Accelerometric Measurements during Speech for Automatic Speech Segmentation and for Professional Voice Education[1]

- A Brief Preliminary Account -

Introduction

During the last two decades, the analysis of jaw movement during speech and mastication has been a topic of research in several laboratories, for example (Cosi and Magno Caldegnetto), in Padova, Italy, (Attina et al., 2004), in France, (Ostry and Munhall, 1994), in Canada, (Fukaya and Byrd, 2005) in USA, etc.

We propose the measurement of the acceleration of the jaw during speech, as a method for speech processes analysis and as an auxiliary instrument for automatic segmentation of the speech. We are not aware of any use of such a method in automatic speech segmentation. Some researches on measurements of the acceleration of the jaw during speech, but using different set-ups, have been previously reported. Those researches had aims different to ours; namely, they were intended to study of speech and mastication mechanisms (e.g., [1-14]).

Beyond tongue and lips movement, the jaw movement is one of the most obvious mechanisms in speech production. Speech production involves vertical displacement of the jaw, in conjunction with the lips opening and closing, to help shaping the appropriate cavities and mouth openings during speech. In our opinion, jaw acceleration measurements should be included in any research on speech physiology and should complement any explanation of speech production. Consequently, we propose that such measurements are routinely included in all voice tests, medical voice databases, and in the training for voice rehabilitation and voice education.

We propose the use of jaw acceleration measurements with two purposes, namely for helping speech segmentation for speech annotation, and as a mean to produce feedback during the professional voice training. The use in segmentation is supported if the experimental findings demonstrate specific patterns of acceleration at the beginning and the end of syllables, or of specific phonemes, or during phoneme transitions. Our findings support such a conclusion. The inference of the use of acceleration measurements in speech education is less trivial. However, one deficiency of uneducated speakers is to concatenate words in a manner that makes speech difficult to understand. Moreover, stress is less clear for such speakers. Both word separation and good emphasis require a more dynamic movement of the jaw. Hence, providing acceleration data as a feedback during the professional voice education could, in principle, enhance the education process.

Method and errors

The measurements have been performed with an accelerometric module (1 axis) developed in our Group of Laboratories by Mr. Marius Hagan, under the author's supervision. The module is based on a microcontroller with a 12-bit AD/C (ADUC type – Analog Devices micro-controller). The precision of the accelerometer module is about 1 mg (the thousandth part of the gravitational acceleration, or 1 cm/s^2). This range is compatible with the expected movements of the jaw. The acceleration measuring module has been described in (Teodorescu et al., 2005, 2006) [15, [16] and will be detailed in another paper.

[1] First version posted on July 24, 2006. This is the third version, posted on August 4, 2006.

Indeed, a transition during speech, say from a plosive consonant to a vowel, requires the opening of the mouth in about 10 ms; the opening is of the order of one centimeter. From the movement equation, $s = at^2/2$, it results that the value of acceleration can be of the order of:

$$1\,cm = at^2/2 = a \cdot (10^{-2})^2/2 \;\Rightarrow 10^{-2} \cdot 10^4 = a \;\Rightarrow a = 10^2\, m/s^2 \approx 10\, g\,.$$

Thus, the jaw may suffer very large accelerations. A range of 1 mg to 10 *g* and a precision of measurements of 1 mg are needed to monitor the accelerations of the jaw. However, we found that the experimental data do not support such large accelerations. Our measurements produced ranges up to 100 mg, similar to the measurements in male spoken German, as reported by (Tillmann & Pfitzinger, 2003), that is up to 2 m/s^2, or 0.2 g. These findings show that the measured accelerations are somewhat incompatible with the duration of a phoneme, yet the movements should be correlated with the phoneme duration during normal speech production. We do not have yet an explanation for this incompatibility, and further research must be performed to find the source of errors either in the measurements, or in our concepts about speech production.

During measurements, the axis of the measurement was almost vertical, with an error of about 10^o due to the positioning constraints. The sampling rate is 100 samples per second. The data is filtered with an MA filter, averaging *n* consecutive bits, with *n* = 5 ..100.

The module, which is small enough (about 40 grams, 50 × 60 × 10 mm) has been fixed in a lateral position on the middle of the mandible, by means of an elastic belt (see Fig. 1). However, this method of attachment has the drawback of allowing the circuit to slightly move with respect to the jaw, because of the elastic belt. Also, vibrations of the accelerometer are possible. A method to firmly attach the accelerometer to the jaw without impeding on the movements has to be devised. Therefore, the preliminary results described here should be considered mere preliminary tests. Data is prone to a systematical error, because the movement of the jaw is not performed according to a single axis; consequently, any movement of the jaw that changes the plane of the jaw results in a change in acceleration in the vertical direction. Because of the same reason, the subject has to preserve the head perfectly vertical and with no tilting during speech (during measurements). This was found virtually impossible by the author. Tilting the head may produce a large acceleration, which is a serious artifact. Future tests will be made with the head of the speaker fixed by a belt to a predetermined position; however, this will be a constraint for the speaker and may induce changes in the natural movements of the jaw during speech.

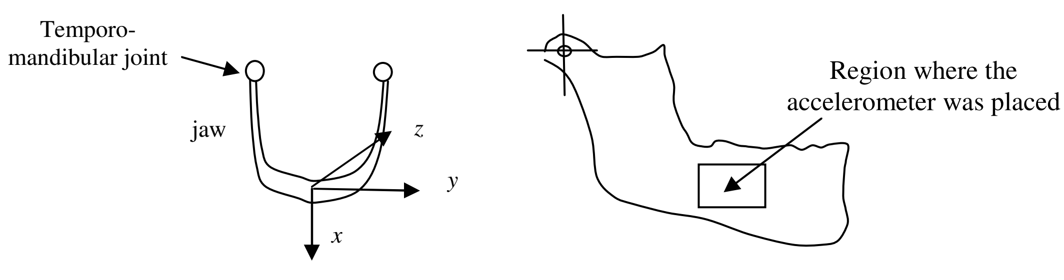

Fig. 1. Measuring coordinate system

The reason we preferred a direct acceleration measurement, instead of space measurement followed by a double differentiation, as proposed by several authors (e.g., [6]), is that differentiation is well known to be prone to noise. Any high frequency noise will be much increased by differentiation, making the result unusable. A similar drawback is encountered when the acceleration signal is used to produce, by double integration, the space (displacement) information; in that case, any offset in the acceleration signal will produce by integration large errors in the space value.

Until now, only a few experiments have been performed. Namely, a single subject (the author) has uttered repeatedly bi-syllabic words in Romanian (mama, "tata", i.e., papa in English), and a sentence of two words, with short, bi-syllabic words ("mama vine" – mother comes; vine mama?). All the words have the structure CV-CV. The utterance of two words (mama, tata) is produced with large jaw movements, because their syllables ("ma", "ta") start with plosives or nasals followed by a strong vowel ("a"). The third syllable in this set of words is "vi-", which starts with a semivowel ("v") and includes a vowel which is less energetic ("i"); thus, this syllable implies smaller movements of the jaw. The fourth syllable, "-ne", starts with a nasal and ends in an energetic vowel, implying strong movements.

Preliminary results

The results are preliminary and the technique used to attach the accelerometer to the jaw has been found unsatisfactory and should be modified in further experiments. The range of the acceleration has been found to be from a few mg to 0.15 g, which is much lower than expected, according to the computations presented above. An example of processed data, for several pronunciations of "mama comes" ("vine mama", in Romanian) is shown in Figure 2. The generated waveforms do not show any obvious regularity and the best similarity between any two curves is low. Similar conclusions can be derived from Figures 3 and 4, which show the acceleration unfiltered and filtered curves for a sequence if uttered words "tata"-"mama".

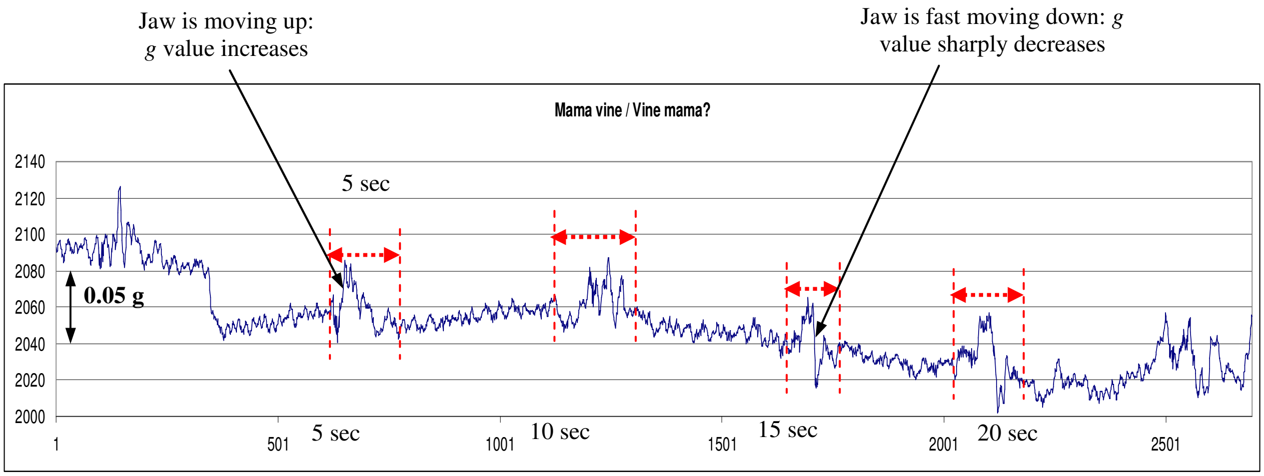

Fig. 2. Acceleration signal in the vertical direction, as measured on the middle lateral section of the jaw, during the repeated utterance of short sentences consisting in two words. Pronunciation segments are marked by arrows

The graphs in Figs. 2-4 should be interpreted taking into account that an increase in the acceleration value corresponds to a movement of the jaw in the upward direction; in that case, to the gravitational acceleration measured by the accelerometer, the inertial acceleration is added. The converse is true for descending jaw movements. From Figs. 2-4, the following information can be immediately derived:

- Approximate duration of the word
- The starting moment of the utterance (mouth opening moment)
- end of jaw movement (which is generally about 0.1 – 0.3 s delayed with respect to the end of the utterance)
- The up (higher *g* value) and down (lower *g* values) movements (see Fig. 2)

- The energetic movements of the jaw; these movements correspond to the beginning of strong vowels, or to movements of energetic lips separation, like in the nasals followed by vowels (“ma”), “ne”), or lips closing when nasals, especially “m”, end a vowel, like in “am”. Notice that “m” requires a closing or opening of the lips, associated with a strong jaw movement, while “n” can be produced with no lips closing, but only with the help of the tongue.
- Strong, vibrating “r” can be distinguished from non-vibrating “r” by the vibration induced to the jaw during uttering of the vibrating “r”.

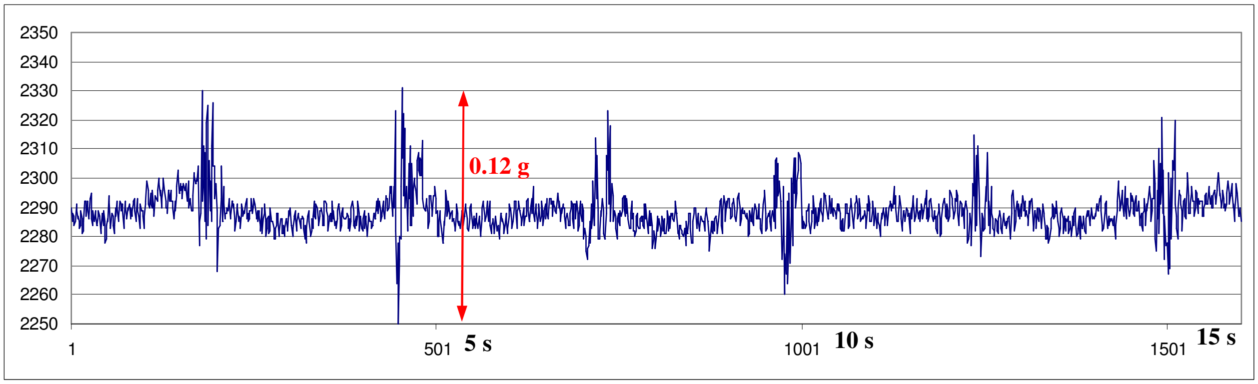

Fig. 3. Acceleration during repeated uttering of the words “tata” and “mama”. Unfiltered signal.

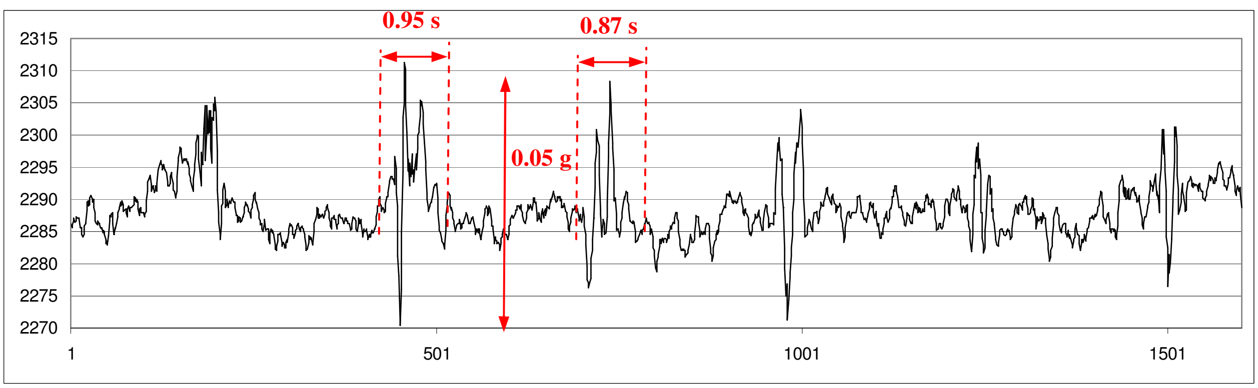

Fig. 4. Acceleration during repeated uttering of the words “tata” and “mama”. Low pass filtered signal.

Low-pass filtering has to be made with care for the errors it can produce. Notice that the MA-type, low-pass filtering, which is either a discrete counterpart of the integration, or a weighted discrete summation, reduces the amplitude of the shorter parts of the waveforms. The reduction, as exemplified by Figures 3 and 4, is more than twice. The peak acceleration in Fig. 3 is about 0.12 *g* (1.2 m/s^2), while it the peak acceleration in Fig. 4 is only 0.05 g.

A direct comparison of several utterances of “vine mama” shows a low resemblance between the waveforms. In Fig. 5, the first three utterances of “vine mama” from Fig. 2 are plot together, with 100 samples each.

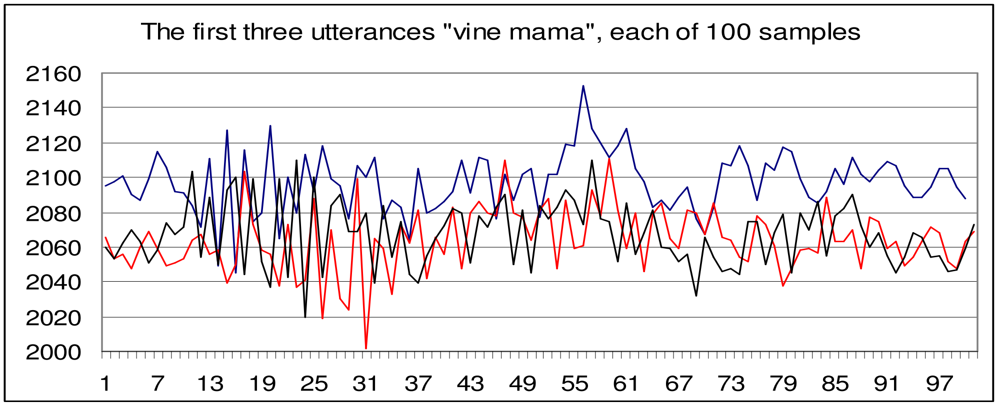

Fig. 5. Comparison of the acceleration signal produced during three utterances of the sentence "vine mama"

The low reproducibility of the waveforms is easy to notice from the amplitude-time plots (see Fig. 2 and Fig. 5), and is further emphasized by the interrelation plots produced as $x(N_1+n) = f\left(x(N_2+n)\right)$ where N_1 is the starting point for one waveform and N_2 is the starting point for the second waveform. Such a graph, for the first two "vine mama" waveforms in Fig. 2 and for $n = 0..50$ is shown in Fig. 6.

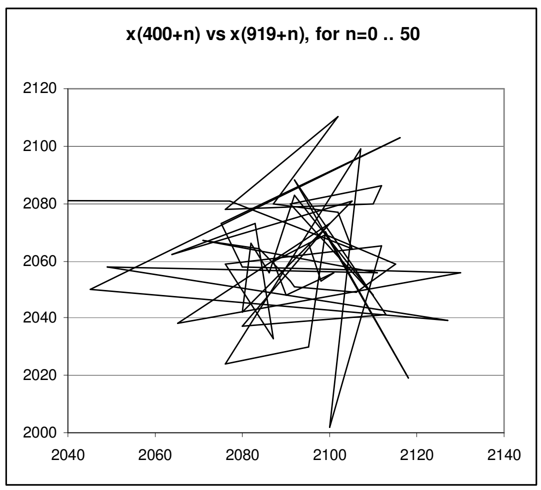

Fig. 6. Plot $x(N_1+n) = f\left(x(N_2+n)\right)$ for the first two "vine mama" utterances, with N_1 = 400, N_2 = 919 and $n = 0..50$

While the results are preliminary and raw, they demonstrate that the use of the jaw acceleration measurement may help speech segmentation. Moreover, they demonstrate the complexity of the jaw movement during the phonation mechanisms.

A measuring protocol proposal

At our best knowledge, no acceleration measuring protocol has been consistently established, based on experimental findings, during previous researches. One positive results of the research has been the establishment of a protocol for jaw acceleration measurements during mastication and speech. The protocol outline is:

1. Position of the subject: sitting, dorsal spine in vertical position, head in a vertical position (no tilt, no inclination, no lateral rotation); head fixed by a belt running around the front, to reduce head movements.
2. Accelerometer positioning – measuring axes: *Oy* axis is horizontal (the intersection of the frontal plane with the transversal plane); *Oz* is the sagittal axis; *Ox* is the vertical axis (intersection of the sagittal plane with the frontal plane), pointed downwards.
3. Acceleration range and sensitivity: 10 mg to 10 *g*.
4. Data transmission: preferably wireless, to eliminate any harness or wire attachment to the acceleration module, thus reducing errors.

The measured acceleration values are of the order of magnitude of 5 to 200 mg. On the other hand, the raw estimations of the average accelerations of the jaw are consistent with the value predicted by mathematical modeling by (Mallett et al.) for frogs during prey capturing. In their quoted paper, these authors determined, for the movement of the jaw of several species of frogs and toads, a displacement of about 0.5 cm in about 0.1 sec. (see Fig. 9 in the quoted paper, vertical movement at the beginning of the movement). That movement amounts in an acceleration of about 2.5 m/s^2, i.e. 0.25 *g*. The acceleration values determined by us are also consistent with other papers in the literature, for example, computations based on the graphs in the paper (Ostry and Munhall) produces similar results.

Further work

The method of measuring the accelerations of the jaw during speech, as used in the reported tests, has to be refined. Future research needs to investigate syllables with the phonological structures VC, CVC, VCC, diphthongs, etc., and phrases with various melodic trajectories (trajectories of the pitch, F0.)

Conclusions

In this preliminary paper, we have proposed to new directions of application of the jaw acceleration measurements during speech, namely, the use in automatic speech segmentation and in speech education and rehabilitation. Only the first application has been analyzed in some detail.

Whatever raw are the data collected until now, they show good promises in segmentation of the syllables whenever a CV-CV structure occurs in speech. Also, the data show that plosives and especially nasals (mainly “m”) are well emphasized by the larger jaw accelerations. Similarly, plosives that can not be produced by simply parting the lips, but need the tongue movement toward the palate, require the jaw movement and therefore produce larger accelerations. The vowel “e” after a plosive elicits larger jaw movements. Moreover, fricatives “sifflants” like “f”, “v”, “s” followed by “E” (but not sh-E) produce significant jaw movements that can help segmentation.

The measurement of the accelerations during speech may not only bring valuable supplementary information for automatic speech segmentation, but it may add extra information on the physiology of speech, including clarification of the mechanisms of jaw movement correlation with lips parting and closing and jaw-tongue movements. More research is needed to clarify the potential of the method.

Acknowledgments. Mr. Marius Hagan, MS, has built the acceleration module based on the recommendation and the technical specification provided by the author. Mr. Hagan also assisted the author with the data acquisition. The data is collected on the author. The tests have been performed with the partial support of a research grant from the CEEX projects, “Program VIASAN; Project: Sistem automat de diagnostic paraclinic in sindromul disfunctional al sistemului stomatognat”). The acceleration measurement module has been developed with the partial aid of the research grant offered by ROSA, Program AEROSPATIAL, Proiect: Sistem de masura accelerometric avansat pentru investigari atmosferice cu sonde planetare. Mrs. Diana Trandabat, Ph.D. student, and Miss Monica Feraru, MS, Ph.D. student have made several typo corrections; Mr. M. Hagan, MS, Ph.D. student has made a valuable critique.

References

[1]. D.M. Shiller, D. J. Ostry, P.L. Gribble, and R. Laboissiere, Compensation for the Effects of Head Acceleration on Jaw Movement in Speech. The Journal of Neuroscience, August 15, 2001, 21(16):6447–6456 (Also, at http://psyc.queensu.ca/~munhallk/02_shiller2001.pdf. Accessed July 28, 2006)

[2]. D. M. Shiller, R. Laboissiere, and D.J. Ostry, Relationship between Jaw Stiffness and Kinematic Variability in Speech. J. Neurophysiol. 88: 2329–2340, 2002

[3]. Abbs JH, Netsell R., An Interpretation of Jaw Acceleration During Speech as a Muscle Forcing Function. J. Speech Hear Res., 1973 Sept., 16 (3): 421-5

[4]. B.L. Pence, J. Munger, and S.L. Thomson, Measurements of the Frequency Response of Skin on the Head and Neck during Speech. ICVPB 2006, Tokyo. The 5th International Conference on Voice Physiology and Biomechanics. Variations across Cultures and Species. July 12-14 2006, Tokyo, Japan. http://www.vocology.jp/icvpb2006/Proceedings/AC4.pdf. (Accessed July 28, 2006)

[5]. D.J. Ostry and K.G. Munhall, Control of Jaw Orientation and Position in Mastication and Speech. Neurophysiology, Vol. 71, No. 4, April 1994. Also http://www.psych.mcgill.ca/labs/mcl/1528.pdf, (Accessed July 28, 2006)

[6]. P. Cosi and E. Magno Caldognetto, Lips and Jaw Movements for Vowels and Consonants: Spatio-Temporal Characteristics and Bimodal Recognition Applications, http://www.pd.istc.cnr.it/Papers/PieroCosi/cp-NATO95.pdf. (Accessed July 28, 2006)

[7]. Maureen Stone and Sandra L. Hamlet, Tongue and Jaw Gestures during the Production of /d/ and Flap in Nonsense Phrases. The Journal of the Acoustical Society of America, November 1981, Volume 70, Issue S1, p. S13

[8]. Teruhiko Fukaya, Dani Byrd, An Articulatory Examination of Word-final Flapping at Phrase Edges and Interiors. Journal of the International Phonetic Association (2005) 35/1

[9]. J. H. Abbs, R. Netsell, An Interpretation of Continuous Articulator Acceleration as a Muscle Forcing Function. The Journal of the Acoustical Society of America, July 1971, Volume 50, Issue 1A, p. 139

[10]. Lakshmi Venkatesh, Christopher A. Moore, Kevin J. Reilly, Kathryn P. Connaghan, Roger S. Steeve, Timing of Oronasal Coupling in typically Developing Children and Children with Speech Delay. http://depts.washington.edu/spchphys/pres/ASHA2003_lv.pdf (Accessed July 28, 2006)

[11]. Eric S. Mallett, Gary T. Yamaguchi, James M. Birch, and Kiisa C. Nishikawa, Feeding Motor Patterns in Anurans: Insights from Biomechanical Modeling. AMER. ZOOL., 41:1364–1374 (2001)

[12]. T.S. Miles, S. C. Flavel, and M. A. Nordstrom, Control of Human Mandibular Posture during Locomotion, J. Physiol., January 1, 2004; 554(1): 216 - 226.

[13]. Hans G. Tillmann and Hartmut R. Pfitzinger, Local Speech Rate: Relationships between Articulation and Speech Acoustics. 15th ICPhS Barcelona, 2003, University of Barcelona, Barcelona, Spain, pp. 3177-3180. ISBN 1-876346-48-5

[14]. Philip Hoole, Articulatory Discriminability of Vowels: Articulator and Corpus Effects. 137th ASA meeting. Berlin March 14-19, 1999. Session 5pSCb5. : Friday Afternoon, March 19. Presentation at http://www.phonetik.uni-muenchen.de/~hoole/pdf/berlin0.pdf. (Accessed July 20, 2006)

[15]. H.N. Teodorescu and Hannes Griebel (Eds.), Mars and Planetary Science and Technology, Ed. Performantica, Iasi, Romania, Sept. 2004. ISBN 973-7994-83-3

[16]. H.N. Teodorescu, M. Hagan, et al., 2005, Research Report, Grant ROSA, Iasi 2005

Notice. This summary has been prepared for presentation in a local seminar and for possible submission in a conference.