently the spontaneous texts were analysed for discourse structure, using an objective method with different markers for different discourse determinants [6], indicating the information status of the concepts on a global level (the division of the discourse in clauses and paragraphs) and on a local level (new, inferrable, evoked, discourse marker). Apart from this, a perceptual evaluation (by 12 listeners) was obtained for the spontaneous versions of the texts of all speakers, in order to compare, in a later stage of the project, the perceived discourse structure and the prominence judgements with the results of the textual analysis of the discourse.

2.2. Measurements

Before explaining the actual measurements, first we will carefully define the terms on speaking rate as we used them in this study.

As said above we chose the interpausal run as our basic unit of analysis. For this purpose the notion 'pause' had to be defined in a rather general way. It turned out that our speakers used three types of pausing: 1) silent pauses, i.e., no speech sounds at all during more than 150 ms; 2) filled pauses, i.e., a hesitation sound ('eeh'), preceded and/or followed by a silence; 3) lengthening of certain words, often by means of a connected hesitation sound ([8] for more details on pausing strategies). In the present study the pauses of type 1) and type 2) defined a run. Words containing type 3) are considered in a specific way: if the filler could be separated from the word it was connected with, the filler was counted as a type-2 pause, but if the filler could not be separated, the whole word was left out of consideration.

As for definitions on speaking rate (so far, apart from the title, we only used this term), we decided to use 'speaking rate' when pauses are included, and the term 'speech rate' when pauses are not included, so reflecting the actually produced speech.

With respect to the acoustic rate characteristics, we measured 'speaking rate' (pauses included) at discourse level (globally over the whole text) and at run level (locally for every run including its following pause), and we measured 'speech rate' (pauses not included) at interpausal run level, for the 'spontaneous' (=retold) version for each speaker (values expressed in seconds). Total pause duration per speaker was also measured for each run and for the whole discourse per speaker. Variability in speech rate is expressed in average syllable duration (ASD) in ms. Since in a number of cases interpausal runs existed of only one syllable, these are left out of the ASD calculations. Number of words and number of syllables were counted for the whole discourse, and number of syllables were counted for each interpausal run, for each of the eight speakers.

3. RESULTS

3.1. Global measures

Table 1 gives an overview of numbers of words and syllables, summed durations of speech and of pauses, average duration per syllable (ASD), total speaking duration (speech plus pauses), and the ratio between pause and speech durations, all broken down per speaker.

A striking aspect, revealing from Table 1, are the high values for the ratio between pause and speech durations. A large proportion, ranging from about a quarter of the total discourse duration for speaker 7 to almost half of the discourse duration for speaker 6, is used for some kind of pausing. Although comparable data on the read-aloud speech material are not yet available at this moment, we will undoubtedly find here a large difference between the two speaking styles. The very global measure of overall correlation between total number of words and number of syllables is high (.96), as could be expected, just as the overall correlation between speech duration (pauses excluded) and number of syllables (.84).

3.2. ASD at run level

Since we are interested mainly in more detailed and local aspects of the average syllable duration (ASD) per run, the data on the ASD-values together with standard deviations and range values, and the correlations between speech duration and number of syllables within a run per speaker (Table 2), will be more revealing. These data will give us a more specific insight into the background of the variability in durational aspects, since they may answer our questions on a probably constant ASD-value, a negatively correlated one, or a variable one, as explained in the introduction.

Table 1. Overview of numbers of words and syllables, summed durations of speech and of pauses, average duration per syllable (ASD), total duration, and ratio between pause and speech durations, broken down per speaker.

speaker	n. of words	n. of syll.	speech dur.(sec)	ASD (ms)	pause (sec)	total dur. (sec)	pause ratio
1	537	709	140	198	84	225	.38
2	459	620	96	154	74	169	.44
3	582	805	141	176	61	202	.30
4	504	677	106	157	62	169	.37
5	491	590	95	161	45	140	.32
6	361	472	94	199	83	177	.47
7	417	571	99	173	32	131	.24
8	511	707	114	161	51	165	.31

Table 2. Overview of ASD-values (in ms) together with standard deviations and range values, and correlations of speech duration and number of syllables per interpausal run, broken down per speaker.

speaker	ASD	st.dev.	max.	min.	range	correl.
1	198	67	483	133	350	.94
2	154	49	404	111	292	.94
3	176	38	295	121	174	.97
4	157	32	293	92	201	.95
5	161	27	234	103	131	.97
6	199	56	394	121	273	.92
7	173	33	308	129	179	.96
8	161	49	425	107	318	.95

It will be clear from Table 2 that the variability in ASD-values is very large, with standard deviations up to 67 ms. But also the variability in number of syllables per run is very large, ranging from 2 to over 40 syllables per run (not displayed in the table). Therefore the possibility that ASD-values remain constant, independently of number of syllables, has to be rejected. Nevertheless, as can be seen in the last column of Table 2, the correlation between speech duration and number of syllables per interpausal run is very high for each of the speakers. So we will have to account for this variability in ASD-values over the various runs.

The next item to be tested is the possibility that a close negative relationship would exist between the number of syllables in a run and the ASD over each run. This might be the case if no time-consuming discourse planning occurs *within* a run and therefore a run could behave like a multisyllabic word. However, calculation of the correlations between number of syllables and ASD over runs per speaker provides low correlation values for each of the speakers (ranging between -.27 and -.47).

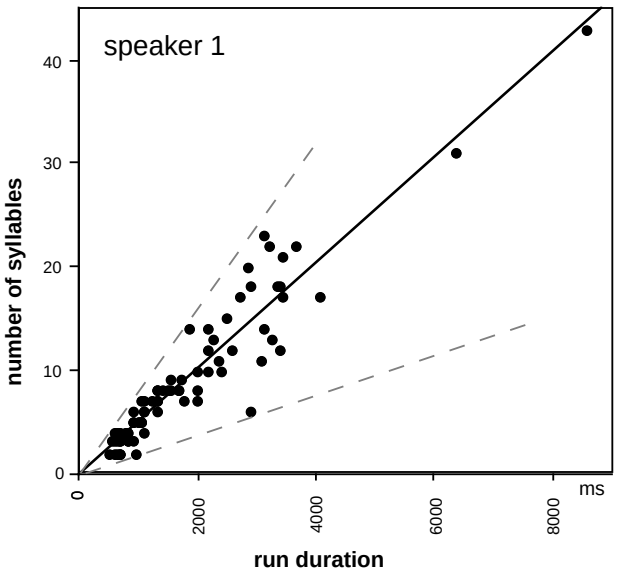

Fig. 1. Scattergram of speech duration and number of syllables per interpausal run, for speaker 1. Dashed lines indicate the variability in ASD.

So generally speaking, long runs containing many syllables do not display shorter ASD-values than runs with only a few syllables (remember that one-syllabic runs have been left out of consideration). Therefore, further explanations will be sought in the structure of the various discourses. To illustrate the problem, we displayed a scattergram of the relationship between speech duration and number of syllables per interpausal run, indicating the variability in ASD as well, for one of our speakers (Fig. 1).

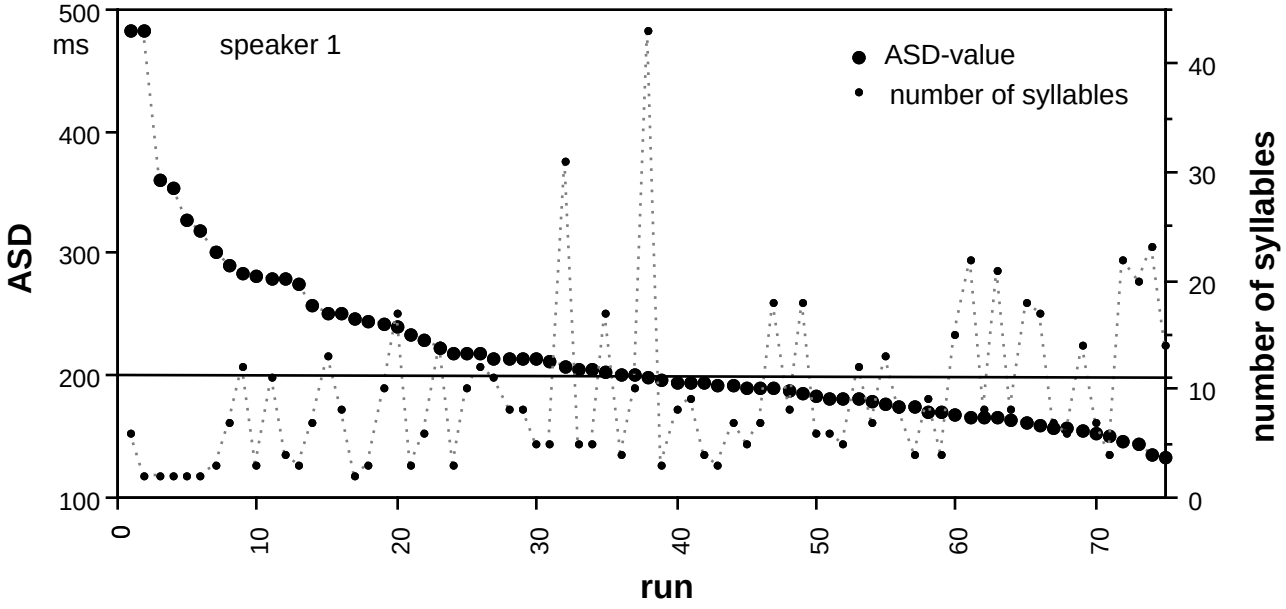

Fig. 2. ASD-values (large dots, in order of descending ASD-values) and concurrent number of syllables per run, for speaker 1. The median ASD-value is indicated by the continuous line.

For the same speaker 1 the relationship between ASD-values (in descending order) and concurrent number of syllables is given in Fig. 2. From these figures it will be clear that in a number of cases indeed a negative relationship exists between ASD-values and number of syllables, but that so many runs display a different picture, probably caused by the specific structure of the discourse at those places.

3.3. ASD and paragraph structure

To further study the relationship between global and local discourse structure on the one hand and the variability in ASD per run on the other hand, we first attended to more global structures of the discourses. For each speaker we investigated whether every first run after a paragraph boundary displayed higher ASD-values than the median ASD, assuming that important information comes at the start of a new paragraph. Results show that indeed there is such a tendency: in 60 % of the runs immediately following a paragraph boundary, the ASD-values exceed median values (with a range from 40 to 80 % over all speakers). However, we think this tendency is not really convincing. Therefore we decided to inspect the quartile with the highest ASD-values and the quartile with the lowest ASD-values per speaker in more detail, in relation to local discourse structure characteristics.

3.4. Highest and lowest ASD-values

When inspecting the two quartiles with the most extreme ASD-values per speaker, a number of striking aspects are met. In the first place runs with a low number of syllables are much more frequent in the first quartile, and runs with a high number of syllables are much more frequent in the last quartile, for all eight speakers. However, since several runs disprove this tendency, no high correlations can be found. Nevertheless the two-, three-, and four-syllabic runs in the first quartile almost all consist of an information status that has been marked as new in the discourse analysis [6], meaning information that has to be put in focus. We expect that these syllables will always be marked as prominent in the perception test (that will be worked out at a later stage). Two-, three-, and four-syllabic runs in the last quartile occur rarely.

Part of the short runs in the first quartile exist of discourse markers, followed by a (long) pause. Moreover, since the one-syllabic runs, that have been left out for processing, almost always exist of discourse markers of long duration followed by a long pause, this group in the discourse analysis may be considered as accountable for slowing down the speaking rate in a discourse to a great extent.

Another striking aspect concerning the runs in the first quartile for each of the speakers is, that these runs in almost all cases concern the main topic of the story, whereas the runs in the last quartile mainly concern expansions.

4. CONCLUSIONS AND DISCUSSION

The main conclusion of our study must be that accounting for variations in speaking rate of what may be considered as 'spontaneous speech', is a very complicated task. At this stage in the study we only used durational measurements related to aspects of discourse structure. Therefore, further acoustic characteristics like intonation and prominence, to be included in the project at a later stage, may account for another part of the variability in speaking rate.

It is clear that for each of the speakers a large variability in average syllable duration over the various interpausal speech runs exists, that no straightforward relationship is found between average syllable duration and number of syllables in a run. The structure of the discourse, when divided in paragraphs, accounts for a small part of the variation, but most explanations can be found when studying the runs with the most extreme ASD-values separately, in relationship to a hierarchical analysis of the discourse in at least topic lines and expansions.

In the near future we will first explore the methods of hierarchical discourse analysis which normally are used for read discourses, and test whether they can be used in a quantitative way for our spontaneously retold stories as well.

Next we will compare the present results on spontaneous speech with comparable measurements on the concurrent re-read versions.

Finally we will combine these findings with data on pausing strategies, on intonation, and on perceived prominence.

5. REFERENCES

1 Crystal, T.H. & House, A.S. "Articulation rate and the duration of syllables and stress groups in connected speech". J. Acoust. Soc. Am. 88: 101-112, 1990.

2 Den Os, E.A. Rhythm and tempo of Dutch and Italian; a contrastive study. Ph.D.-thesis Utrecht University, 1988.

3 Koopmans-van Beinum, F.J. "The role of focus words in natural and in synthetic continuous speech: acoustic aspects". Speech Communication 11: 439-452, 1992.

4 Koopmans-van Beinum, F.J. & Pols, L.C.W. "Naturalness and intelligibility of rule-synthesized speech supplied with specific spectro-temporal features derived from natural continuous speech". Proceedings ICSLP 94, vol. 4: 1787-1790, 1994.

5 Nooteboom, S.G. Production and perception of vowel duration. A study of durational properties of vowels in Dutch. Ph.D.-thesis Utrecht University, 1972.

6 Van Donzel, M.W. & Koopmans-van Beinum, F.J. "Evaluation of discourse structure on the basis of written vs. spoken material". Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Vol. 3: 258-261, 1995a.

7 Van Donzel, M.W. & Koopmans-van Beinum, F.J. "Prominence judgements and textual structure in discourse". Proceedings of the Institute of Phonetic Sciences Amsterdam 19: 11-23, 1995b.

8 Van Donzel, M.W. & Koopmans-van Beinum, F.J. "Pausing strategies in discourse in Dutch". These Proceedings, 1996.