A comparative study on modeling and controlling emotional acoustic parameters in neural networks based Japanese and Spanish speech synthesis

Jaime Lorenzo-Trueba[1,a) Shinji Takaki[1,b)] Junichi Yamagishi[1,2,c)]

Abstract: In neural network based speech synthesis, adding a one-hot vector to the input is an easy, intuitive but useful way to model multiple speakers or multiple language. In this paper, we use the one-hot vector for modeling and controlling emotional acoustic parameters. We have used Spanish and Japanese databases having the multiple acted emotional speech uttered by professional speakers in our experiment and will show the performance of the one-hot vector approach.

Keywords: Emotional speech synthesis, Deep neural network, recurrent neural networks

1. Introduction

Speech synthesis is a technique that is aimed at generating natural-sounding intelligible speech from arbitrary tests. This has been studied for a long time, and the most popular successful approaches that have been successful include unit selection (US) techniques and statistical parametric techniques such as hidden Markov models (HMMs). Neural network (NN)-based methods are currently being actively investigated because they are able to significantly improve the quality and naturalness of synthetic speech compared to traditional approaches [1]. First, feed forward (FF) deep neural network (DNN) systems have been proposed as a replacement for decision tree approaches in HMM-based speech synthesis, which have demonstrated better proficiency at handling large amounts of speech data than traditional approaches [2]. Long short-term memory (LSTM)-based recurrent neural networks (RNNs) have recently been adopted and it has been reported that they provide even better naturalness and prosody of synthetic speech due to their capabilities for modeling the long-term dependencies of speech [3].

However, there have been few publications on expressive speech synthesis and DNN and hence the advantages of DNN-based approaches for expressive speech synthesis are not very clear. We think that this is partially because existing expressive speech databases normally have fewer speech data than normal reading speech data, i.e., typically less than 1,000 sentences per emotion or speaking style, which may not be sufficient to train high-quality DNN based acoustic models. However, it has been reported that expressive speech synthesis techniques improve the usability of speech interfaces and spoken dialogue systems [4], and it is therefore meaningful and important to analyze how DNNs may improve the performance of expressive speech synthesis systems [5].

When there are less or somewhat limited speech data available for each emotion, we can pool speech data in different emotions together and estimate an emotion/style mixed model for HMM-based speech synthesis systems [6]. We can also consider using adaptation techniques [7] so that the acoustic model for a new emotion can be effectively estimated using small amounts of speech data. We can also use several other techniques such as interpolation [8] or transplantation [9] to control and manipulate emotional synthetic speech. However, what can we achieve in expressive speech modelling in DNN speech synthesis?

With that objective in mind, the present research was aimed at analyzing if DNN-based approaches were capable of generating recognisable and controllable emotional synthetic speech comparable to what was achievable with traditional HMM-based systems when using the same amount of speech data. We propose the use of an emotional one-hot vector as an additional auxiliary input to DNN-based acoustic models to attain that purpose, and we tried to build common DNNs that shared speech data across multiple emotions and jointly model all the different emotions that were included in a training corpus. Although the emotional one-hot vector is simple, it is very convenient for generating synthetic speech because we can hopefully select and/or control the vector during synthesis to manipulate the emotions of synthetic speech. There are has been little relevant research, although some examples are those by Li and Zen [10], who have built multi-language and multi-speaker models by sharing data across languages and speakers. Luong et al. [11] shared speech data across speakers and built multi-speaker models from over 100 speakers using speaker code vectors. We measured and compared the emotional

[1] National Institute of Informatics, Tokyo, Japan
[2] The University of Edinburgh, Edinburgh, UK
a) jaime@nii.ac.jp
b) takaki@nii.ac.jp
c) jyamagish@nii.ac.jp

modelling accuracy of feedforward and LSTM-based RNN-based systems using the emotional one-hot vector with HMM adaptation methods by means of a perceptual evaluation in terms of emotion recognition rates, perceived emotional strength and perceived speech quality. Finally we also wanted to evaluate how different languages and different amounts of training data impacted the control capabilities of the considered systems, for which we carried out an additional similar evaluation.

The paper is structured as follows: The proposed systems are explained in Section 2. Section 3 introduces the emotional speech corpus considered for the development of the proposed systems and the perceptual evaluation design is explained in section 4. Finally, Sections 5 and 6 present the results of the perceptual evaluations. Section 7 draws some global conclusions and discusses work we expect to do in the future.

2. DNN-based Emotional Speech Synthesis Systems

We needed to train NN-based emotional speech synthesis models that could effectively handle a set of small amounts of speech data in different emotions in this research, because the quantity of speech data in each emotion may be too limited, as we described earlier. If we were to train a DNN model on a small number of utterances, it is obvious that the results would not be satisfactory. As such, the main objective was to find a way of training the model using all emotional data at the same time, while being able to produce each of the required emotions at synthesis time. In addition to that, we also wanted to control the strength of the produced expressive synthetic speech so that we could emphasize or de-emphasize it according to our interests, similarly to what interpolation techniques have achieved in HMM-based systems.

We propose the addition of an emotional one-hot vector code that represents an acoustic emotional category of an utterance as the additional auxiliary inputs to DNN-based acoustic models to achieve that purpose, as outlined in Figure **Fig. 1**. If we have C-representative emotions, the one-hot vector $\boldsymbol{e}_i$ for the i-th emotion is defined as $\boldsymbol{e}_i = (e_1, e_2, ..., e_C)$, where each value e_c is given by:

$$e_c = \begin{cases} 0 & (c \neq i), \\ 1 & (c = i). \end{cases} \qquad (1)$$

where each subscript indicates each emotional category of the speech utterance that was used. This simply means that only the i-th element will be set to 1 for emotion i, and the rest of the elements will be 0. If annotation for the degree and strength of emotional speech data is available, we can replace the binary values to the numerical scores that represent the emotional strength. At synthesis time, the generation of synthetic speech in the required emotion can be achieved by inputting text features and the one-hot vector where the component of the target emotion is activated.

We can expect that this will also allow us to produce synthetic speech in intermediate or emphasised emotions by specifying and controlling the components of the emotional one-hot vector. For instance, we can set the value of the one component of the one-hot vector to 0.5, and we can expect that the degree and strength of the corresponding perceived emotions of synthetic speech will be de-emphasized, and if we set the value to 2.0, it may emphasise the degree and strength of perceived emotions of synthetic speech, as could be done with multi-regression HMM systems [12].

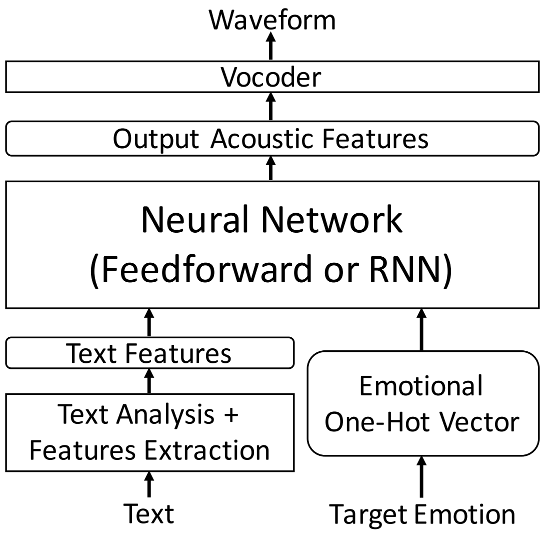

Fig. 1 Flowchart for the proposed DNN speech synthesis system using the emotional one-hot vector.

We implemented feed forward-based and recurrent neural network-based speech synthesis systems using this emotional one-hot vector. The following subsections describe the details of the architectures we used.

2.1 Emotional adaptive HMM-based system

First of all we implemented a HMM-based system for purposes of comparison. We had several options for the HMM system and we chose HMM adaptive systems in this research because this was one of the most frequently evaluated systems in past research [13], [14]. All the emotional data was used to train an average model using speaker adaptive training (SAT) [15]. Then, the different emotional models were adapted using constrained structural maximum a posteriori linear regression (CSMAPLR) [7]. The same linear regression functions were then used for manipulating and controlling the emotional strength of synthetic speech. For more details, please refer to [16].

The HMM speaker models were trained with three feature streams with their Δ and Δ^2 coefficients: logarithm of the F_0 (1 coefficient), 60 Mel-cepstral coefficients and aperiodicity bands (25 coefficients). For the synthesis part we used the STRAIGHT vocoder [17].

2.2 FF-based neural network system

The FF-based system we used had five layers and 1024 units per layer. The sigmoid function was used for all the units for both hidden and output layers. They were randomly initialized and trained to minimise mean square error using stochastic gradient descent. The linguistic features consisted of 253 features (including the emotional one-hot vector) with the phoneme boundaries estimated by forced alignment of a HMM model and they

Table 1 Description of the Japanese emotional speech database. Durations are including silences and expressed in minutes. Rhythm excludes silences and is expressed in phones per second. Total rhythm shows the average rhythm of the database.

Emotion	Sentences	Duration	Rhythm
Neutral	1200	147 min	10.39 phones/sec
Happy	1200	133 min	10.90 phones/sec
Sad	1200	158 min	9.04 phones/sec
Calm	1200	154 min	9.05 phones/sec
Insecure	1200	141 min	9.88 phones/sec
Excited	1200	136 min	10.51 phones/sec
Angry	1200	148 min	9.26 phones/sec
Total	8400	1017 min	9.86 phones/sec

were normalised to zero-mean unit-variance. The acoustic features consisted of 259 dimensions: 60 Mel-cepstral coefficients, interpolated log F_0 and the voiced/unvoiced parameter, 25 band aperiodicity coefficients and also their Δ and Δ^2. We used the WORLD vocoder [18] to generate the speech waveform [*1] .

2.3 RNN-based based neural network system

The RNN system utilized two feed forward layers with 512 nodes per layer and two bi-directional recurrent layers with 256 LSTM units per layer. This LSTM-based RNN system was implemented using the CURRENNT toolkit [19]. The acoustic and linguistic features are the same as those of the feed forward system, and the networks were trained based on random initialization and stochastic gradient descent with early stopping and parallel sentence training with 20 sentences to accelerate the training time of each epoch [19]. WORLD vocoder was also used for generating the waveforms.

3. Emotional Speech Corpus

3.1 Spanish Corpus

For the Spanish evaluation we have used the Spanish Expressive Voices (SEV) speech corpus [20]. The database consists of two professional speakers, a male and a female, who read various texts in four acted emotions (anger, happiness, sadness and surprise) plus neutral speech in an acoustically treated room. All emotional speech utterances were recorded using the same set of 489 utterances, which were carefully designed so that they would provide no emotional information and were phonetically balanced. On average, each emotional speech subset consisted of around 30 minutes of speech data, meaning a total of 2445 utterances (about 150 minutes of data). A comparison of US and HMM synthesis built on the SEV corpus has been reported by Barra-Chicote et al. [21].

3.2 Japanese Corpus

For the Japanese evaluation we recorded our on data in a professional recording studio. We recorded three pairs of acted emotions for a professional female speaker: happy - sad, calm - insecure, excited - angry plus neutral read speech. A detailed description of the amounts of data per emotion can be seen in table **Table 1**. The phone alignments used to estimate rhythm were obtained from the results of HMM forced alignment.

When designing the recording sentences we wanted to consider more conversational sentences instead of typical news texts, so that they would be easier to interpret for the voice actress. Most of the sentences were neutral and shared across emotions (see **Table 2**), but in the case of novel sentences, which could potentially possess emotional context, were filtered so that the recorded emotion was in accordance to the text.

[*1] The differences between the STRAIGHT and WORLD vocoders are only a few patented operations and their synthetic speech sounds similar.

Table 2 Description of the Japanese emotional database recording sentences. Shared column specifies if the sentences were shared across emotions.

Source	Sentences	Shared
News	101	Yes
Novel	313	No
Ted Talks	196	Yes
Car Navigation System	200	Yes
MULTEXT [22]	191	Yes
Phonetically Balanced	199	Yes
Total	1200	Depends

4. Perceptual Evaluation

The main objective of this perceptual evaluation was to compare the overall performance and controllability of the three emotional speech synthesis systems that were described earlier. The perceptual evaluation measured three aspects of emotional synthetic speech: speech quality, emotional strength, and emotion identification rates. All systems of the same language were evaluated in the same perceptual test.

4.1 Evaluation design

The evaluations were carried out by means of a Web interface, where a single audio sample was presented to the listener without any additional information about the synthesized text or emotion. The samples could be played on demand by the listeners. First, listeners were asked to to identify the emotion conveyed by the utterance from an open list of the synthesized evaluations also including the ”other” and ”neutral” option. Then, they were asked to rate the perceived emotional strength on a 5-points MOS scale and then to rank their perceived speech quality also on a 5 points MOS scale. Systems were presented to the listeners in random order without repetitions. Because the evaluation also took emotion identification rates into consideration, the utterances were chosen to be devoid of any emotional context so that they did not influence the evaluation.

4.1.1 Spanish Evaluation

Three emotions (anger, happiness and surprise) were evaluated using a number of different control factors. Since the values of the emotional one-hot vector and the values used for the HMM systems may not have corresponded to the same degree of intended emotions, we manually adjusted the values so that all the systems became comparable (we used 0.5, 1.0, 1.5, and 2.0 for the FF and RNN systems, and 0.5, 0.75, 1.0, and 1.25 for the HMM system). This meant that we had a total of 40 systems (4 control ratios x 3 emotions x 3 systems + 3 neutral systems + 1 natural read speech voice sample). We adopted the Latin-square approach [23] that requires 40 different utterances to be evaluated for all the systems.

4.1.2 Japanese Evaluation

In this case six emotions (happiness, sadness, calmness, inse-

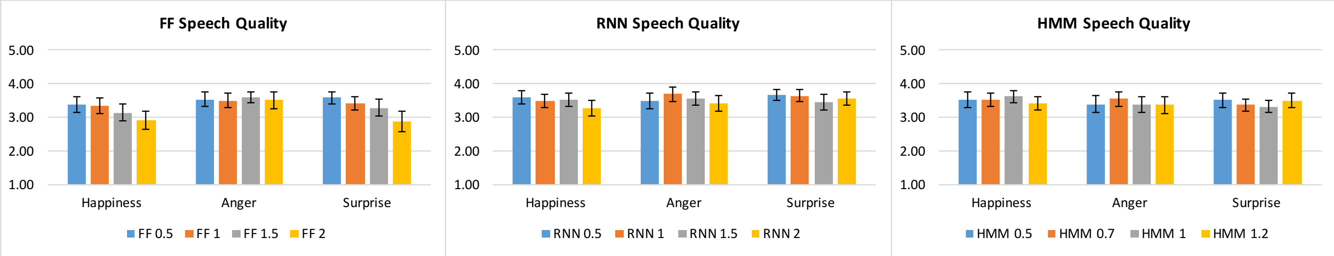

Fig. 2 Speech quality results. Numbers in legends refer to control vector components. Bars indicate 95% confidence intervals.

curity, excitement, anger) were evaluated with four control factors (0.5, 1.0, 1.5 and 2.0), plus neutral speech. In this case only the RNN-based system was evaluated as we had already established with the Spanish systems evaluation that it provided the best results. In total we evaluated 25 systems (4 control ratios x 6 emotions + 1 neutral system).

5. Spanish Evaluation Results

A total of 40 native speakers of Spanish took part in the evaluation. A total of 1600 utterances were evaluated and analyzed in three variables: speech quality, perceived emotional intensity, and emotion identification rates. The following subsections explains the details.

5.1 Speech quality

Fig. 2 has the average scores for speech quality with 95% confidence intervals based on Student's t-distribution. First, we can observe that almost all the systems provided very similar results, apart from the FF2 of surprise, which was statistically significantly worse than its FF0.5. We believe that this is still a good result considering that we had trained the NN-based systems by only using 490 utterances per emotion.

Due to the emotional code, we could share the speech data of multiple emotions and could robustly train a system that can generate synthetic speech that was comparable to conventional well-trained HMM systems in terms of speech quality.

5.2 Perceived emotional intensity

Fig. 3 has the average scores for the perceived emotional intensity with 95% confidence intervals based on Student's t-distribution. We can clearly see from the results that the increasing values of the emotional one-hot vector of the RNN-based system yielded higher scores for perceived emotional intensity and this was similar to or better than those for the HMM-based approach. However, the trends and patterns for the FF system were not clear.

Although this was mostly a trend that we could identify in the average values (statistically significant differences could be observed between HMM 1 and HMM 1.2 for surprise), we considered this to be a promising result. There are two reasons for this. The first is that the modeling of multiple emotional synthetic speech and the manipulation of emotional synthetic speech were done using a single model without any additional post-processing. The second is that we did not use any manually annotated emotional strength values; although all the emotional data were valued at 1.0 as the emotional strength, we were able to control the perceived emotional strength of synthetic speech to some extent. We therefore believe that training these systems with a properly labelled database will significantly help in the task of controlling the intensities of perceived emotions.

5.3 Emotion identification rates

Fig. 4 plots the emotion identification rates in percent. We can see the results for each emotion in the three systems we propose. The vertices represent the control vector components (0.5, 1.0, 1.5, and 1.5 correspond to the 0.5, 0.75, 1.0 and 1.25 control factors in the HMM system). The identification rates axis is represented by the grey squares, with the smallest being 20% and the largest 100%, in steps of 20%. Finally, the dots represent the emotion identification rate itself.

We can see several interesting aspects from the results. First, the emotion identification rates for the FF-based system are worse than those of the baseline HMM-based system for happiness and surprise. The RNN-based system provides comparable or better emotion identification rates for happiness and anger. We can particularly see that the RNN system improves the recognition rates for happiness very significantly, which occurred due to a reduction in confusion with surprise. The RNN system, on the other hand, had somewhat worse emotion identification rates than the HMM system for surprise. This generally lead us to believe that we were capable of controlling the emotional expressiveness of synthetic voices due to the emotional one-hot vector, at least within some ranges.

6. Japanese Evaluation Results

A total of 16 native Japanese listeners took part in the evaluation of the Japanese systems. As mentioned in section 4.1, we have only considered the LSTM system here as we had already confirmed from previous experiments that we could achieve the best results with the RNN-based system.

6.1 Speech quality

In **Fig. 5**, we see the results for speech quality with 95% confidence intervals based on Student's t-distribution. Apart from the sad case, we can see that even if we set the emotion control factor higher than 1, we always obtain good scores higher than 3 in the MOS scale, without significant degradation introduced by the emotional control vector manipulation. We believe that the decrease in speech quality for sadness is due to problems in F0 estimation that introduced several artifacts.

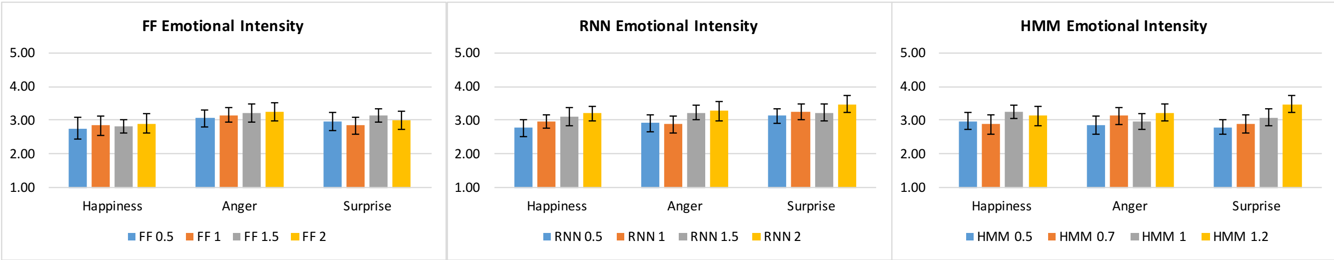

Fig. 3 Spanish evaluation perceived emotional intensity results. Numbers in legends refer to control vector components. Bars indicate 95% confidence intervals.

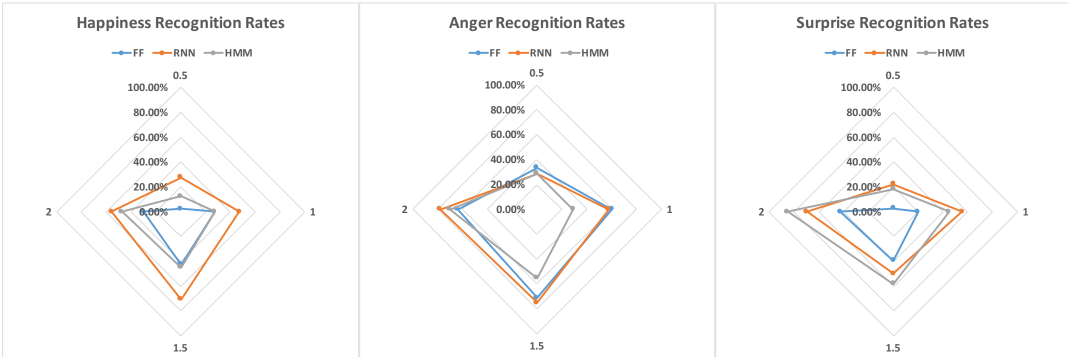

Fig. 4 Radio plots for emotion identification rates, in percent.

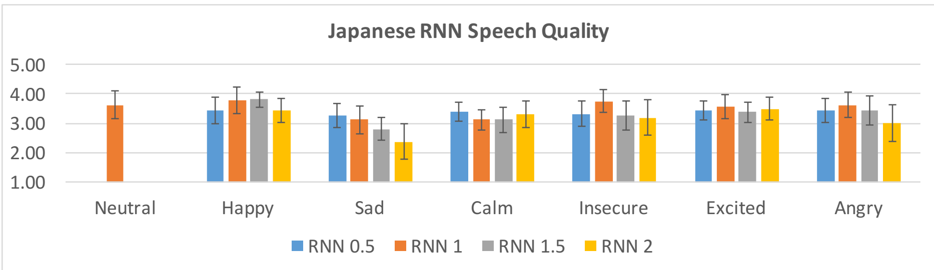

Fig. 5 Japanese evaluation speech quality results in MOS scale. Numbers in legends refer to control vector components. Bars indicate 95% confidence intervals.

6.2 Perceived emotional intensity

Fig. 6 presents the perceived emotional intensity evaluation results with 95% confidence intervals based on Student's *t*-distribution. We see that de-emphasizing the emotional intensity is possible for all the emotions, but, it is not always possible to emphasize the emotional intensity of all the emotions precisely. But we see that at least for happy, insecure, excited and angry there is a significant improvement in the perceived emotional intensity from the RNN 0.5 system to the RNN 1.5 or 2 system. This proves that the proposed system is somehow capable of enhancing the expressive capabilities of the synthesizer.

6.3 Emotion identification rates

Fig. 7 shows the results for the Japanese evaluation of emotion identification rates. No confidence intervals are shown because all the differences were not statistically significant. However, the identification results are always higher than the random threshold (approx. 14%), and are very high for happy, insecure and angry. It indicates that speech in the emotions were clearly and distinctively produced by the speaker. However modifying the emotional vector did not change the emotion identification rates.

7. Conclusions and Future Work

This paper proposed and discussed our evaluation of FF and RNN-based emotional speech synthesis systems, where a common neural network was trained using speech data in multiple emotions jointly together with the emotional one-hot vector as an additional auxiliary input. Using the neural network-based acoustic model, we also investigated whether the emotional one-hot vector was useful for manipulating and controlling the perceived emotional intensity of synthetic speech to be generated. We also compared the capabilities of the proposed system across language. Perceptual evaluation tests on emotion recognition rates, perceived emotional strength, and perceived speech quality were carried out to compare the performance of neural network systems with conventional HMM-based systems.

The experimental results revealed that reasonable quality of synthetic speech can be constructed by using the emotional one-hot vector and mixing data of multiple emotions together. Moreover, we demonstrated that we could control the intensity of the perceived emotions of synthetic speech by manipulating the emotional one-hot vector. These are very promising results because

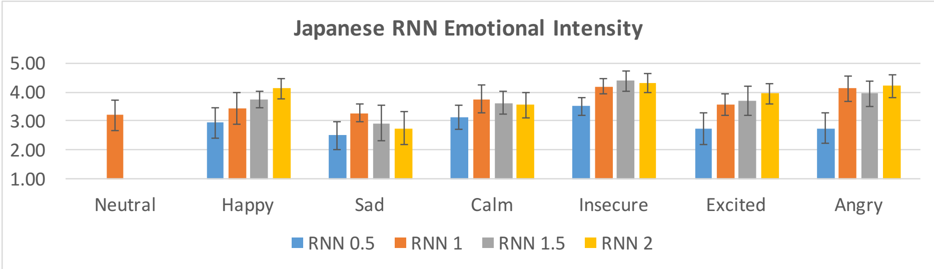

Fig. 6 Japanese evaluation perceived emotional intensity results in MOS scale. Numbers in legends refer to control vector components. Bars indicate 95% confidence intervals.

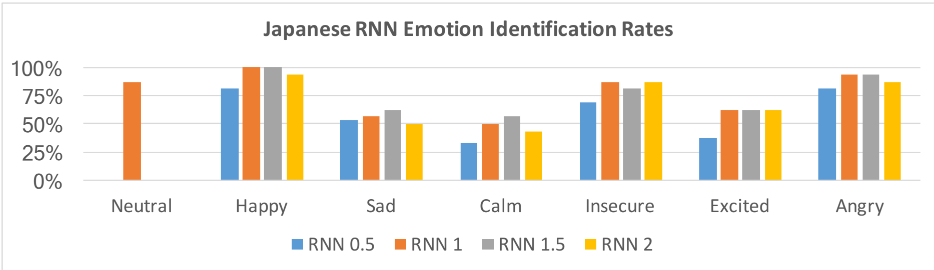

Fig. 7 Japanese evaluation emotion recognition results in percents. Numbers in legends refer to control vector components.

we did not use any manually annotated emotional strength values for our database. We also found that the RNN-based system resulted in more accurate emotional identification rates. When considering the differences between languages, we can see how both Japanese and Spanish follow the same trends even when considering database differences.

Our future work includes the use of an emotional database with labelled emotional strength values, which we suspect will significantly improve the proposed RNN system. We also want to consider emotion representations othe than one-hot vectors.

References

[1] Zen, H.: Acoustic Modeling in Statistical Parametric Speech Synthesis-From HMM to LSTM-RNN (2015).

[2] Wang, X., Takaki, S. and Yamagishi, J.: A Comparative Study of the Performance of HMM, DNN, and RNN based Speech Synthesis Systems Trained on Very Large Speaker-Dependent Corpora, *9th Speech Synthesis Workshop (SSW9)* (2016).

[3] Fan, Y., Qian, Y., Xie, F.-L. and Soong, F. K.: TTS synthesis with bidirectional LSTM based recurrent neural networks., *Interspeech*, pp. 1964–1968 (2014).

[4] Lutfi, S. L., Fernández-Martínez, F., Lorenzo-Trueba, J., Barra-Chicote, R. and Montero, J. M.: I Feel You: The Design and Evaluation of a Domotic Affect-Sensitive Spoken Conversational Agent, *Sensors*, Vol. 13, No. 8, pp. 10519–10538 (2013).

[5] Watts, O., Wu, Z. and King, S.: Sentence-level control vectors for deep neural network speech synthesis, *Proc. Interspeech* (2015).

[6] Chen, L., Gales, M., Wan, V., Latorre, J. and Akamine, M.: Exploring Rich Expressive Information from Audiobook Data Using Cluster Adaptive Training, *Interspeech 2012, 13th Annual Conference of the International Speech Communication Association. Portland, Oregon. September 9-13* (2012).

[7] Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K. and Isogai, J.: Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm, *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 17, No. 1, pp. 66–83 (2009).

[8] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T. and Kitamura, T.: Speaker interpolation in HMM-based speech synthesis system., *EUROSPEECH* (1997).

[9] Lorenzo-Trueba, J., Barra-Chicote, R., San-Segundo, R., Ferreiros, J., Yamagishi, J. and Montero, J. M.: Emotion transplantation through adaptation in HMM-based speech synthesis, *Computer Speech & Language* (2015).

[10] Li, B. and Zen, H.: Multi-Language Multi-Speaker Acoustic Modeling for LSTM-RNN based Statistical Parametric Speech Synthesis (2016).

[11] Luong, H.-T., Takaki, S., Henter, G. E. and Yamagishi, J.: ADAPTING AND CONTROLLING DNN-BASED SPEECH SYNTHESIS USING INPUT CODES, *Under review for ICASSP 2017* (2017).

[12] Nose, T., Kato, Y. and Kobayashi, T.: Style estimation of speech based on multiple regression hidden semi-Markov model., *INTERSPEECH*, pp. 2285–2288 (2007).

[13] Latorre, J., Wan, V., Gales, M. J., Chen, L., Chin, K., Knill, K. and Akamine, M.: Speech factorization for HMM-TTS based on cluster adaptive training., *INTERSPEECH* (2012).

[14] Nose, T. and Kobayashi, T.: An intuitive style control technique in HMM-based expressive speech synthesis using subjective style intensity and multiple-regression global variance model, *Speech Communication* (2012).

[15] Anastasakos, T., McDonough, J. and Makhoul, J.: Speaker adaptive training: A maximum likelihood approach to speaker normalization, *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, Vol. 2, IEEE, pp. 1043–1046 (1997).

[16] Lorenzo-Trueba, J., Watts, O., Barra-Chicote, R., Yamagishi, J., King, S. and Montero, J. M.: Simple4all proposals for the ALBAYZIN evaluations in speech synthesis, *Proc. Iberspeech*, Vol. 2012 (2012).

[17] Kawahara, H., Estill, J. and Fujimura, O.: Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT, *Proc. MAVEBA*, pp. 13–15 (2001).

[18] M. Morise, F. Y. and Ozawa, K.: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, *IEICE transactions on information and systems*, Vol. E99-D, No. 7, pp. 1877–1884 (2016).

[19] Weninger, F., Bergmann, J. and Schuller, B.: Introducing CURRENNT–the Munich open-source CUDA RecurREnt neural network toolkit, *Journal of Machine Learning Research*, Vol. 16, No. 3, pp. 547–551 (2015).

[20] Barra-Chicote, R., Montero, J. M., Macias-Guarasa, J., Lufti, S., Lucas, J. M., Fernandez, F., D'haro, L. F., San-Segundo, R., Ferreiros, J., Cordoba, R. and Pardo, J. M.: Spanish Expressive Voices: Corpus for Emotion Research in Spanish, *Proc. of LREC* (2008).

[21] Barra-Chicote, R., Yamagishi, J., King, S., Montero, J. M. and Macias-Guarasa, J.: Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech, *Speech Communication*, Vol. 52, No. 5, pp. 394–404 (2010).

[22] Shigeyoshi, K., Tatsuya, K., Kazuya, M. and Toshihiko, I.: Preliminary study of japanese MULTEXT: a prosodic corpus, *International Conference on Speech Processing, Taejon, Korea*, pp. 825–828 (2001).

[23] Gao, L.: *Latin Squares in Experimental Design*, Michigan State University (2005).