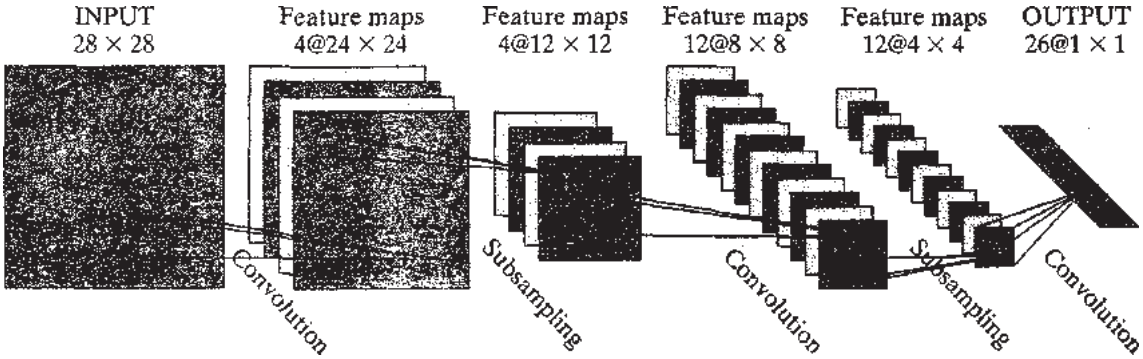

Figure 1. Convolutional neural network for image processing

B. **Experiment**

Experiment was conducted to figure out if an offered fee for a job is satisfactory enough for each applicant. We took 20 participants and interviewed them by Skype. The video series we received were analyzed by CNN real-time. As output, we get the percentage of "happy" and "disappointed" images. On this basis, we made a conclusion if an offer is suitable for each interviewee.

Results and discussion

When we offer the fee, lower than in resembling vacancy on market CNN discovers prevailing percentage of "disappointed" images. It shows us that the applicant found job offer unsatisfying. In the opposite situation, when we offer fee a little bigger than on market, the interviewed find the suggestion attractive for considering.

Using emotion facial recognition during job interview is potentially extremely effective and can be used to improve communications between company manager and applicant. We can see what emotions a person experiences and analyze its answers in different way. CNN is a competent method of processing image data.

The study was produced with the financial support of the Russian Foundation for Fundamental Research. (project № 14-07-00603 " Intelligent cognitive systems for decision support")

Bengio Y. 2009. Learning deep architectures for AI, Now Publishers Inc.

Haykin S. 1999. Neural Networks: A Comprehensive Foundation, 2nd Edition, Prentice-Hall.

Miyakoshi, Yoshihiro and Kato Sh. 2011. Facial Emotion Detection Considering Partial Occlusion Of Face Using Baysian Network. Symposium on Computers & Informatics.

STIMULSTAT: A DATABASE FOR LINGUISTIC AND PSYCHOLOGICAL STUDIES ON RUSSIAN LANGUAGE

S.V. Alexeeva[1], N.A. Slioussar[1, 2], D.A. Chernova[1]
mail@s-alexeeva.ru, slioussar@gmail.com, chernovadasha@yandex.ru
[1]St. Petersburg State University (St. Petersburg, Russia), [2]School of Linguistics, HSE (Moscow, Russia)

Experimental studies conducted by linguists, psychologists and other researchers identified a large list of word properties that play a role for speech production and comprehension. They include lemma and form frequency, word length, the number of syllables, stress pattern, frequency of letters and syllables the word consists of, whether the word has homonyms, homographs or orthographical neighbors, whether it has multiple senses etc. Of course, various grammatical characteristics (part of speech, inflectional paradigm etc.) are also important.

Taking these properties into account in new studies became an important problem. For several languages, databases with search tools were designed to solve this problem. Among them are the English lexicon project (http://elexicon.wustl.edu) (Balota et al. 2007); DlexDB for German (http://www.dlexdb.de) (Heister et al. 2011); the CELEX Lexical Database for Dutch, English and German (http://wwwlands2.let.kun.nl/members/software/celex.html); N-Watch (Davis 2005) and MRC databases (Colheart 1981) for English; Lexique (http://www.lexique.org) (New et al. 2004) for French; BuscaPalabras (Davis, Perea 2005) for Spanish and E-Hitz (Perea et al. 2006) for Basque.

As for Russian, some required characteristics are not represented in electronic dictionaries and databases at all, while the others are scattered across frequency lists, grammatical or explanatory dictionaries or others resources that usually do not have a user-friendly interface for filtering words. In this paper, we present StimulStat, the first searchable

database for Russian that includes more than 50000 most frequent Russian words (> 1 900 000 word forms) characterized according to more than 50 properties that were demonstrated to play a role in linguistic and psychological research. The database can be used for stimulus selection in experimental studies of Russian and provides a lot of information that may be relevant in other linguistic domains because one can easily calculate how words with many combinations of different characteristics are distributed, which patterns are widespread and which are infrequent.

The database represents such parametres as lemma-based and word form-based statistics (length in letters, length in syllables, frequency etc), phonological information (stress position, stress pattern — CVCCV for the word *кошка* 'cat' — etc.), morphological information for each part of speech (including Zalyznyak index), semantic information (number of meanings) and information about different types of orthographical neighbors of the word — visually similar words like *сок* 'juice' vs. *ток* 'current' (substitution neighbors), *бук* 'beech' vs. *куб* 'cube' (transposition neighbors), *карат* 'carat' — *кара* 'penalty' (addition/deletion neighbors), *абориген* 'local native' vs. *бор* 'pine forest' (part/whole neighbors) etc. (Andrews 1997). We also counted the uniqueness point (OUP) or the letter position in the word where that word is differentiated from other words. Recent research found that naming and lexical decision times for words with an early orthographic uniqueness point (OUP) were faster than for words with a late OUP (Kwantes & Mewhort 1999a, Lindell, Nicholls, & Castles 2003). We found out that the mean lemma-based length in Russian corrected by the frequency is 5 and the mean frequency is 18,5 ipm.

The main linguistic sources of our database are the Russian word frequency dictionary (Lyashevskaya, Sharov 2009), the digital versions of the Grammatical dictionary of the Russian Language by A.A. Zaliznyak (Zaliznyak 1977) and the Explanatory dictionary of the Modern Russian Language by T.V. Efremova (Efremova 2000), the morphological parser pymorphy2 (https://pymorphy2.readthedocs.org/en/latest/) which is based on the OpenCorpora dictionary (opencorpora.org) (Bocharov et als. 2013).

These parametres are to be taken into account in a number of psycholinguistics experiments investigating word recognition (using such methods as lexical decision task, naming task, perceptual identification task, semantic categorization task) and reading processes and can be checked with the help of our database which has already been used in (Frolova 2014).

The database is available at http://stimul.cognitivestudies.ru

This project is supported by the grant 14-04-12034 from the Russian Foundation for Humanities.

Andrews S. 1997. The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. Psychonomic Bulletin & Review, 4, 439-461.

Balota D.A., Yap M.J., Cortese M.J., Hutchison K.A., Kessler B., Loftis B., Neely J.H., Nelson D.L., Simpson G.B., Treiman R. 2007. The English Lexicon Project. Behavior Research Methods 39, 445-459.

Bocharov V.V., Alexeeva S.V., Granovsky D.V., Protopopova E.V., Stepanova M.E., Surikov A.V. 2013. Crowdsourcing morphological annotation. Computational Linguistics and Intellectual Technologies: DIALOG 2013, vol. (12) 19. Russian State Humanitarian University, Moscow, 109-114.

Coltheart M. 1981. The MRC Psycholinguistic Database. Quarterly Journal of Experimental Psychology 33A, 497-505.

Davis C.J. 2005. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. Behavior Research Methods 37, 65-70.

Davis C.J., Perea M. 2005. BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. Behavior Research Methods 37, 665-671.

Efremova, T. 2000. The New Explanatory Dictionary of the Russian Language // ed. T. Efremova. Moscow: Russian language. In Russian. (original title: Efremova, T., (2000). Novy slovar russkogo yazyka. Tolkovo-slovoobrazovatelny. Moskva: Russkij yazyk.)

Frolova A.M. 2014. Aspects of speech segmentation: an experimental study with reference to Russian (master thesis) [Osobennosti segmentatsii ustnoy rechi: eksperimentalnoye issledovaniye na material russkogo yazyka (magisterskaya dissertatsiya)], St.Petersburg State University (http://phil.spbu.ru/ucheba-1/zaschita-magisterskih-rabot-v-2014-g/kafedra-obschego-yazykoznaniya)

Heister J., Würzner K. — M., Bubenzer J., Pohl E., Hanneforth T. et al. 2011. DlexDB — eine lexikalische Datenbank für die psychologische und linguistische Forschung. Psychologische Rundschau 62, 10-20.

Kwantes P. Mewhort D. 1999a. Evidence for sequential processing in visual word recognition. Journal of Experimental Psychology: Human Perception and Performance 25 376-381

Lindell A., Nicholls M., Castles A.E. 2003. The effect of orthographic uniqueness and deviation points on lexical decisions: Evidence from unilateral and bilateral-redundant presentations. Quarterly Journal of Experimental Psychology 56A 287-307.

Lyashevskaya O., Sharov S. 2009. The frequency dictionary of modern Russian language Moscow: Azbukovnik. In Russian (Lyashevskaya O., Sharov S. (2009). Chastotny slovar sovremennogo russkogo yazyka (na materialah Natsionalnogo korpusa russkogo yazyka. Moskva: Azbukovnik)).

New B., Pallier C., Brysbaert M., Ferrand L. 2004. Lexique 2: A new French lexical database. Behavior Research Methods 36, 516-524.

Perea M., Urkia M., Davis C.J., Agirre A., Laseka E., Carreiras M. 2006. E-Hitz: A word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). Behavior Research Methods 38, 610-615.

Zaliznyak A.A. 1977. Grammatical Dictionary of the Russian Language. Moscow, Russia. In Russian. (original title: Zaliznyak, A., A., (1977). Grammaticheskii slovar russkogo yazika: Moskva, Rossiya.)