Machine learning for detection of anticipation nucleus

Július Zimmermann, LICOLAB, Pavol Jozef Šafárik University in Košice
Eva Kiktová, LICOLAB, Pavol Jozef Šafárik University in Košice

Anticipation is highly beneficial for an interpreter: it enables prediction of what the speaker has in his mind and how his speech will continue. The detection of anticipation nuclei is a challenging task for interpreters. Their ability to detect them is strongly dependent on their personal skills and experience. This prediction – anticipation – is based on semantics, syntax, and intonation. In our study, a machine learning approach based on statistical learning methods is applied for the detection of anticipation nuclei via prosody. Perceptually identified anticipation nuclei from French speeches were used to train the statistical models. The models created were then evaluated in the testing process. For the efficient use of the database, cross-validation was performed. The results obtained suggest that machine learning algorithms can be effectively used to detect a very fine suprasegmental phenomenon – the anticipation.

Keywords: *interpreting, anticipation, prosody, machine learning, detection, k-NN*

1. Introduction

A common interaction between people is realized in the verbal and nonverbal planes. The information content of these planes is naturally fused and provides complex information. Sometimes an entire sentence cannot be told; nonetheless, we can capture all of the information. This important human capability, the ability to anticipate future content, is routinely used in everyday conversations. This ability significantly improves the process of interpreting (Seeber 2001; Butz et al. 2007). During interpreting, an interpreter analyses and fuses several information planes of speech – prosody, semantics, syntax of language, etc. These planes work together to help the interpreter decode the final utterance (Kleiber & Sock 2006; Solé et al. 2007).

In this study, a specifically modified speech at the suprasegmental level of time, frequency and intensity will be investigated in order to detect a specific part of an utterance called an anticipation nucleus. The presence of this anticipation nucleus allows listeners to predict how the rest of the utterance will continue.

Anticipation plays an important role, not only for interpreting, but also for human-machine interaction (Koppula & Saxena 2016; Matos 2008; Ondáš & Pleva 2019). Anticipation allows the robot (system) to predict what the user will say, thus eliminating the need for verification when the prediction holds. It also allows the robot to take the initiative to propose the predicted next event. In this way, anticipation ensures a more natural and fluid communication with the machine/robot (Dominey et al. 2008; Rious et al. 2008).

The ability to anticipate remains a human privilege, but the development of artificial intelligence allows the application of different algorithms to learn how anticipation, theme detection, sentiment detection, etc., can be realized. In our work, machine learning algorithms are applied to learn the prosody information for anticipation nucleus detection.

In general, the presence of the anticipation nucleus (focus) is not a logical value, but is rather expressed by a grade of anticipation. This corresponds to the reality that not every anticipation nucleus is confirmed clearly by the perception tests performed by several independent interpreters (Paľová & Kiktová 2019). An anticipation moment detected only once has a low grade of anticipation – in other words, it has a low probability of anticipation compared to an anticipation moment that was confirmed several times through perception tests.

2. Theoretical background

Generally, the anticipation nucleus (focus) is a specifically modified sonantic nucleus of a syllable placed in front of a pause (a pause before a next utterance) (Paľová & Kiktová 2019). Sonantic nuclei can be modified by changes in fundamental tone and intensity. Our previous experiments, performed at LICOLAB (Language, Information and Communication Laboratory) at Pavol Jozef Šafárik University in Košice, showed that in French, tone modulation is the factor most relevant to anticipation, followed by intensity change.

The tone modulation of the anticipation nucleus is achieved by the short time increase of fundamental tone between the penultimate and the ultimate syllable. This is illustrated in Figure 1. The second way the anticipation nucleus can be formed is on the ultimate syllable only. It is the case of diphthong or hiatus (Figure 2). Examples of phonetic transcriptions, oscillograms, and extracted f0 values for the two aforementioned realizations of anticipation nuclei are depicted in Figures 1 and 2.

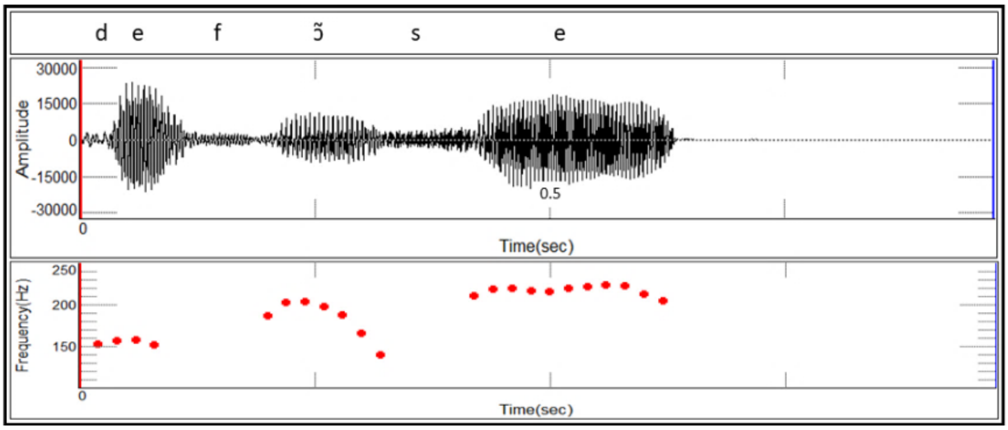

Figure 1: The anticipation nucleus – two syllables

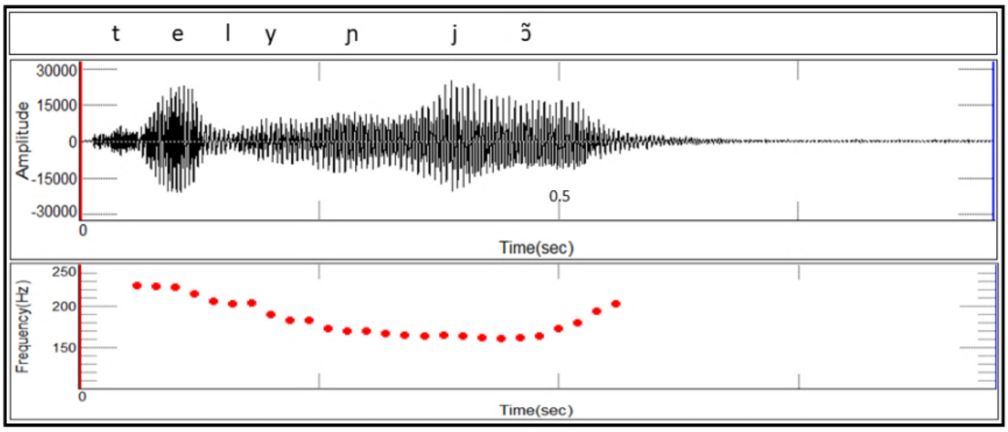

Figure 2: The anticipation nucleus – one syllable

3. Feature extraction

Used acoustic data included speeches of European Parliament members in French (74 speakers, both men and women). The representative sample of 100 sentences was randomly selected. Acoustic data in mono mode with 44.1 kHz sampling frequency and 16-bit resolution were analysed in the Multi-Speech software, in which intensity and fundamental frequency were extracted from segments divided into two groups: the first with anticipation and the second without anticipation (i.e. background). The duration of the analysed acoustic segments belonging to anticipation nucleus was from 0.5 s to 1.3 seconds. The principal block scheme of feature extraction is depicted in Figure 3.

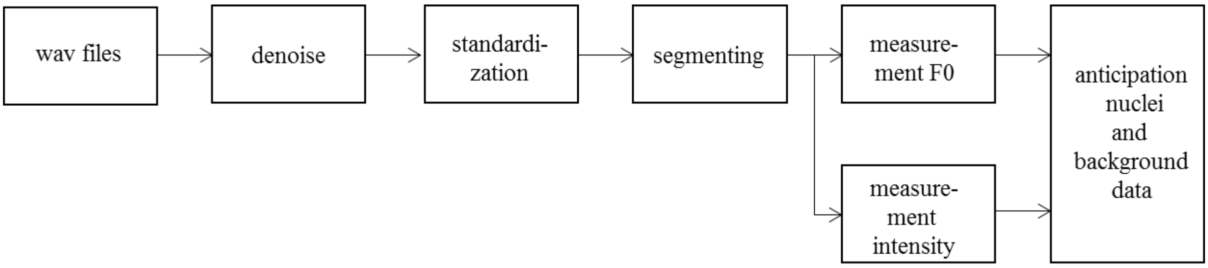

Figure 3: Feature extraction block scheme

In the preprocessing phase, the noise was removed from the acoustic signal by Multi-Speech software routine and the signal amplitude was normalized to an effective value of -17 dB. From such prepared signals, ultimately, 60 segments with anticipation and 60 segments with background sounds (from speeches by 14 women and 46 men) were identified. After the described processing, the intensity values in dB and fundamental frequency in Hz were extracted from 25 ms frames with 20 ms frame shift. A parametric representation of sounds with an anticipation nucleus and sounds without an anticipation nucleus were created using this method in Multi-Speech software.

4. Description of experiment

The feature vectors of extracted intensity contour and fundamental frequency were extended with the response values – the tag for anticipation (ant) or the tag for background (bcg) – according to the results of perceptions. Such prepared data in matrix form were processed in MATLAB 2019a software using the Statistics and Machine Learning Toolbox (MathWorks 2019). It can be used to describe, analyse, and model data; next, it provides different techniques for feature reduction; and also support various techniques for supervised and unsupervised machine learning. In our experiments, different supervised learning techniques were applied to a relatively small amount of data. For this reason, a five-fold cross-validation for training and testing, in which data were randomly divided into five equal parts, was performed. One part was always used for testing and the remaining parts were used in the training process. In this way, it is possible to train and evaluate the classifier effectively even on a relatively small sample of data. The final result represents the average performance from five partial recognition scores obtained from each fold.

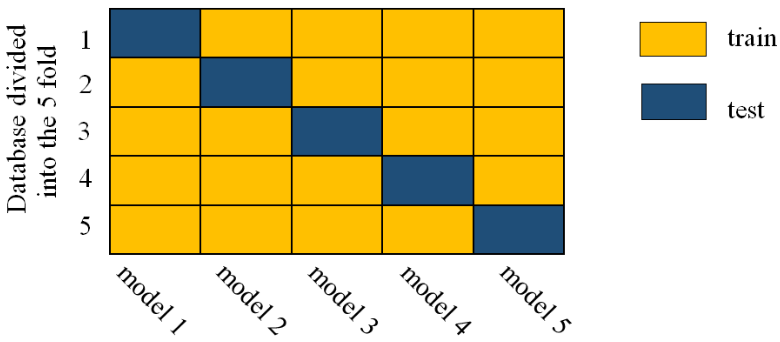

Figure 4: Five-fold cross-validation: division of database for training and testing process

The settings depicted in Figure 4 were repeated eight times and the best recognition results were obtained with the k-nearest neighbours (k-NN) classifier (Altman 1992), belonging to the group of ensemble algorithms running in Matlab environment. The framework for ensemble learning uses various methods that can meld results from many weak learners into one high-quality ensemble predictor (MathWorks 2019). In (k-NN) algorithm, input multivariate vectors are placed in the n-dimensional space in the training process. In the test phase, k-nearest vectors can be found for the classified vector according to the Euclidean metric. The Euclidean distance between vectors m_e is calculated using the formula:

$$m_e(\vec{a},\vec{b}) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2},$$

where, $\vec{a}$ and $\vec{b}$ are two vectors having the same number of elements. The test vector is classified into the class to which most of the closest neighbours belonged.

5. Results

The following results were achieved using the k-NN classifier: from the 60 segments with anticipatory nucleus, 57 segments (95%) were correctly classified. On the other side, 53 segments (88%) were correctly classified from the 60 background segments without anticipation. Thus, the average success rate was 91.5%.

The depiction of the Receiver Operating Characteristic (ROC) curve in Figure 5 shows true positive rate versus false positive rate for the currently selected trained classifier. It represents the best achieved result.

A false positive rate (FPR) of 0.12 indicates that the current classifier assigns 12% of the observations incorrectly to the positive class. A true positive rate of 0.95 indicates that the current classifier assigns 95% of the observations correctly to the positive class.

Area under the curve (AUC) expresses the classification ability of a given classifier. The closer the AUC is to 1, the more accurate the classifier. AUC values of approximately 0.5 represent an ineffective classifier, because the created classifier has no discrimination capacity to distinguish between the positive class and the negative class.

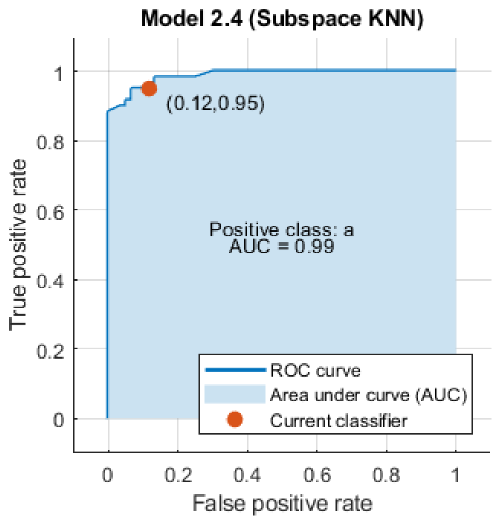

Figure 5: ROC curve shows true positive rate versus false positive rate for the currently selected trained classifier

6. Conclusion

The aim of this study is to test the possibility of detecting the presence of anticipatory nuclei by machine learning algorithms, i.e. to determine whether a mathematical model can perform a similar activity as the brain (learning and testing). The results confirmed that the model can perform this task and it can be effective for the detection of such fine suprasegmental phenomenon – the anticipation. We achieved a true positive rate (TPR) of 95%, which corresponds to the high discrimination ability based on the chosen prosody parameters (the fundamental tone and the intensity). Despite this result, the human brain's activity and methods of thinking and decision making, considering all consequences, are still a unique human skill.

In this paper, we investigated only one level of anticipation – prosodic. However, in real conditions, the interpreter analyses and uses all available information coming not only from speech (a level of semantics, syntax, and prosody), but also from nonverbal behaviour. In the effective anticipation process, all obtained knowledge is helpful, and an interpreter can choose the information that will be a ground truth for the next utterance (i.e. the rest of information serves for confirmation of an anticipated event). If the information coming from all levels is consistent, the anticipation process can be very effective.

In the future, we would like to perform the same experiments with large amounts of data, as well as investigate the impact of prosody on the anticipation process.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under contracts No. APVV-15-0307 and APVV-15-0492.

References

Altman, Naomi S. 1992. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician* 46(3). 175–185.

Butz, Martin V. & Sigaut, Olivier & Pezzulo, Giovanni & Baldassarre, Gianluca (eds.). 2007. *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior*. Berlin & Heidelberg: Springer-Verlag.

Dominey, Peter Ford & Metta, Giorgio & Nori, Francesco & Natale, Lorenzo. 2008. Anticipation and initiative in human-humanoid interaction. *Humanoids 2008 – 8th IEEE-RAS, International Conference on Humanoid Robots*. 693–699.

Kleiber, Georges & Sock, Rudolph. 2006. Ce + N + relative: Sémantique et prosodie. *Lingvisticae Investigationes* 29(2). 251–273.

Koppula, Hema S. & Saxena, Ashutosh. 2016. Anticipating Human Activities Using Object Affordances for Reactive Robotic Response. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38(1). 14–29.

Matos, Ecivaldo. 2008. Anticipation and Human-Computer Interaction: Design, Anticipation and Interaction. *IX Congreso Internacional Interacción Persona-Ordenador*. 463–466.

MathWorks. 2019. *Statistics and Machine Learning Toolbox*™. User's Guide by The MathWorks. (https://ch.mathworks.com/help/stats/index.html)

Ondáš, Stanislav & Pleva, Matúš. 2019. Anticipation and its applications in human-machine interaction. In Barančíková, Petra & Holeňa, Martin & Horváth, Tomáš & Pleva, Matúš & Rosa, Rudolf (eds.), *Proceedings of the 19th Conference Information Technologies - Applications and Theory (ITAT 2019)*, 152–156. CEUR-WS.

Paľová, Mária & Kiktová, Eva. 2019. Prosodic anticipatory clues and reference activation in simultaneous interpretation. *XLinguae* 12(1). 13–22.

Rious, Vincent & Dessante, Philippe & Glachant, Jean-Michel. 2008. Anticipation for efficient electricity transmission network investments. *First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future*. 1–6.

Seeber, Kilian G. (ed.). 2001. *Intonation and anticipation in simultaneous interpreting*. Genève: Université de Genève Ecole de traduction et d'interprétation.

Solé, Maria-Josep & Beddor, Patrice Speeter & Ohala, Manjari (eds.). 2007. *Experimental Approaches to Phonology*. Oxford: Oxford University Press.

Július Zimmermann
LICOLAB (Language Information and Communication Laboratory)
Department of Slovak Studies, Slavonic Philologies, and Communication
Pavol Jozef Šafárik University in Košice
Slovakia
julius.zimmermann@upjs.sk

Eva Kiktová
LICOLAB (Language Information and Communication Laboratory)
Department of Slovak Studies, Slavonic Philologies, and Communication
Pavol Jozef Šafárik University in Košice
Slovakia
eva.kiktova@upjs.sk

In SKASE Journal of Theoretical Linguistics [online]. 2020, vol. 17, no. 4 [cit. 2020-10-27].Available on web page http://www.skase.sk/Volumes/JTL45/pdf_doc/01.pdf. ISSN 1336-782X