

#24



Multiagent model of perceptual space formation in the process of mastering linguistic competence.

Zalimkhan Nagoev¹, Irina Gurtueva^{*1} (gurtueva-i@yandex.ru).

¹Kabardino-Balkarian Scientific Center of Russian Academy of Sciences, I. Armand Street, 37-a, 360000, Nalchik, Russia.



Zalimkhan Nagoev



Irina Gurtueva



ABSTRACT

We propose a simulation model of early development for language competence. It is a model of phonemic imprinting that describes the process of mapping audio stimuli into classes of elementary units of a language, taking into account the influence of social factors. The supervised machine-learning algorithm used in the model was developed using the features of speech addressed to children. This model will allow us to explore the features of phonetic perception, the cognitive mechanisms that underlie language development, highlight the main factors affecting the duration of the period of plasticity. On that basis we can build perceptual maps, create diagnostic tools to describe the sensitive period, which will facilitate the study of the stages of its opening and closing. The model can also be used to create speech recognition systems that are resistant to a variety of accents and effective when used in noisy environment.

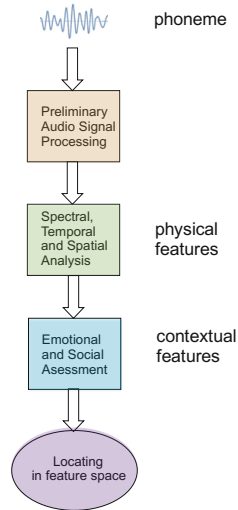
1. INTRODUCTION

Recent progress in speech recognition and related technologies over the past ten years is due to deep learning methods application, including convolutional and recurrent neural networks, to availability of training material, and increase in computing power of computers made. The named factors have brought developments in the sphere of speech systems to a completely new level [1]. To be exact word error rate in automatic conversational telephone speech transcription has been decreased from 22% in 2002 down to 5.1% in 2010. That is why many experts quite reasonably claim that human parity in speech recognition has been achieved. But it is very important to specify the performance metric, task, and human gold standard references chosen for comparative analysis, to describe experiment conditions (language, genre etc.) and indicate test data.

However, taking into account the fact of human parity achieving in a number of speech recognition tasks, modern systems are not perfect still and not all problems in speech recognition are solved. For example, recognition in conditions of high noise and cocktail-party situations have yet to be solved. In addition, it is important to qualify differences between errors made by automatic system and human. Comparative analysis of errors made by automatic recognizer and human in conversational telephone speech transcription showed that the most frequent error of speech system was the substitution of interjection, filling pauses and signaling turn-taking, by interjections of backchannel acknowledgements, signaling, on the contrary, to continue. Speech system tend to make it in 45 case versus 5 cases, when human tend to do it.

Differences in the quality of errors come down to distinguishing between pragmatic functions as modern speech technologies do not use dialogue context modeling. In our opinion, it is necessary to create a system based on human cognitive functions to take into account the context.

2. Stages of processing an audio message by a multi-agent phonemic imprinting model



3. The Algorithm for Perceptual Space Building

Pre-processing converts audio signal into the following set of signatures [3, 4]:

$$\{F_0, F_1, F_2, \Delta t\},$$

where F_0 - the fundamental frequency, F_1, F_2 - the first and second harmonics, Δt - phoneme duration.

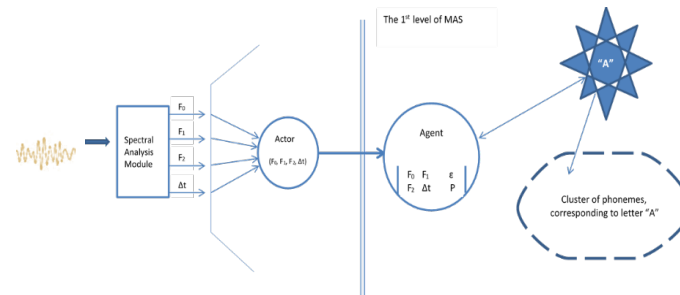
The feature vector is fed to the input of a multi-agent recursive cognitive architecture, where neural factories create agents, determine their type and place them in the corresponding feature space area.

The training material is selected taking into account the features of speech addressed to children [5]. The multi-agent system appends two features to the agent knowledge base - emotional evaluation, and also identifies it as a prototype / non-prototype on the base of comparative assessment of the analyzed phoneme duration and existing statistical estimates of the maternal phoneme duration. The agent knowledgebase is appended:

$$\Phi_{\text{power}} = \{(F_0, F_1, F_2, \Delta t), K_{\text{letter}}, \varepsilon, P\},$$

where K_{letter} - classifying contractual relationship between agent-actor and agent-letter, $\varepsilon \in [0, 1]$ - emotional evaluation, $P(0, 1)$ - prototype/non-prototype.

4. Multiagent model of perceptual space of vowels for native language



5. Machine learning procedure imitating the process of phonemic awareness of a human while mastering native language

The agent, implementing a program determined by its own genome rules addresses the question to the expert. Based on the expert's response, a contract is concluded between a first-level agent-actor and an agent-letter characterizing this class. The prototype/non-prototype feature and the emotional evaluation determine not only the spatial localization of an agent in the multi-agent perceptual space, but also the initial life time of the agent in the system, which is then controlled by a stepwise memory function. In the case of a long period of inactivity of the agent, it is excluded from the system. The settings of the agent's life expectancy parameter will allow us to study the problems of the efficiency of children's mastery of new knowledge, as well as the problems of memorization.

DISCUSSION

The concept is based on the analysis of experimental data from behavioral studies and model ideas about speech recognition mechanisms from the point of view of psycholinguistics and neurolinguistics, fundamental ideas of developmental psychology, cognitive sciences, and the study of maternal speech [5-9]. It allows to track the mechanism of formation of human auditory patterns. It reflects the fact that there is a steady change in the perception of speech, due to the influence of the surrounding linguistic environment [2]. That is, having completed native language acquisition, adults learning a new language are not able to absorb the statistical properties of the phonetic units of the new language implicitly - attention and cognitive efforts are determined by the structure of the mastered category. Increased attention and mental effort are needed to process stimuli that are sharply inconsistent with existing phonetic categories.

It is important to note that the prototype/non-prototype feature and the emotional evaluation determine not only the spatial localization of an agent in the multi-agent perceptual space, but also the initial life time of the agent in the system, which is then controlled by a stepwise memory function. The settings of the agent's life expectancy parameter will allow us to study the problems of the efficiency of children's mastery of new knowledge, as well as the problems of memorization.

The initial parameters of the phonemic prototype system formed at the stage of preliminary training will be refined in the process of speech recognition of a specific speaker. The adjustment will be carried out by shifting the coordinates of the phonemic prototype in the feature space towards the last pronounced sound. It will help to create automatic speech recognition system that turns to speaker and to solve the problem of robustness in noisy environment.

CONCLUSION

The model of phonemic imprinting imitates audio stimuli mapping into classes of minimal language units taking into account the influence of social factors. The machine-learning algorithm used in the model was developed using the features of maternal speech. This approach will allow to build perceptual maps, diagnostic tools for sensitive period description. The model can be used to design automatic speech recognition systems that take into account dialogue context. It can also help to turn to a speaker that, in its turn, make the system resistant to a variety of accents and effective when used in noisy environment.

ACKNOWLEDGMENTS & REFERENCES

The research was supported by the Russian Foundation of Basic Research, grants No. 18-01-00658, 19-01-00648.

1. Stolcke, A., Droppo, J.: Comparing Human and Machine Errors in Conversational Speech Transcription (2017).
2. Strange, W.: Speech Perception and Linguistic Experience (1995).
3. De Cheveigne, A., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music (2002)
4. Maher, R. C., Beauchamp, J. W.: Fundamental frequency estimation of musical signals using a two-way mismatch procedure (1994).
5. Fernald, A., Kuhl P.: Acoustic determinants of infant preference for Motherese Speech (1987).
6. Doupe, A. J., Kuhl, P. K.: Birdsong and human speech: Common themes and mechanisms (2008).
7. Vygotsky, L. S.: Thinking and Speech (2019).
8. Chomsky, N.A.: A Review of Skinner's Verbal Behavior (1967).
9. Haikonen, P.: The Cognitive Approach to Conscious Machines (2003).