INTRASPEAKER VARIATION ON THE SEGMENTAL LEVEL: A TRANSCRIPTION-BASED APPROACH

A.P.A. Broeders* & W.H. Vieregge**

*National Forensic Science Laboratory, Rijswijk, Netherlands
**Dept. of Language & Speech, University of Nijmegen, Netherlands

ABSTRACT

This paper discusses a set of procedures which may be used to examine intraspeaker variation on the segmental level. The primary tool employed for this purpose is the consensus transcription. A variation index is proposed which captures the amount of intraspeaker variation around the modal realization of each variable. The procedures described should provide a principled approach to the investigation of intraspeaker variation with special relevance to the subject of speaker identification.

1. INTRODUCTION

While it is generally recognized that intraspeaker variation poses a major problem in speaker identification, comparatively little is known about the way in which this type of variation manifests itself in the speech of individual speakers. It is not clear, for example, whether speakers differ consistently in the amount and nature of intraspeaker variation associated with their speech. In recent years, there has been a marked increase in the number of studies dealing with inter- and intraspeaker variation, many of them undertaken with the prime object of answering questions in the field of speech technology. In spite of the current interest in speaker characteristics, there is still a remarkable scarcity of data at even the most basic level about the actual extent of variability in the speech of individual speakers. The present study seeks to develop a systematic approach to this question. However, unlike many other studies in this field, ours is not inspired by issues arising from speech technology and may therefore be of only marginal interest to it. We are aiming to devise an approach which is primarily relevant to auditory speaker identification. The primary tool employed for this purpose is the consensus transcription. Presented below are the preliminary results of this approach.

2. PURPOSE OF THE STUDY

Our main objective is to gain a better understanding of the magnitude and nature of intraspeaker variation through the use of a consensus transcription. Some of the questions we would like to answer are: Do some speakers consistently exhibit more variation than others?; Is it possible to express speaker variation in quantitative terms, and if so, how much material is required to arrive at a reliable index of intraspeaker variation?; Are some variables more consistent than others?; Is variation constant against time?

In order to investigate these questions non-contemporary speech samples were collected from 6 speakers of Dutch and subsequently transcribed according to the principles outlined below.

3. CONSENSUS TRANSCRIPTION

The concept of the consensus transcription is not new. Shriberg et al. [2] recommend it as a procedure which can be used to eliminate errors due to inattention and other shortcomings of the transcriber. They found that, of the corrections made by transcribers in a consensus transcription, 90% of those in vowel segments and 80% of those in consonant segments were considered by the transcribers to be due to inattention on their part during the original transcription process. Also, Ting et al. [3] have shown that within a group of transcribers mutual corrections lead to greater agreement between transcribers.

In the present instance, all speech samples were first transcribed by pairs of Language & Speech Pathology students of the University of Nijmegen - all of them qualified speech therapists - as part

Speaker	<3>	Svara.	Voicing	Elision	<ts>
MJ	3 (.39),3	0 (.79),1	+ (.50),2	0 (.67),2	ts (1.00)
JK	˚3(.78),2	ə (.96),1	+ (.78),2	0 (1.00),0	ts (1.00)
MK	ʃ˚(.94),1	0 (.62),1	+/v(.39),1	0 (.79),2	ts (1.00)
NO	˚3(.33),3	ə (.79),1	- (.95),1	0 (1.00),0	ts (.97),1
NR	ʃ˚(.83),2	ə (.88),1	- (.39),2	0 (1.00),0	ts (.97),1
WS	˚ʃ(.33),4	0 (.58),1	v/-(.39),1	0 (1.00),0	s (.90),2

The same statistics were determined for the various contexts of the variables 1, 2 and 6. They are omitted here for reasons of space.

The descriptive statistics presented above give a first indication of the various degrees of intraspeaker variability encountered in the material produced by the six speakers. They will make it possible to examine any changes in the realization of the variables with time. More specifically, we will be able to determine whether the modal realization changes or remains constant in both qualitative and quantitative terms. What is less satisfactory about the format used so far is the amount of information it contains about the non-modal realizations. It tells us how many realizations there are in addition to the mode and what their combined relative frequency is but it would be more interesting to know whether they are very similar to the mode in qualitative terms or very different. In other words, we would like to be able to develop a variation index which can capture the degree of similarity between the modal and non-modal realizations. The solution proposed here is one based on the use of a distance matrix as developed by Vieregge & Cucchiarini [4]. A weighted variation index Vw can be calculated by means of the following formula:

$$Vw = \sum_{i=1}^{N} Ri \,.\, di$$

Here, Ri stands for the relative frequency of the various non- modal realizations and di for the articulatory distance between a realization Ri and the Mode, calculated on the basis of the number of articulatory features in terms of which the two realizations differ. The value of the index is arrived at by summing the products of the relative frequency of each non-modal realization and its distance measure. It will be clear that the weighed variation index Vw represents a measure of the articulatory variation around the mode which is superior to the gross variation index obtained by summing the relative frequencies of the non-modal realizations because it takes account of the articulatory difference between the mode and the non-modal realizations.

7. CALCULATION OF THE VARIATION INDEX

We will illustrate the calculation of the variation index for one of our variables, <3>. Between them, the 6 speakers used 10 different realizations of this variable. The following matrix was used to calculate the differences:

	ʃ+	3+	3˚+	ʃ	3	3˚	z˚	z	ʃ−	ʃ˚ˇ
r1	ʃ+	1.0	0.5	0.5	1.5	1	2	2.5	1	0.5
r2	3+		0.5	1.5	0.5	1	2	1.5	2	0.5
r3	˚3+			1	1	0.5	1.5	3	1.5	0
r4	ʃ˚+				1	0.5	2.5	3	0.5	1
r5	3					0.5	2.5	2	1.5	1
r6	3˚						2	2.5	1	0.5
r7	z˚							0.5	3	2.5
r8	z								3.5	2
r9	ʃ−									1.5
r10	ʃ˚ˇ									

of the final project of a 120-hour course in phonetic transcription taught by the second author. They were instructed to produce a consensus transcription in accordance with the IPA conventions [1] which, they were told, would later be assessed by their teacher. The final version of the consensus transcription, which forms the basis of the present study, was produced by the two authors. After several tuning sessions, during which a number of minor notational problems were ironed out and maximum uniformity in transcriptional practice was achieved, the authors worked through the student-made transcriptions on an individual basis. However, apparent inconsistencies in the author versions were carefully re-examined to produce the ultimate consensus transcription used for this study.

4. COLLECTION OF MATERIALS

The speech samples were produced by 6 educated speakers of standard Dutch, all employed by the University of Nijmegen and living in the Nijmegen area, though originally hailing from various parts of the country. The amount of regional accent in their speech varied from mild to reasonably strong. There were three women and three men, their ages ranging from 25 to 50. The six speakers read three texts on each of three days, with a one- week interval. On each day, the three texts were read three times in succession at three points in time, i.e. at 9am, 1pm and 5pm, giving a total of 9 readings per speaker per day, and a grand total of 27 readings for each speaker for the three days. Although the texts were different, they were identical in terms of the variables under investigation, so that in effect 27 tokens of each instance of all variables are available for analysis. However, the preliminary results presented below are based on a subset of 6 non-contemporary readings from the total of 27 readings.

5. VARIABLES INVESTIGATED

Nine segmental variables were investigated. They were selected on the basis of their expected variability in Dutch. They are (n = the number of tokens per reading):

1. <r>, in four contexts, viz.:
   r1: C - n=3
   r2: - (C) # n=6
   r3: # - V n=3
   r4: V - V n=4
2. <x>, in two contexts, viz.:
   x1: - r n=2
   x2: # - V n=3
3. <z> n=5
4. <v> n=5
5. <ʒ> n=3
6. Svarabhakti, in two contexts, viz.:
   S1: l - n=3
   S2: r - n=1
7. Assimilation of voice before /b/ and /d/ n=3
8. Elision of /n/ after schwa n=4
9. <ts>, as in Dutch *politie* (English *police*) n=5

6. DESCRIPTIVE STATISTICS

In order to arrive at a first overall measure of the degree of intraspeaker variation, the following statistics were determined per variable and per speaker over the six non-contemporary readings:
1. the Mode, (M), i.e. the most common realization of the variable;
2. the relative frequency of the mode, (fM);
3. the number of realizations other than the mode, (p), any number of hapax legomena (i.e. unique realizations) being counted as 1.

They are expressed below in the format M (fM),p, or M1/M2 (fM),p, for a bimodal distribution. (The conventions used for the Voicing variable are + for voicing, - for devoicing and v for a media realization.)

Speaker	<r>	<x>	<z>	<v>
MJ	ɣ̊ (.57),6	x (.73),3	z̥ (.50),3	f/v (.33),2
JK	ʁ̊ (.36),6	ɣ (.57),4	z (.93),1	v (.77),2
MK	r (.59),7	χ (.90),1	s (.93),1	f (.97),1
NO	R (.25),10	x (.60),4	z (.57),5	f (.57),2
NR	R (.39),6	x (.87),2	s (.60),4	f (.63),3
WS	ɣ̊ (.35),9	x (.57),3	z̥ (.40),3	v (.53),3

For a full discussion of the principles underlying the distance matrix the reader is referred to Vieregge & Cucchiarini [4]. Suffice it to say here that measures used to calculate the distances are based on the articulatory difference between the sounds. Note that the value 0 is assigned to the distance measure between the realizations r3 and r10, the devoiced realization of a voiced fricative and the voiced realization of its voiceless counterpart.
Speaker MJ's raw variation score for this variable is (.39),3. There were 7 instances of the modal realization r3, 4 occurrences of r1, 4 of r2 and three hapax legomena, r5, r7 and r10. The variation index is then calculated as follows:

Vw = (.222 x .5) + (.222 x .5) + (.058 x 1.5) + (.058 x 1) + (.058 x 0) = .37

It is interesting to compare this index with the combined relative frequency of the variation around the mode, which was .61. Below, the variation index Vw is given for the remaining 5 speakers, followed by the raw variation index V.

	M	Vw	V
JK	ʒ̬(.78)	.26	.22
MK	ʃ̬(.94)	.03	.06
NO	ʒ̬(.33)	.39	.67
NR	ʃ̬(.83)	.17	.17
WS	ʃ̬(.33)	.51	.63

It appears that the weighed variation index Vw can deviate quite considerably from the raw variation index, especially if the mode has a low frequency of occurrence, as in the case of speakers NO and WS. While the relative frequency of the mode is the same for these speakers, NO's weighed variation index is considerably lower, which reflects the greater similarity to the mode of NO's non-modal realizations.

8. CONCLUSION

As observed in the introduction, the results presented above are based on a small portion of the available data. The emphasis here has been on some of the procedures used to describe intraspeaker variation in a systematic fashion. The consensus transcription is proposed as the most suitable format for the initial analysis of the speech samples collected. The use of a distance matrix based on articulatory differences between realizations affords a principled approach to a further, quantitative analysis of the variation encountered in the material. Major problems remain to be resolved before a meaningful comparison is possible of the readings produced at different times. It is this comparison which should provide answers to the central question of the consistency of intraspeaker variation patterns.

REFERENCES

[1] ROACH, P.J. (1989), "Report on the 1989 Kiel Convention", *Journal of the International Phonetic Association, 19*, 67-80.
[2] SHRIBERG, L.D. et al. (1984), "A Procedure for Phonetic Transcription by Consensus: A Research Note", *Journal of Speech and Hearing Research, 27*, 456-465.
[3] TING, A. et al. (1970), "Phonetic Transcription: A Study of Transcriber Variation", *Report*, Wisconsin Research and Development Center, Madison.
[4] VIEREGGE, W.H. & C. CUCCHIARINI (1988), "Evaluating the Transcription Process", *Proceedings of the 7th FASE Symposium Speech 88*, Edinburgh, 73-80.