RESEARCH ARTICLE OPEN ACCESS

Rule-based Prosody Calculation for Marathi Text-to-Speech Synthesis

Sangramsing N. Kayte[1], Monica Mundada[1], Dr. Charansing N. Kayte[2], Dr. Bharti Gawali*
[1,*]Department of Computer Science and Information Technology Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
[2]Department of Digital and Cyber Forensic, Aurangabad, Maharashtra

ABSTRACT
This research paper presents two empirical studies that examine the influence of different linguistic aspects on prosody in Marathi. First, we analyzed a Marathi corpus with respect to the effect of syntax and information status on prosody. Second, we conducted a listening test which investigated the prosodic realisation of constituents in the Marathi depending on their information status. The results were used to improve the prosody prediction in the Marathi text-to-speech synthesis system MARY.
Keywords – Prosody, Marathi MARY, MULI corpus, GToBI.

I. INTRODUCTION

The prediction of appropriate prosody is a crucial task for the synthesis of speech. Generating inadequate prosody seriously hampers intelligibility and naturalness [1]. To a certain extent, the problem can be avoided when using corpus-based synthesis, by selecting units from the appropriate parts of a sentence and thus indirectly generating the correct prosody as recorded in the corpus. However, more recent attempts to generate expressive speech including emphasis or focus require the explicit modelling of prosody [2-4]. As a basis for modelling expressive speech, it is thus necessary to be able to predict unexpressive speech from linguistic features.

The problem of prosody prediction is by no means solved. Too little is known about the multitude of factors and their interactions that influence the prosodic realization of a sentence. Factors reported to be relevant include part of speech, position of the word in the sentence, sentence type, various aspects of syntactic structure, and information structure. This list clearly is not exhaustive. Given this large number of potentially relevant variables, a statistically based investigation would have been attractive, not least because it could have provided us with an estimate of the relative importance of the various factors. However, we could not follow the statistical approach because our Marathi MARY text-to-speech (TTS) system [5] [6]. For that reason, we pursue a rule-based approach, which allows for a very controlled prediction of prosody, and which has the advantage that findings can be interpreted which is often not the case in statistically trained prediction systems.

The paper is structured as follows. The first section formulates a number of concrete assumptions regarding the links between a variety of linguistic factors and prosody, based on the existing literature. The second section describes the analysis of the Marathi corpus MULI, testing some of the assumptions made in the first section, notably regarding the effects of part of speech, syntax and information status on prosodic realization.

The third section presents a listening test which investigates in more depth the effect of a constituent's information status on its preferred prosodic realization.

II. CORPUS ANALYSIS

In order to verify the validity of the assumptions derived from the literature in the previous section, we carried out an analysis of the MULI corpus.

A. The Corpus

We analyzed the corpus elicited in the MULI (Multilingual Information structure) [7] project, which examined the means with which information structure is realized in Hindi and Marathi. The Marathi part of the corpus contains 1000 sentences stemming from the economics section of the Marathi newspaper Frankfurter Rundschau. The text was spoken by one speaker. As the material is also part of the TIGER Treebank [8], the corpus already contained detailed syntactic information. Some special syntactic information was added, mainly word order information like fronting or extra position. Prosodic annotation followed the GToBI conventions. The annotation of information status is based on the taxonomy of [9], which distinguishes the statuses "brand new" and "unused", representing new information, "evoked", representing given information and "inferable". In the case of inferable

information, the type of bridging relation between anaphor and antecedent was also annotated. Additionally, information about lexical relations between anaphor and antecedent was added. Even though the corpus must be considered very small for our purposes, it appears to be the only Marathi corpus available for which both GToBI and information structure are annotated [13][20-27].

B. Method

We tested the various assumptions that arose from the literature survey as summarized in the previous section, using the MMAX framework. For each assumption, we carried out frequency counts of the values of the prospective linguistic predictor variables and the predicted prosodic variables [14].

C. Results and discussion

Part of speech was confirmed as an important predictor for accentuation. Content words frequently carry an accent (71%), and function words are mostly not accented (10%). Proper nouns (85%), adjectives (80%), nouns (79%) and numbers (90%) show the highest accentuation rates.

The surface position of a word has an effect on the type of accent realised on it. In prenuclear position, rising accents (L+H*) are frequent (55%). In nuclear position, falling (H+L*) (65%) and low accents (L*) (38%) were realised more frequently. The H* accent appears in both prenuclear (53%) and nuclear (31%) position.

The assumption that finite verbs in verb second position are never accented could not be confirmed in the corpus. The probability for finite verbs in this position to be accented (47%) was only marginally lower than the general probability for finite verbs to be accented (37%).

Following the hypothesis that the rightmost element within a phrase is accented, we investigated the prosodic realisation of the rightmost element in chunk phrases. In fact, the rightmost element carries an accent very frequently (85%), but the part of speech of a word has more influence on its accentuation: the content words in chunk phrases that are not the rightmost ones, also carry an accent frequently (83%).

The assumption that the most deeply embedded verb within a verbal sequence always carries an accent could not be confirmed. The probability for embedded infinitives and participles of full verbs to carry an accent (76%) is approximately the same as the general probability for infinitives and participles of full verbs to be accented (71%). The MULI corpus does not contain any auxiliary verbs appearing in embedded position.

As objects carry accents with roughly the same frequency as subjects (objects: 86%; subjects: 88%), the tendency for subjects to be accented was not confirmed.

An interaction between the Marathi and prosodic phrasing could be observed. In about 53% of the cases in which contains three words, a prosodic boundary is realised. Furthermore, an increasing length of the accompanied by an increasing likelihood for the realisation of a boundary after the similar observation was made for chunk phrases: with the increase in the length of a chunk phrase, the probability that the chunk is followed by a prosodic boundary also becomes higher. If the chunk phrase contains more than four words, the realisation of a boundary is more probable (55%) than the absence of a boundary [15].

Regarding information status, we observed that nouns representing new information are frequently accented (brand new: 91%, unused: 93%), but the same holds for inferable (89%) and evoked information (91%). Thus the assumed influence of information status on the prosodic realisation of nouns could not be confirmed in the corpus. Note, however, that personal pronouns, which always represent evoked information, are only accented in 11% of the cases.

The effect of lexical and bridging relations was difficult to interpret, because the number of occurrences in the MULI corpus was small and the number of possible relations large. All relations show a similar distribution across accent types, but more data would be needed to consolidate this observation.
The assumption that given information is deaccented if anaphor and antecedent share the same grammatical role could not be confirmed. This type of given information was accented in 92% of the cases [10][11].

When examining the number of prosodic boundaries following new (38%) vs. given (46%) or inferable (39%) information in the corpus, the assumption that new information is more often followed by a boundary could not be confirmed.
The examination of contrastive constituents revealed that they are always accented, most frequently with an L+H* (46%) or an H* (49%) accent. Thus the L+H* accent seems to be an appropriate accent for expressing contrast.

III. LISTENING TEST

The apparent conflict between the findings of our corpus analysis and the literature prompted us to gather complementary information regarding the role of information status for the accentuation of constituents, which we investigated by means of a listening test.

The experiment by [7], which confirmed the accentuation of given information in Marathi, tested only constituents in sentence final position. However, in the MULI corpus, the major part of given and

inferable constituents appeared at the beginning or in the middle of the sentence and were frequently accented. For Hindi, it was found that grammatical subjects at the beginning of a sentence are more often accented than constituents with other grammatical functions, independently of their information status [9]. Still, the same authors found less accents for given subjects when the grammatical role and surface position of the antecedent was the same.

We therefore designed a listening test in order to investigate the preferred prosodic realisation of grammatical subjects at the beginning of a sentence, more specifically in the Marathi depending on their information status and on the grammatical role and position of the antecedent of given information.

For information status, we distinguished only new vs. given information. The given constituents were always coreferent with their antecedent. Each test stimulus consisted of two sentences, so that the given constituents always referred to an antecedent appearing in the immediately preceding sentence. We formulated sentences from three different text genres (news, literary style, familiar context) to test whether the genre has an effect on the prosodic preferences.
Our hypothesis is as follows. New information carries an accent and is possibly followed by an intermediate boundary; given information whose antecedent does not have the same grammatical role and position within the sentence is also accented; and given information whose antecedent has the same grammatical role and surface position is deaccented[16-17].

A. Stimuli

In each of the three genres, we designed three target sentences with a consisting of a two-word noun phrase. Using the di-phone synthesis system MARY [6], we generated three prosodic versions of each target sentence: an accented version with a following intermediate boundary, an accented version without boundary, and a deaccented version. As previous findings consistently suggested the L+H* accent as appropriate in prenuclear position, we used this accent type for the accented versions. For each target sentence, we created three context sentences.

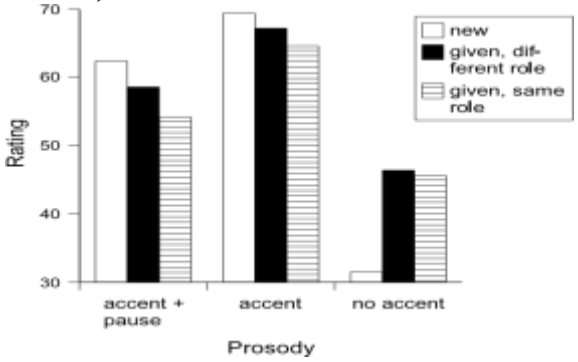

Figure 1: Relation between information status and prosody

Two context sentences contained the same information as the constituent of the target sentence so that the constituent in the target sentence refers to given information. In one version, anaphor and antecedent had the same grammatical role and surface position, in the other version, they had different roles and positions. In the third version, the constituent of the target sentence was not already mentioned in the context sentence and thus was new. The target sentence needed to be minimally adapted to be plausible as a follow-up to the different context sentences.

B. Method

We presented the 30 sentence pairs (2 genres x 2 target sentences x 2 information statuses) in written and in auditory form using the tool 'Rating Test'. For every sentence pair presented in written form on the computer screen, three auditory versions were presented via headphones. Participants were allowed to listen to the versions as often as they wanted to. Just as in training – with two practice trials – they were asked to judge the appropriateness of the sound of the second sentence, especially with respect to the context, i.e. to the content of the first sentence. They were instructed to make their judgments independent of the segmental quality of the speech synthesis. 20 native speakers of Marathi took part in the experiment [10][12][16-17].

C. Results and Discussion

We used SPSS to conduct several analyses of variance.

The analyses showed that the accented version without boundary was always judged to be most appropriate, closely followed by the version with a boundary. This effect was independent of the information status of the constituent in the (see Figure 1). Insofar, the strong formulation of our hypothesis cannot be confirmed.

Nevertheless, a weaker effect in the hypothesized direction was found. It can be seen from Figure 1 that the deaccented version was considered clearly unacceptable for new information while showing medium acceptability for the two types of given information. Beside this, the accented version with boundary, which is the most marked one, received the highest ratings if realised in case of new information. This interaction between information status and prosody is highly significant. There was no significant interaction between text genre and any other factors [16-17][20-27].

IV. CONCLUSION

We investigated the interaction between different linguistic factors and prosody in Marathi. By analyzing a spoken Marathi corpus, we could show that some of the assumptions made in the literature,

mainly for Hindi, can also be confirmed for Marathi, but a considerable amount of the assumptions could not. In particular, we could not find a relation between information status and prosody. As the existence of such a relation was experimentally confirmed for sentence final constituents in Marathi, we investigated the preferred prosodic realisation of constituents in sentence initial position, depending on their information status. We found that the accentuation of these constituents is always preferred in Marathi, both for new and for given information, but that the accentuation of given information is also acceptable to some degree. By implementing our findings in the TTS system MARY, we obtained better speech synthesis results.

As the analyzed corpus is not very large and is spoken by one speaker only, our results cannot claim to be representative of Marathi prosody in general. Further investigations of spoken language with respect to the factors that influence the prosodic realisation in Marathi are needed. New findings in the field could be used to improve the prosody prediction in TTS systems.

References

[1] Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014.

[2] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" International Journal of Computer Science and Information Technologies (IJCSIT)

[3] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis System for Hindi" International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015

[4] W. Hamza, R. Bakis, E. Eide, M. Picheny, and J. Pitrelli. The IBM Expressive Speech Synthesis System. In Proc. of the 8th International Conference on Spoken Language Processing, Jeju, Korea, 2004.

[5] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis Systems for Marathi Language" OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 5, Ver. I (Sep –Oct. 2015), PP 76-81e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[6] M. Schr¨oder and J. Trouvain. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. International Journal of Speech Technology, 6:365–377, 2003.

[7] S. Baumann, C. Brinckmann, S. Hansen-Schirra, G. Kruijff,I. Kruijff-Korbayov´a, S. Neumann, and E. Teich.Multi-Dimensional Annotation of Linguistic Corpora for Investigating Information Structure. In Proc. of Frontiers in Corpus Annotation Workshop at HLT-NAACL 2004.

[8] S. Brants, S. Dipper, S. Hansen, W. Lezius, and G. Smith.The TIGER Treebank. In Proc. of the Workshop on Treebanks and Linguistic Theories, pages 24–41, Sozopol,Bulgaria, 2002.

[9] E. Prince. Towards a Taxonomy of Given-New Information.In P. Cole, editor, Radical Pragmatics, pages 223–255. Academic Press, 1981.

[10] Sangramsing Kayte, Monica Mundada, Santosh Gaikwad, Bharti Gawali "Performance Evaluation Of Speech Synthesis Techniques For English Language " International Congress on Information and Communication Technology 9-10 October, 2015

[11] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte " Performance Calculation of Speech Synthesis Methods for Hindi language IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov - Dec. 2015), PP 13-19e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[12] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis System for Hindi" International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015

[13] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Implementation of Marathi Language Speech Databases for Large Dictionary" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 40-45e-ISSN: 2319 –4200, p-ISSN No. : 2319 –4197

[14] Sangramsing Kayte, Dr. Bharti Gawali "Marathi Speech Synthesis: A review" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 3708 – 3711

[15] Sangramsing Kayte, Monica Mundada "Study of Marathi Phones for Synthesis of Marathi Speech from Text" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-10) October 2015

[16] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte" Speech Synthesis System for Marathi Accent using FESTVOX" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.6, November2015

[17] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte "Screen Readers for Linux and Windows – Concatenation Methods and

Unit Selection based Marathi Text to Speech System" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.14, November 2015

[18] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte " Performance Evaluation of Speech Synthesis Techniques for Marathi Language " International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[19] Sangramsing Kayte, Monica Mundada, Jayesh Gujrathi, " Hidden Markov Model based Speech Synthesis: A Review" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[20] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte" Speech Synthesis System for Marathi Accent using FESTVOX" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.6, November2015

[21] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte "Screen Readers for Linux and Windows – Concatenation Methods and Unit Selection based Marathi Text to Speech System" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.14, November 2015

[22] Sangramsing Kayte, Monica Mundada,Dr. Charansing Kayte " Performance Evaluation of Speech Synthesis Techniques for Marathi Language " International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[23] Sangramsing Kayte, Monica Mundada, Jayesh Gujrathi, " Hidden Markov Model based Speech Synthesis: A Review" International Journal of Computer Applications (0975 – 8887) Volume 130 – No.3, November 2015

[24] Sangramsing N. Kayte ,Monica Mundada,Dr. Charansing N. Kayte, Dr.Bharti Gawali "Approach To Build A Marathi Text-To-Speech System Using Concatenative Synthesis Method With The Syllable" Sangramsing Kayte et al.Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part-4) November 2015, pp.93-97

[25] Sangramsing N. Kayte, Dr. Charansing N. Kayte,Dr.Bharti Gawali* "Grapheme-To-Phoneme Tools for the Marathi Speech Synthesis" Sangramsing Kayte et al.Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part -4) November 2015, pp.86-92

[26] Sangramsing Kayte "Duration for Classification and Regression Tree for Marathi Text-to-Speech Synthesis System" Sangramsing Kayte Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part-4)November2015

[27] Sangramsing Kayte "Transformation of feelings using pitch parameter for Marathi speech" Sangramsing Kayte Int. Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 5, Issue 11, (Part -4) November 2015, pp.120-124